



DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

DEDUCTIVE BELIEF DYNAMICS

Theoretical Foundations and Experimental Study of an Argumentative Model of the Belief System

Daniel Rivera Guajardo
DOCTORAL THESIS || 2024



UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

Daniel Rivera Guajardo

**DEDUCTIVE BELIEF DYNAMICS:
THEORETICAL FOUNDATIONS AND EXPERIMENTAL
STUDY OF AN ARGUMENTATIVE MODEL OF THE BELIEF
SYSTEM**

TESIS DOCTORAL

DIRIGIDA POR:

Dr. Josep Demestre Viladevall

Dr. Teun A. van Dijk

Departamento de Psicología



UNIVERSITAT ROVIRA I VIRGILI

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo



UNIVERSITAT ROVIRA I VIRGILI
Departament de Psicologia

Carretera de Valls, s/n
Campus Sescelades
43007
Tel: 977 55 80 75

FAIG CONSTAR que aquest treball, titulat "Deductive belief dynamics: theoretical foundations and experimental study of an argumentative model of the belief system", que presenta Daniel Rivera Guajardo per a l'obtenció del títol de Doctor, ha estat realitzat sota la meva direcció al Departament de Psicologia d'aquesta universitat, i que la tesi compleix tots els requisits per poder optar a la Menció de Doctorat Internacional.

HAGO CONSTAR que el presente trabajo, titulado "Deductive belief dynamics: theoretical foundations and experimental study of an argumentative model of the belief system", que presenta Daniel Rivera Guajardo para la obtención del título de Doctor, ha sido realizado bajo mi dirección en el Departamento de Psicología de esta universidad, y que la tesis cumple con los requisitos para poder optar a la Mención de Doctorado Internacional.

I STATE that the present study, entitled "Deductive belief dynamics: theoretical foundations and experimental study of an argumentative model of the belief system", presented by Daniel Rivera Guajardo for the award of the degree of Doctor, has been carried out under my supervision at the Department of Psychology of this university, and that it fulfils all the requirements to be eligible for the International Doctorate Mention.

Tarragona, 13/05/24

El/s director/s de la tesi doctoral
El/los director/es de la tesis doctoral
Doctoral Thesis Supervisor/s

DEMESTRE
E
VILADEVALL, JOSEP
(FIRMA)

Firmado digitalmente por DEMESTRE VILADEVALL, JOSEP (FIRMA)
Fecha: 2024.05.13
12:53:41 +02'00'

Firmado por 30f8b1bd-****-4715-9074-294881ddf1b2 el día 13/05/2024 con un certificado emitido por 30f8b1bd-b344-4715-9074-294881ddf1b2

Josep Demestre Viladevall

Teun van Dijk

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

Agradecimientos

A partir de muchas charlas informales con amigos, surgió en mí la intuición de que debía existir una especie de lógica conversacional que rigiera los procesos de cambio de opinión entre los interlocutores en situaciones comunicativas. Este trabajo nace de esa intuición.

Al principio, pensé que el estudio de los mecanismos cognitivos implicados en la persuasión debía abordarse desde la psicolingüística. Gradualmente, me incliné hacia los procesos de comprensión del discurso, atraído por sus posibilidades teóricas y metodológicas para abordar la persuasión. Sin embargo, los estudios del discurso me condujeron a la retórica y la argumentación. En ese punto, comprendí que primero debía entender en qué consisten las creencias y la credibilidad, lo que me llevó de la argumentación al razonamiento. Así, las creencias, la credibilidad y su dinámica han acabado constituyendo la temática central de esta tesis.

En este sentido, quiero agradecer a mis directores, el Dr. Josep Demestre de la URV y el Dr. Teun A. van Dijk de la UPF, por darme la oportunidad de desarrollar mi propia investigación, aunque finalmente se alejara tanto de sus respectivos campos de estudio. Gracias por su cercanía, orientación, dedicación, ayuda y generosidad intelectual y personal. En particular, quiero agradecer a Teun nuestras charlas sobre cognición, discurso y creencias, siempre planteando preguntas desafiantes y produciendo valiosos insights para mi investigación.

Agradezco sinceramente al Dr. Pere Joan Ferrando su ayuda con mis múltiples dudas metodológicas de diseño experimental y análisis; y con dudas teóricas sobre probabilidad. Gracias por tu buena predisposición y por tu mente preclara que ha servido en no pocas ocasiones para ordenar mis propias ideas. Agradezco también al Dr. José Eugenio García-Albea y, pese a su ausencia, a la Dra. Rosa Sánchez-Casas su cercanía y amabilidad, que me inspiraron para tomar la decisión de aventurarme en esta andadura.

I also thank Dr. Tobias Richter for giving me the opportunity to spend time with his group in Würzburg. His work has been a constant source of inspiration for mine. I am grateful for our conversations about discourse validation processes, as well as for his kindness and support during my stay there. And, of course, to Wienke, Catharina, and Frank for making my stay at the university very pleasant and fun, and for teaching me the local pragmatic norms of courtesy. You know. And outside the university, I'm very thankful to Laith, Andee, and above all, to Pascal.

También quiero mostrar mi agradecimiento a todos los integrantes del GIP de la URV, tanto en lo profesional como en lo personal. Agradezco a Pilar todo su apoyo en múltiples aspectos, de carácter docente, logístico, y por haberme ayudado a revisar algunos fragmentos de este trabajo. Mi agradecimiento es mayor aún en lo personal. Otras personas de este grupo que me han ayudado con la investigación por mera amistad son Marc y Juan. Gracias a los dos por haberse tomado el

esfuerzo y las horas para programar mis experimentos (también los que no incluyo aquí). De no ser por vosotros, no creo que los hubiera podido llegar a implementar.

Agradezco mucho también esos cafés junto a Marc, Roger, Gavi, Juan y Dani H. Un trabajo que te guste no está completo sin gente maja y divertida a tu alrededor. Agradezco nuestra amistad, nuestras conversaciones profesionales y personales, nuestros congresos y nuestras birras, belgas o normales. Ya ves... lo hemos pasado bien, sí señor. En estas relaciones intelecto-personales también quiero incluir a Lluís, Rubén, Fabia, Daniela, Mireia, Laia, Miriam, Marga, Josep Lluís, Camil, Teresa V., Ángel A., Harry, Yao y a mi antiguo amigo Potix ya ausente. Los agradecimientos no están completos sin incluir a otros miembros de la facultad que siempre me han ayudado con aspectos administrativos y logísticos como Joan, Esther, Toni, Ana, Pi y Mayca.

Gracias también a mi Silvia, mi madre, por un lado, y a Pedro, mi padre, por el otro, por todo el apoyo que me han dado durante todos estos años.

Y por supuesto, y por encima de todo, agradezco enormemente a mis hijos Iván y Víctor, y a mi mujer Sveta, a quienes este escrito ha robado momentos de intimidad, su incondicional apoyo y tantas cosas que no pueden llegar a ser expresadas con palabras. A vosotros quiero dedicar todo el esfuerzo que he invertido en este trabajo durante todo este tiempo.

Listado de abreviaturas: teoría, experimentos y elementos textuales en orden alfabético

Abreviaturas sobre conceptos teóricos

AC	Afirmación del consecuente
e	Evidencia.
ELM	Elaboration Likelihood Model.
FFA	Función de fuerza argumental.
h	Hipótesis.
LBS	Local Belief System
ME	Memoria episódica.
MP	Modus Ponens.
MS	Memoria semántica.
MT	Modus Tollens.
NA	Negación del antecedente.
Nc	Nodo de información candidato a entrar en el sistema de creencias del individuo (equivalente a RC).
Nj	Nodo de información justificado (i.e., creencia previa).
O – A – (que) P	Organismo – Actitud – Proposición: “Juan (O) cree (A) que María está en el salón (P)”
P (P S)	La probabilidad de que un predicado sea verdadero dado que un sujeto es verdadero (es decir, la probabilidad condicionada del predicado dado el sujeto).
P (SP SC)	La probabilidad de que una afirmación compuesta por un sujeto y un predicado sea verdadera, dada toda la información que uno considera verdadera, es decir, dado su sistema de creencias (es decir, la probabilidad condicionada de la verdad de una oración dada un sistema de creencias particular).
RC	Representación global candidata a entrar en el sistema de creencias del individuo (equivalente a Nc).
RG	Representación global (compuesta de representaciones particulares).
RGn	El subíndice puede ser sustituido por un número que sirva de identificador para una representación global en cuestión.
RTp	Relación tópica.
S – A – R	Sujeto – Adhesión – Representación.
S – A – (que) P	Equivalente a O – A – (que) P, donde S se refiere a sujeto.
SC	Sistema de creencias.
SLC	Sistema local de creencias (activo en la memoria de trabajo).
S-P	Sujeto-Predicado.
T	Teorías intuitivas implícitas que rigen un dominio de creencias o conocimientos.
TCM	Teoría Computacional de la Mente.

<i>Tp</i>	Tópico (equivalente a RTp).
TRM	Teoría Representacional de la Mente.

Abreviaturas sobre los experimentos

Experimento 1	
C1	Grupo control emparejado con G1.
C2	Grupo control emparejado con G2.
G1	Grupo experimental 1.
G2	Grupo experimental 2.
S	Strong.
W	Weak.
Experimento 2	
H	High.
L	Low.
HH	Warrant H and Data L.
HL	Warrant H and Data L.
LH	Warrant L and Data H.
LL	Warrant L and Data L.
RDI	Relative Disbelief Index.
RBI	Relative Believability Index.
L- Conditions	HL, LH & LL

Abreviaturas sobre elementos textuales

ap.	Apéndice.
App.	Appendix.
cap.	Capítulo.
chap.	Chapter.
pt.	Parte / Part.
sec.	Sección.
sect.	Section.

ÍNDICE

Abstract.....	1
Presentation	5
INTRODUCCIÓN.....	9
PARTE I: ANÁLISIS HISTÓRICO Y FILOSÓFICO DE CONCEPTOS FUNDAMENTALES	19
1. Argumentación	21
2. Concepciones de la Probabilidad	45
3. Filosofía de las creencias: ontología, clasificación y organización en sistemas de creencias.....	65
4. Fundamentos de la psicología cognitiva para el estudio de la credibilidad.....	83
PARTE II: HACIA UN MODELO TEÓRICO INTEGRAL DEL SISTEMA DE CREENCIAS Y LA DINÁMICA DE LA CREDIBILIDAD	117
1. Dinámicas de adaptación en sistemas de creencias: un modelo teórico de interacción, validación y cambio.....	119
2. Componentes del sistema de creencias.....	145
3. Inferencias explicativas como mecanismos de la justificación	179
4. Tipos de relaciones explicativas en el sistema de creencias	185
5. Deducción en el sistema de creencias I: fundamentos probabilísticos	209
6. Midiendo credibilidades como probabilidades.....	217
7. Deducción en el sistema de creencias II: validez y consistencia probabilísticas	227
PARTE III: TRABAJO EXPERIMENTAL.....	245
1. EXPERIMENT 1	247
1.1. Theoretical Foundations	247
1.2. The Present Study.....	254
2. EXPERIMENT 2	279
2.1. Theoretical Foundations	279
2.2. The present study	292
DISCUSSION	339
APÉNDICE A: EXPERIMENTO 1	369
APÉNDICE B: EXPERIMENTO 2	423
APÉNDICE C: FUNDAMENTOS TEÓRICOS - entimemas, tópicos y esquemas de argumentación.....	451
REFERENCIAS	511

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

Abstract

Beliefs are essential to our mental life, as they reflect how we interpret reality and how this interpretation influences our daily interactions and decisions. They are fundamental in shaping our expectations and how we perceive and react to the world around us. Despite their importance, cognitive psychology has not developed a unified framework that integrates beliefs, their credibility, and their dynamics in a model of belief system.

This thesis aims to offer a first approach to a new theoretical model that defines and structures the concept of belief, addressing how they are formed, adapted and modified. It proposes that the human cognitive system naturally seeks to create a coherent and unitary worldview, which is crucial for adapting and functioning effectively. It distinguishes between two types of beliefs: foundational, which are independent and form the basis of the belief system, and derived, which are built upon the beliefs already existing in the system. The belief system is described as a network where there are core beliefs that are robust and well-connected, and peripheral beliefs that are more flexible and susceptible to change. Key elements of this system include the subjectivity of the individual, the content of the belief and how this representation attaches to that subjectivity through justification relations and coherence connections, facilitating the transition from non-belief to belief.

This model considers the belief system as an argumentative map where justifications and refutations are essential for credibility and its dynamics. Validation of beliefs occurs through explanatory relationships which transform accepted information into premises and new information into conclusions. This model focuses primarily on inductive and deductive reasoning, where the former creates general knowledge from specific experiences and the latter uses this knowledge to integrate new information, ensuring that degrees of credibility from premises are transmitted to conclusions. Internal coherence is crucial for the validity of the belief system, but due to computational limitations, this coherence is mainly maintained in what we call the 'local belief system', active in working memory. This is where contradictions can be detected and corrected to maintain consistency, especially when integrating new information. According to the model, consistency is achieved by multiplying the credibilities of premises to align new information with existing information in terms of credibility. The transmission of credibility within the system depends on the connection between an established belief and new information, through a valid justification. If such a connection is missing or the credibility of any of the components is low, the transfer of credibility will be compromised. This is called the 'the weakest part epistemic rule'.

This model also introduces the 'standard of argumentation', which states that the credibility of a new representation can only increase if the premises from which it is derived have a higher

credibility. Arguments that follow this principle are known as 'enthymemes', and their effectiveness depends on the combination of credibilities of the premises exceeding the initial credibility of the conclusion.

This thesis validates the proposed belief dynamics by means of two experiments focused on the deductive aspect of the model, which conceives the belief system as an argumentative system. It investigates how credibility is transmitted between justified and initially non-credible representations, using terms and concepts from argumentation theory. 'Data premises' representing facts and 'warrant premises' as relational rules structure the arguments, described as enthymemes when the credibility of the premises exceeds the initial credibility of the conclusion. This approach allows the effectiveness of the model to be measured and confirmed, using argument structures to assess how credibility is conveyed within the belief system.

The first experiment used a test-posttest design with a control group to assess the effectiveness of enthymemes in increasing the credibility of initially disbelieved statements. Claims were evaluated in the absence as well as in the presence of premises. In addition, the arguments could contain a warrant premise believed or disbelieved by the participants. Thus, it was examined whether arguments with credible warrants influenced the transmission of credibility from the data premise, which was always believed, to the conclusion. The results showed that arguments with believed warrants were significantly more effective in increasing the credibility of critical statements compared to the other conditions, thus validating the effectiveness of the model and the enthymemes in the process of argumentation and belief change.

The second experiment investigated how the credibilities of premises are combined, testing the multiplication hypothesis, which is central to validating the dynamics of beliefs proposed by the model in terms of probabilistic validity and consistency.

In the experiment, the credibility of the warrants (high vs. low) and data premises (high vs. low) was manipulated, establishing four distinct credibility conditions for the arguments. This allowed for an evaluation of how different levels of premise credibility affected the acceptance of conclusions, whose linguistic content was kept identical across all argument credibility conditions. Participants first assessed the credibility of each premise individually and then the credibility of the critical claims, considered as conclusions of the formulated arguments. A control group evaluated these same claims in isolation to establish a baseline of a priori credibility.

The results showed that the credibility of the conclusions varied systematically according to the credibility of the premises. In arguments that formed enthymemes, the credibility of the conclusion was higher than in other conditions and the a priori credibility. The results also demonstrated a close relationship between the credibility predicted by the multiplication hypothesis and those observed by the participants across all items. The weakest part epistemic rule also received support, acting as the upper limit of the credibility of the conclusions of the enthymemes, and exerting a

strong influence on the credibility of the conclusions in other conditions. In these other conditions, the a priori credibility acted as the upper limit and in all conditions, multiplication was the lower limit of the credibility adopted for the conclusions of the arguments.

The validation of these aspects of the model emphasizes the importance of a coherent argumentative structure in the formation and modification of beliefs within the belief system.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

Presentation

The present doctoral thesis aims to provide a significant contribution to the study of beliefs from a cognitive psychology perspective. The thesis is structured in three main parts, in order to address in a relatively exhaustive and detailed way the different aspects of the research carried out.

The three general parts of the thesis are as follows:

- **Part I:** Fundamental Concepts. This section provides a broad theoretical background covering several disciplines relevant to the study of beliefs. In particular, argumentation (logic, dialectic and rhetoric), conceptions of probability (epistemic and aleatory; qualitative and quantitative), philosophy of belief and justification networks (foundationalism and coherentism), and the psychology of reasoning (formal and everyday).
- **Part II:** Development of a theoretical model on beliefs, credibility and the belief system. One of the main contributions of the research is presented here, proposing a theoretical model that attempts to address the above concepts and the main internal mechanisms of the system.
- **Part III:** Presentation of two experiments that test the key mechanisms of the proposed theoretical model. One of them is more focused on examining structural aspects of the transmission process, while the other is centred on the credibility computation mechanisms derived from the theoretical model proposed in Part II.

I am not counting in these three parts the first introductory chapter of the thesis, nor the general discussion and conclusions of the thesis.

In order to obtain an international doctoral degree, the thesis has been written bilingually. The sections written in English include the abstract, this presentation, as well as Part III, and the general discussion and conclusions. The abstract and the general discussion and conclusions are written in English in accordance with the minimum requirements of the regulations for obtaining the international mention. The fact that this presentation and Part III are in English is due to the possibility that some members of the panel may not understand Spanish. In this way, it is easier for all members of the examining board to have access to significant parts of the thesis. On the other hand, the drafting of Part III in English is a first step towards its subsequent adaptation for possible publication.

The Spanish parts cover the introductory chapter of the thesis, as well as Parts I and II in their entirety. The choice to write these sections in Spanish is based on several reasons. Spanish allows me to express myself with greater precision and fluency, especially when dealing with theoretical and complex concepts. The writing of the introduction and, above all, Parts I and II demand

precision, as they deal with the theoretical framework and conceptual development of the thesis. Writing these parts in my native language allows ideas and concepts to be conveyed more precisely. Thus, since the thesis regulations allow it, I have made this decision for a more solid and reliable presentation of the theoretical content.

In any case, in the introduction to each of the experiments in Part III I have explained complementary parts of the model presented in Part II. In this way, if any member of the examining board did not have access to the contents in Spanish, he or she could get an idea of the theoretical model, which is essential to understand well where the hypotheses of the experiments come from. It is true that this lengthens the length of the thesis by repeating ideas that have already appeared in Spanish in Part II, but it ensures that all members of the examining board can properly understand the content and development of the research.

As the research covers several disciplines, Part I is relatively extensive. This is due to the need to establish a comprehensive theoretical framework that facilitates the understanding of the issues addressed in the later parts of the thesis. In this sense, despite the considerable size of the thesis, it is important to note that the key original contributions are to be found in Parts II, III, and the general discussion, all comprised between pages 119 and 367. The detailed, interdisciplinary theoretical introduction, as well as the extensive appendices, contribute significantly to the total number of pages, but should not be considered a central part of the original contribution.

Thus, since it is not my intention to hijack the reader's attention for so many pages, I have thought it appropriate to offer some itineraries to guide the reader to facilitate an agile and focused reading through the thesis.

Recommended Itineraries

My own contributions, which form the central body of the thesis, are Parts II and III. Part I has several objectives, which are detailed in the introduction to the thesis. In order to cut to the chase and not to overwhelm the reader, the most relevant sections of each chapter of Part I are detailed below:

- **Chapter 1:** The most relevant sections are 1, 3.1, 4.2.1, 4.2.2 and section 5. Throughout these sections the origin of the discussion between the certain, demonstrative and conclusive and the uncertain and probable is established. The understanding of logic, dialectic and rhetoric in the classical world and its current resurgence is also presented, as well as Toulmin's model, which will serve as inspiration for several aspects of the thesis. Finally, some key concepts of classical rhetoric are presented, such as topical relations, argumentation schemes, *éndoxa* and the concept of *enthymeme*.
- **Chapter 2:** The relevant sections on conceptions of probability are 2, 3 and 6 (pre-theoretical probability), and 9 and 11.2 (modern probability). Throughout these sections, the aim is to show the natural relationship of probability to argumentation and credibility, already in the ancient world, as well as to observe its successive transformations up to the present day. In this way, the relations of aleatory probability to epistemic probability are

described, the relations between deductive and inductive probability are explained, and today's conception of subjective probability is presented.

- **Chapter 3:** The most relevant section for the thesis is section 3. This section presents the two main philosophical proposals concerning the transmission of justification in belief systems: foundationalism and coherentism. Their main differences and similarities are presented. In fact, they agree on several of the mechanisms on the transmission of credibility. The model presented in this thesis is better aligned with certain postulates of foundationalism and, in the same way, it admits some of the mechanisms defended by coherentism.
- **Chapter 4:** This chapter continues the debate between the formal (certain, logical) and the informal (fallible, probable). The sections relevant to the theoretical model developed are 8, 13, 14.1, 14.3, 14.4 and 15.2. Sections 15.4 to 15.6 deal with the role of memory in accessing rebuttals to arguments and the time course of their retrieval. To some extent, these issues underlie the theoretical model presented. However, they are not directly related to it because, in this thesis, I am not concerned with online processing of credibility. Therefore, these sections can be skipped. Sections 16 and 17, on the other hand, are relevant since they establish the bridge between the studies on rebuttal retrieval and probabilistic reasoning, and lead to the development of the theoretical model.

These sections are the main introduction to the following parts. Throughout Parts II and III, I recurrently mention the relevant sections as concepts introduced earlier appear. Thus, one can go through the sections just mentioned before entering Parts II and III. Or, one can go directly to those Parts and move backwards as the relevant sections are mentioned. Or, one may choose to read more deeply into the aspects of argumentation, probability, philosophy or psychology, depending on the reader's particular background and considerations.

On the other hand, as already mentioned, Parts II and III form the central body of the thesis and represent the original contributions of this work. The theoretical model presented in Part II outlines the adaptive and communicative functions of the belief system, its basic constituents, and the sources of belief justification. Furthermore, it discusses its argumentative and explanatory nature, as well as the inductive and deductive mechanisms in belief formation and the transmission of credibility. The role of induction in the formation of schemas understood as rules and general beliefs is explained, and some considerations are made about the adequacy of measuring credibility as a subjective probability. Finally, it focuses on the deductive mechanisms of the system in the transfer of credibilities, which will be the central theme of the experiments presented in Part III of the thesis.

Experiment 1 is a first empirical approximation to the model. The design resulted in data analysis equivalent to two identical experiments, with the results being replicated in both. In this sense, the second analysis and its discussion are very similar to the first one. Experiment 2 addresses several aspects related to the validity of the theoretical model. The main objective is to validate the

hypothesis of multiplicative computation in the transmission of credibilities. Several consequences follow from this hypothesis, the relevance of which may depend on the reader's objectives and vary in their degree of specialization. There will be hypotheses, with their results and discussions, which may be relevant only to those interested in argumentation (e.g., standard argumentation). Other issues may appeal only to those concerned with specific aspects of probabilistic and plausibilistic argumentation and reasoning, such as the discussion of the Theophrastus rule (weakest link rule or the MIN rule). Additionally, other issues may be relevant only to those concerned with probabilistic validity and consistency and their relation to the limits of the a posteriori credibility of a given assertion in terms of the premises of the argument. All these questions are very specific and detailed and may respond to more particular interests.

After the general discussion, from page 369 onwards, there are appendices with the materials used in the experiments, some additional analyses and an extensive and detailed appendix on the concepts of enthymemes, topical relations, discourse coherence and argumentation schemes. We consider that the presentation of all these concepts in the main body of the thesis (especially in the corresponding sections of Part I) is sufficient to follow the development of Parts II and III of the thesis. This appendix has been included only in case the reader feels that an assumption requires further justification or explanation, or if he or she simply wishes to know more about the topic. Thus, it is not considered a really necessary text, which is why it is included as an appendix. Finally, from page 511 to the end, there are the bibliographical references.

INTRODUCCIÓN

La mayoría de las veces que uno trata de dar una explicación acerca de su propia conducta o de la de otras personas suele hacerlo en términos de ciertos estados mentales como sus deseos, sus intenciones o sus creencias. Por ejemplo, al tratar de explicar algo como “por qué Juan fue de Tarragona a Barcelona en tren y no en bus o en coche” podríamos decir que lo hizo “porque Juan creyó que ir en tren era más rápido que ir en bus y más barato que ir en coche y que quería optimizar dinero y tiempo”, o algo por el estilo. Y este mismo ejemplo podría aplicarse también en primera persona.

Los deseos, las intenciones y las creencias juegan un papel fundamental en la vida mental humana, y muchos modelos y teorías los toman como predictores relativamente fiables de las acciones de los individuos. Compramos un producto, votamos a un partido, emitimos un veredicto o un diagnóstico, discutimos y defendemos nuestras ideas sobre la base de estos parámetros; base a la que, sin duda, se añadirían otros ingredientes más. Veremos también que las creencias, además de predecir conductas, pueden funcionar como buenos predictores de otras creencias.

El concepto de creencia ha sido ampliamente investigado en epistemología y en filosofía de la mente; y, desde una perspectiva computacional, también en Inteligencia Artificial. Sin embargo, en psicología ha recibido una atención comparativamente menor, pese al uso generalizado que se hace del término. Es posible que se haya considerado que el estudio acerca de qué es una creencia sea una cuestión metafísica u ontológica o, por lo menos, una cuestión más propia del análisis conceptual filosófico que una tarea del psicólogo. Pero, a poco que se sepa de las diferentes áreas que aborda la psicología, se puede ver que las creencias están implicadas aquí y allá. Son un componente del autoconcepto, de las terapias de reestructuración cognitiva, del proceso de disonancia cognitiva, del proceso motivacional, de las actitudes, de los estereotipos sociales, de la persuasión, están implicadas en la comunicación y el discurso, en la toma de decisiones, en el razonamiento e incluso en la percepción; y esto no pasa de ser un listado breve. Así pues, aun asumiendo que pudiera ser lícito evitar investigar qué es una creencia, es relevante tratar de dar con algunas de sus propiedades básicas para poder entender mejor cómo funcionan, dada su

ubicuidad en la cognición. Si bien tal vez podría argüirse que las creencias como tales no son mensurables, sí que podría serlo una propiedad básica inherente a toda creencia: su credibilidad. Y, desde luego, medir constructos mentales y manipular sus magnitudes para saber cómo funcionan sí que es una tarea adecuada para la psicología.

Por otro lado, las credibilidades son variables, cambiantes. Lo que uno creyera ayer, puede que haya dejado de creerlo hoy, y lo de hoy, mañana. Y del mismo modo, uno puede pasar de no creer algo en un momento dado a creerlo posteriormente. Además, al menos intuitivamente, la credibilidad parece darse en grados. Uno puede estar totalmente convencido de la verdad o de la falsedad de una proposición, de la corrección o incorrección de una idea, de la adecuación o inadecuación de la relación entre una representación subjetiva y el mundo real. Pero también puede dudar a su respecto, puede estar más o menos seguro, puede creerlo hasta cierto punto. La credibilidad se mueve entre la certeza y la incertidumbre.

1. Objetivo general de la investigación

El objetivo más general del presente trabajo es avanzar en la comprensión de los mecanismos básicos implicados en los cambios de credibilidad, tratando de averiguar qué tipos de cálculos de credibilidad realiza el sistema cognitivo, así como examinar los procesos mentales que permiten realizar dichos cálculos.

Para poder hacerlo, habrá que construir un marco teórico en el que situar el estudio de la credibilidad, habrá que determinar de qué elementos depende su dinámica y habrá que hallar un método empírico que permita manipularla y medirla.

2. La argumentación retórica como herramienta para el estudio de la credibilidad

Tradicionalmente, se ha considerado que la credibilidad que otorga un individuo a un contenido dado p varía en función de tres elementos: de la experiencia personal directa del individuo con ese contenido (he percibido p), de procesos comunicativos con otros individuos acerca de ese contenido (me han dicho que p) y de procesos individuales de razonamiento e inferencia que relacionan el contenido en cuestión con otros contenidos previos (a partir de saber que r y s , infero que p). Y en todos esos casos, uno podrá sostener o no que p con un grado de seguridad e incertidumbre variable. En la presente investigación cobrará especial interés el último punto referido al razonamiento sobre contenidos previos para determinar la credibilidad de un contenido dado. En este sentido, conocer los procesos del tipo de razonamiento implicado en los cambios de credibilidad será una tarea central en esta investigación.

En cualquier caso, será necesario partir de un cuerpo teórico y metodológico especializado en el estudio del cambio racional de credibilidades. En este sentido, los conocimientos provenientes de la argumentación retórica servirán de base para detectar las variables de interés en este trabajo. Y detectando las variables de las que depende el cambio racional de credibilidades será posible aislarlas y manipularlas a fin de examinar sus efectos. La argumentación retórica es un proceso comunicativo. Debido a esto, también el penúltimo punto mencionado en el párrafo anterior acerca de los procesos comunicativos y discursivos será de gran relevancia para el desarrollo teórico. Así, muy brevemente y yendo a lo que nos interesa para este trabajo, la argumentación retórica prescribe que el hablante deberá ofrecer razones al oyente que *justifiquen* su cambio de credibilidad.

3. El entimema como argumento retórico de la presente investigación

Decir, simplemente, que me serviré de los conocimientos procedentes de la argumentación retórica es aún demasiado general. Hay que concretar un poco más. En este sentido, la clase de argumentos que se empleará como herramienta básica para lograr los cambios de credibilidad a lo largo de los experimentos de esta tesis son los llamados *entimemas*, también conocidos como *silogismos retóricos*. No es este aún el momento de hacer una explicación detallada del concepto de entimema. Pero es pertinente describirlo someramente (con la advertencia de que todo lo que diga será matizado en el cuerpo de la tesis) y poner un ejemplo.

Un entimema es un silogismo. Y un silogismo es un tipo de razonamiento lógico. *Grosso modo*, tradicionalmente se ha sostenido que los razonamientos lógicos pueden ser deductivos o inductivos. En la retórica clásica se ha considerado que el razonamiento retórico inductivo se corresponde con la ejemplificación, mientras que el razonamiento retórico deductivo se corresponde con el entimema. Un entimema es un silogismo retórico. Y la retórica tiene primordialmente (aunque no exclusivamente) la función comunicativa de lograr la persuasión acerca de algún asunto por parte de una audiencia y, tradicionalmente, lo hace por medio del discurso. Así, al establecer una tesis, trata de lograr un aumento de credibilidad en el contenido de esa tesis, mientras que, al tratar de refutarla, perseguirá conseguir una disminución de credibilidad por parte de la audiencia acerca de ese contenido. En esta investigación sólo emplearé entimemas que pretendan lograr un aumento de credibilidad.

Si bien un principio básico para la comunicación de información nueva consiste en tomar la información ya conocida por parte de los oyentes como punto de partida del discurso, en los entimemas se tomará la información ya aceptada y creída por los oyentes como premisas o puntos de partida del argumento. Como ya se indicó más arriba, la credibilidad es gradual y se mueve entre la certeza y la incertidumbre. En este sentido, las premisas de los silogismos retóricos no se corresponderán necesariamente con verdades objetivas, sino con las creencias sostenidas por la audiencia; creencias sostenidas con una fuerza variable. Es decir, que, en el silogismo retórico, el

contenido de las premisas no es necesariamente verdadero, sino meramente probable. Y esa probabilidad ni siquiera depende del propio contenido aducido en la premisa, sino de la que le otorgue la audiencia.

Imaginemos que, en el curso de una conversación cotidiana, alguien nos dice lo siguiente: “... los avances en medicina, al final, acabarán provocando un aumento del número de personas con enfermedades”. Supongo que la mayoría de nosotros, por lo menos, de los miembros de la sociedad a la que pertenezco, mostraríamos nuestro disentimiento acerca de esta idea. Entonces, nuestro interlocutor podría exponer el siguiente razonamiento: “... parece bastante claro que los avances en medicina están produciendo un aumento del número de ancianos. Y, en general, los ancianos son un colectivo de personas que suelen tener una u otra enfermedad”. Estas dos últimas frases funcionarían a modo de premisas de la conclusión “...Por tanto, los avances en medicina acabarán causando un aumento del número de personas con enfermedades”.

Este argumento constituiría un entimema o silogismo retórico sólo si aceptamos en cierto grado lo que se está diciendo en las premisas, es decir, sólo si el contenido de las premisas encaja con lo que creemos. Imaginemos que encaja bastante con lo que cree la mayoría de los miembros de una comunidad *x*, pero no así con lo que cree una comunidad *y*. En ese caso, el argumento expuesto sería un entimema en *x*, pero no lo sería en *y*. Y el hablante habrá sido un buen comunicador retórico en *x* por haber detectado las opiniones generales de los oyentes, y no lo habrá sido en *y*, por lo contrario. Los entimemas parten de lo que los oyentes consideran probable, derivando en conclusiones a las cuales les otorgarán también cierta probabilidad mediante el razonamiento. Y da igual si, por ejemplo, la proporción objetiva de ancianos con alguna enfermedad es del 90%, del 50% o del 1%; no importan tanto aquí las supuestas probabilidades objetivas. Los entimemas operan con probabilidades subjetivas. Por lo tanto, en esta tesis voy a sostener que las credibilidades se miden como probabilidades subjetivas y que se comportan como tales.

Esta es, pues, la acepción original principal de entimema. A pesar de ello, no es la más extendida.

Como mencionaremos en más de algún lugar, el concepto de entimema que ha llegado hasta nuestros días es el de un silogismo al que le falta una premisa. Retomando el ejemplo anterior, podría ser tanto:

1. Los avances en medicina están produciendo un aumento del número de ancianos. Por tanto, los avances en medicina acabarán produciendo un aumento del número de personas con enfermedades.

Como:

2. Si aumenta el número de ancianos, entonces aumentará el número de personas con enfermedades. Por tanto, los avances en medicina acabarán produciendo un aumento del número de personas con enfermedades.

Pese a que debo, al menos, presentar estos *entimemas parciales*, no nos ocuparemos de ellos en esta tesis. En todo caso, cabe apuntar que, en las conversaciones cotidianas se emplean más los entimemas parciales con premisas implícitas que los completos, que exponen todas las premisas del silogismo. Sostengo que esto es así por una cuestión de economía cognitiva y agilidad discursiva, pero sin efectos relevantes en lo que a la credibilidad se refiere. En todo caso, algunos autores han apuntado la posibilidad de que el hecho de dejar una premisa implícita aumenta los efectos persuasivos de los entimemas. Otros se han interesado más por la cuestión acerca de cómo se reconstruyen estos entimemas en la mente de los oyentes. Todas estas propuestas son interesantes, pero no tendrán cabida en la presente investigación.

4. Controversias teóricas de cada uno de los conceptos relevantes

Ninguno de los supuestos que adoptamos en este trabajo está fuera de discusión. Por ejemplo, cuando se habla de probabilidad se piensa en cálculos matemáticos, en frecuencias relativas y en el azar, y no tanto en cuestiones epistemológicas como el conocimiento, las creencias o las credibilidades. Cuando se habla de retórica, se piensa más en características ornamentales y estilísticas de un texto discursivo hablado o escrito, o de un mensaje visual, así como en la generación de inclinaciones emotivas en la audiencia. Pero no en estructuras con un componente lógico sobre las que discurren ciertas formas de razonamiento que computan credibilidades. Cuando se habla de lógica, y en particular de lógica deductiva, se piensa en un tipo de razonamiento concluyente y de corte racionalista. Este es definido con precisión matemática y es abstracto, exclusivamente formal. Se corresponde con la lógica clásica bivalente, en la que las premisas de los argumentos son verdaderas o falsas en función de que sus símbolos constituyan fórmulas bien formadas de ese lenguaje formal. Además, las reglas de combinación de símbolos y fórmulas son insensibles a los contenidos materiales o sustanciales del argumento. Sin embargo, la lógica de los entimemas hace deducciones sobre representaciones mentales con contenidos cotidianos, donde las premisas conllevan cierta carga de incertidumbre y permiten excepciones, derivando en un razonamiento no concluyente. Además, la aceptabilidad de las conclusiones es aquí muy sensible a los contenidos del argumento. Es, en fin, un razonamiento menos matemático (en su sentido racionalista) y más cotidiano y discursivo. Finalmente, cuando se habla de entimemas, si se piensa en algo, las más de las veces será en un silogismo cualquiera al que le falta una premisa, y no tanto en un argumento retórico cuya finalidad es la de persuadir o convencer.

Así pues, se abordará la discusión de estos temas a lo largo de la revisión teórica de la primera parte de esta tesis, defendiendo que la retórica alberga primariamente un tipo de razonamiento, que los entimemas computan credibilidades operando mediante un tipo de deducción, y que las credibilidades pueden ser medidas y funcionan como probabilidades subjetivas.

5. Objetivos específicos de la investigación

Una vez dicho todo esto, presento una serie de objetivos más específicos. El primero de ellos consiste en tratar de esbozar un marco teórico para el estudio de la credibilidad desde la perspectiva de la psicología cognitiva. Por otro lado, dado que no existe ningún otro trabajo experimental que haya puesto a prueba la capacidad de los entimemas para cambiar credibilidades, otro de los objetivos será traerlos de la teoría al laboratorio para comprobar si funcionan para tales propósitos. Si este objetivo se satisface, entonces podremos avanzar al siguiente. Este consiste ya en investigar de forma sistemática los cómputos de credibilidad que se desprenden del modelo teórico esbozado. Esto implica estudiar los procesos de razonamiento desencadenados por la argumentación con entimemas, examinando la posibilidad de que existan relaciones regulares de credibilidad entre las premisas y la conclusión.

6. Estructura de la tesis

Estos objetivos se irán abordando a lo largo de los siguientes capítulos de esta tesis, que se dividen en tres partes:

La *primera parte* de la tesis comprende un análisis histórico, filosófico y psicológico de los conceptos fundamentales que servirán como base para el modelo teórico propuesto. En el primer capítulo, se explorarán conceptos propios de la *argumentación* como el silogismo, el entimema, el éndoxa, la relación tópica y el Modelo de Toulmin. Este capítulo será relativamente amplio, ya que algunos conceptos podrían ser menos familiares para los lectores procedentes de la psicología. El segundo capítulo se centrará en el análisis histórico y filosófico del concepto de *probabilidad* y sus interpretaciones. En el tercer capítulo, se examinarán las principales propuestas filosóficas sobre el concepto de *creencia*, los tipos de creencias y los *sistemas de justificación*. El cuarto capítulo explorará las diferentes propuestas en *psicología del razonamiento* y su relación con la credibilidad. Estos capítulos están conectados por un análisis histórico del debate entre lo verdadero y lo verosímil, lo estático y lo dinámico, lo abstracto y lo concreto, así como las formas y los contenidos.

La *segunda parte* de la tesis aborda uno de los objetivos centrales de este estudio: presentar un modelo teórico sobre el sistema de creencias. Se busca definir qué es una creencia, sus componentes y la naturaleza de la credibilidad, así como explorar las relaciones en los sistemas de creencias, con énfasis en los procesos inductivos y deductivos. Se examina la medición del grado de creencia mediante probabilidades subjetivas y se profundiza en el papel de la deducción en el sistema de creencias, detallando sus características principales, operaciones en la combinación de la credibilidad y sus consecuencias. De estas últimas consideraciones surgirán las principales hipótesis que serán puestas a prueba en los experimentos subsiguientes.

La *tercera parte* de esta tesis se enfoca en el segundo gran objetivo: analizar empíricamente aspectos clave del modelo de sistema de creencias propuesto anteriormente y evaluar la

aplicabilidad de algunas prescripciones de la retórica y la teoría de la argumentación. Se llevarán a cabo dos experimentos. En el primero, se evaluará si los entimemas aumentan la credibilidad que se unas proposiciones críticas y se investigará la función de los componentes del argumento en este proceso. El segundo experimento profundizará en las hipótesis y predicciones derivadas del modelo teórico sobre el cómputo de credibilidades y sus consecuencias. Se discutirán los resultados a la luz del modelo teórico propuesto. Además, se podrán en relación con la teoría de la argumentación y la psicología del razonamiento. Específicamente, se explorará cómo estos resultados se vinculan con los intervalos de coherencia en la deducción probabilística y las variantes de la Regla de Teofrasto en el razonamiento plausibilístico y la psicología de la argumentación.

La conclusión de la tesis resume el modelo teórico, los experimentos y los resultados más importantes. Se realiza una *discusión* crítica que abarca las implicaciones teóricas, empíricas y aplicadas, especialmente en relación con la persuasión. Se identifican y se analizan las principales limitaciones tanto teóricas como metodológicas, y se proponen posibles direcciones para futuras investigaciones.

Los *apéndices* no solo incluyen los materiales del experimento y las instrucciones (también traducidas al inglés), sino también análisis teóricos y resultados secundarios. Un apéndice extenso analiza los conceptos de entimema, esquemas de argumentación y su relación con los esquemas cognitivos, así como las relaciones tópicas y su conexión con las relaciones de coherencia propuestas en psicología del discurso. Aunque se hacen referencias internas a este apéndice a lo largo de la tesis, no es esencial para entenderla. Su propósito es respaldar las suposiciones hechas en la tesis y proporcionar información adicional sobre estos conceptos para el lector interesado.

7. Contribuciones generales de esta tesis al estudio científico de la credibilidad

Desde los inicios de los trabajos de Ulrike Hahn y Mike Oaksford hace dos décadas sobre razonamiento bayesiano y argumentación, el interés de los psicólogos del razonamiento por la argumentación ha ido en aumento. Esta popularidad también se ha extendido a los estudios psicológicos del discurso. Sin embargo, pocas investigaciones han explorado en profundidad el concepto de credibilidad que subyace a estos procesos.

Esta tesis se adentra en el estudio de la credibilidad desde una perspectiva de la psicología cognitiva. Comienza con una fundamentación teórica y filosófica que abarca algunos aspectos de la lógica clásica y la argumentación retórica, pasando por el razonamiento cotidiano hasta la probabilidad subjetiva y los sistemas de justificación, evitando un abordaje demasiado técnico.

Una de las aportaciones principales de esta tesis reside en el desarrollo de un modelo teórico sobre la credibilidad. Para ello, se ha intentado articular conocimientos de diversas disciplinas como la argumentación, la filosofía, la psicología del discurso y la psicología del razonamiento. El modelo aspira a ser exhaustivo al abordar los conceptos, la génesis, la estructura y la dinámica de las

creencias y la credibilidad. Además, pretende ser un marco teórico preliminar para futuras investigaciones empíricas sobre las creencias. La estructura del modelo permite distinguir entre diferentes componentes y procesos en el sistema de creencias, proporcionando un punto de partida para organizar el plan de trabajo en este campo y facilitar la relación interdisciplinaria.

La segunda aportación principal se centra en el trabajo experimental que pone a prueba los pilares centrales del sistema en su aspecto deductivo, específicamente evaluando cómo se combinan las credibilidades durante la transmisión. Esta evaluación es importante, ya que la firmeza de estos pilares depende de dicha combinación. En este sentido, la función del trabajo empírico es, por un lado, mostrar la operatividad del modelo para derivar en predicciones precisas. Por el otro, proporcionar un respaldo empírico firme al modelo teórico, un requisito indispensable para considerarlo como una línea de investigación viable en el campo.

En la realización de este abordaje empírico, nos enfocamos en la estructura argumentativa del entimema como un medio clave para modificar credibilidades. Los entimemas fueron ya definidos por Aristóteles como "el cuerpo de la persuasión". Este estudio los recupera como herramienta para entender la credibilidad desde una perspectiva cognitiva, explorando su potencial para la modificación efectiva de credibilidades mediante deducciones sobre probabilidades. A través de esta metodología, se ha desarrollado una fórmula matemática simple que predice con alta precisión la credibilidad de las conclusiones argumentales, facilitando así la comprensión de cómo las premisas influyen en la aceptación de conclusiones. A partir de esto ha sido posible generar también otros indicadores acerca de la fuerza argumental, y de las condiciones de credibilidad que deben reunir las premisas para producir un aumento de credibilidad de la proposición que pretenden justificar.

Finalmente, los materiales utilizados en esta investigación son innovadores, ya que se basan en las creencias reales de los participantes sobre situaciones cotidianas, proporcionando insights auténticos sobre la credibilidad real en contextos no ficticios. Este enfoque permite obtener una medida más precisa y significativa de la credibilidad y su impacto en el razonamiento diario.

8. Limitaciones generales de la presente investigación

Evidentemente, las limitaciones de este trabajo son también muchas. En la vida cotidiana, nuestras creencias pueden influir sobre nuestra conducta de muchos modos, dependiendo de muchas más variables de las que pueden ser manipuladas en un laboratorio en un momento dado.

En primer lugar, muchas creencias se forman a partir de la experiencia directa con el objeto de la creencia. La frecuencia de exposición a p , y la frecuencia relativa de las observaciones personales en las que p es q , sin duda, afectan a las creencias que uno se vaya a formar acerca de p . Este es un estadio de la generación de credibilidades que no será cubierto en la parte empírica de esta tesis; en todo caso, este tema será abordado en el desarrollo del modelo teórico, además de que existe

una bibliografía bastante extensa de razonamiento inductivo que de un modo más o menos directo aborda estas cuestiones. Pero estas creencias acerca de p formadas por inducción, en la vida cotidiana, van a estar inherentemente condicionadas por el contexto físico, social y cultural de los individuos que adquieren la creencia. Resulta evidente que muchas de mis creencias sobre, por ejemplo, los pájaros, serían distintas si yo hubiese nacido y crecido en el Amazonas en lugar de en Groenlandia; y así con muchas otras cosas. En este caso, la bibliografía ya no es tan extensa.

Por otro lado, la comunicación en general, y los medios de comunicación periodísticos en particular influyen en gran manera sobre las opiniones que sostengan las personas. Eso se produce a través de discursos de una extensión mucho mayor de la que constan nuestros entimemas, que no dejan de ser argumentos aislados formados por tres frases cada uno, redactadas, en ocasiones, con un estilo forzado y artificial. En entornos más naturales, esos discursos están, a su vez, enlazados con el contexto social y cultural, e incluso, ideológico de los individuos. La interpretación de los argumentos depende del propio contexto en el que se procesan: del dónde, del cuándo, de quién lo dice y hacia quién va dirigido el discurso en orden a afectar a sus creencias. Pero, por otro lado, sobre todo va a depender de la representación mental particular de ese contexto, con todos sus parámetros, que se formen los oyentes o lectores.

Nada de eso va a ser abordado aquí. Como se ha dicho más arriba, los materiales experimentales son silogismos contruidos sobre las creencias reales de los participantes acerca de cuestiones de su entorno cotidiano a fin de que resulten lo más realistas, naturales y ecológicos posible para nuestros fines. En este sentido, se ha tenido en cuenta a la audiencia para la elaboración de los argumentos en función de lo que ellos creen o no. Este ha sido nuestro modo de dar cabida al contexto de los participantes. Pero, desde luego, los presentes materiales experimentales no pueden ser un sustituto de los discursos habituales del día a día. El necesario control de las variables experimentales para los presentes objetivos lo complica. No se han manipulado variables del contexto de procesamiento para averiguar sus efectos; el contexto experimental se mantuvo constante. Ni se han tomado medidas acerca de las representaciones de dicho contexto por parte de los individuos. No se ha manipulado ninguna variable relativa a la fuente de los argumentos, como podría ser la credibilidad de la fuente comunicativa, variable que muy probablemente debe intervenir siempre en los procesos cotidianos de modificación de credibilidades. Ni se han tenido en cuenta las diferencias individuales de los participantes en variables como, por ejemplo, la emotividad. Esta tesis se limita ofrecer unos primeros pasos en el desarrollo de un modelo teórico cognitivo sobre las creencias, y en el estudio de la modificación racional de credibilidades, dejando fuera variables de tipo emocional y actitudinal que son indispensables para obtener un panorama más completo y natural sobre el funcionamiento de la credibilidad.

Bajando a un nivel más técnico, también es relevante advertir que la investigación de esta tesis se centra en un punto muy acotado: el cambio de credibilidades con entimemas de la primera figura sin cuantificadores con el esquema de argumentación denominado “de la causa al efecto”. Aún es

pronto para entender esta retahíla de palabras, puesto que aún no he presentado varios de estos conceptos. Pero lo que pretendo señalar con ello es que no se puede decir que esta investigación sea acerca del cambio de credibilidades sin más, de forma general. Es sólo acerca de un modo muy local de cambio de credibilidades. El cambio de credibilidad puede suceder por vías no racionales. Cuando son racionales, puede ser por vías no deductivas. Cuando son deductivas, podría ser que se llevasen a cabo mediante otra estructura que no fueran entimemas. Y dentro de los entimemas, podría tenerse en cuenta el efecto de los cuantificadores. O podrían compararse distintas estructuras formales (figuras) que los entimemas pueden tomar. Y, por último, hay muchos más esquemas de argumentación que el escogido para esta tesis. Por otro lado, como se ha apuntado más arriba, los materiales empleados en la parte experimental de la tesis son argumentos aislados, que poco tienen que ver el uno con el otro, pasando de un tema a otro en cada ítem. Son argumentos compuestos por tres frases que los participantes, situados en una cabina experimental ante una pantalla, van leyendo y evaluando. Desde luego, no se pretende que la lectura de un silogismo puntual acerca de un tema particular vaya a tener un impacto que reestructure el sistema de creencias del individuo. Aunque, en cualquier caso, esto es algo que tampoco ha sido comprobado en esta tesis. Aquí no se ha llevado a cabo ningún experimento en el que se haya vuelto a llamar a los participantes unos días después de haber hecho el experimento para ver si perduran las nuevas credibilidades o si han vuelto a los niveles originales.

Por último, pese a la importancia que tiene la fuerza con la que uno está adherido a una creencia para la predicción de la acción, esto es algo que tampoco ha sido introducido en la presente investigación. Si bien es cierto que en las primeras formulaciones de los grados de creencia como probabilidades subjetivas se hacía en términos de conductas de apuestas, y que las probabilidades subjetivas han estado estrechamente vinculadas a la teoría de la utilidad, a las teorías de preferencias y de la acción racional, esta tesis se circunscribe al ámbito más restringido y tradicional de la psicología cognitiva, tratando de entender cómo funciona la credibilidad en relación a los objetivos epistémicos de los individuos.

PARTE I: ANÁLISIS HISTÓRICO Y FILOSÓFICO DE CONCEPTOS FUNDAMENTALES

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

CAPÍTULO 1

Argumentación

Este capítulo se compone de cinco secciones generales. La primera, dedicada a los primeros planteamientos sobre el funcionamiento del razonamiento humano como herramienta epistémica. La segunda, dedicada a la lógica concluyente en su progresiva formalización y sistematización hasta quedar constituida como parte indivisible del razonamiento matemático. La tercera, dedicada a la lógica no concluyente como cuerpo central del razonamiento cotidiano implicado en los procesos dialécticos y retóricos, y la progresiva transformación de estas dos disciplinas hasta su práctica desaparición. La cuarta sección está dedicada a la revitalización de ambas a mediados del siglo XX de la mano de Perelman y Toulmin, exponiendo brevemente su modelo para, posteriormente, pasar a comentar la posterior consolidación de los estudios de la argumentación. En esta consolidación, se han retomado conceptos de la retórica y la dialéctica clásica como los conceptos de tópicos, esquemas de argumentación, éndoxa y entimema, todos ellos relevantes para el modelo que se desarrollará en la segunda parte de esta tesis.

A lo largo de esta tesis, se sostiene que las modificaciones de la credibilidad asignada a una proposición se llevan a cabo mediante procesos de razonamiento cotidiano y de comprensión del discurso. En el capítulo anterior, se presentó la argumentación retórica como el cuerpo de conocimientos que más ha estudiado el cambio racional de credibilidades, mucho antes de que la psicología se interesara por ello. Aunque hoy en día la relación de la retórica con los estudios del discurso es indiscutible, su vertiente racional es menos conocida, y aquí se considera fundamental. De la retórica surgen los conceptos mencionados de entimema, éndoxa y esquema de argumentación.

Es importante conocer las primeras propuestas para el estudio del razonamiento y sus relaciones con la credibilidad. Originalmente, el estudio de la lógica era el estudio del razonamiento humano. Una parte de la lógica trataba del razonamiento exacto, utilizado para hacer demostraciones, preservando la verdad mediante pasos inferenciales. Otra parte abordaba el

razonamiento aproximativo, que opera en contextos cotidianos, cambiantes e inciertos, especialmente en contextos comunicativos donde se pretende convencer a un interlocutor. El primer tipo de razonamiento buscaba producir creencias verdaderas justificadas, mientras que el segundo producía creencias justificadas independientemente de su verdad. En ambos casos, el razonamiento estaba ligado a las creencias y a la justificación.

A lo largo de las siguientes páginas, se detallará cómo esta concepción inicial se fue desdibujando con el tiempo, y cómo los componentes de la lógica y la argumentación se fueron separando a lo largo de los siglos. Esto ayudará a comprender por qué hoy la lógica se entiende como una rama de las matemáticas, y la retórica como una disciplina de la estilística discursiva, sin que ninguna de las dos sea vista como un método para el estudio psicológico de las creencias. La lógica concluyente o demostrativa evolucionó hacia niveles de mayor abstracción, alejándose de los elementos del lenguaje natural para convertirse en una disciplina formal. Esta formalización de la lógica tuvo un impacto significativo en la comprensión del conocimiento y la ciencia desde Descartes. Y según varios filósofos, esa nueva forma de entenderlos, basada en criterios de verdad y de validez, ejerció una gran influencia en cómo se interpretó su contraparte no concluyente o cotidiana. El razonamiento cotidiano, por su parte, se asimilaba al razonamiento probable implicado en la dialéctica y la retórica. Después de Platón, la dialéctica se concibió como un método para debatir asuntos en los que no había un conocimiento exacto, donde ganaba quien lograba la aceptación de su tesis por parte del oponente. La retórica, por otro lado, se enfocaba en los medios de convicción pertinentes para cada caso, utilizados en el discurso público dirigido a una audiencia. En ambos casos, las premisas no se basaban en la verdad, sino en lo que el interlocutor o la audiencia aceptaba, razonando sobre ese material cotidiano y cambiante.

Durante el auge de los formalismos, ambas disciplinas fueron denostadas por su incapacidad de llegar a resoluciones concluyentes y su capacidad de cambiar creencias. La retórica empezó a seccionarse en sus diversos componentes, donde su componente de razonamiento primero se fue a la dialéctica, y posteriormente desapareció con ella. Lo que perduró fue su interpretación estilística. Pero esta tendencia empezó a cambiar debido a dificultades en la lógica formal, que impulsaron la reivindicación de la lógica no concluyente de la dialéctica y la retórica, dando origen a los estudios de la argumentación. Estas voces contemporáneas reivindicaban lo racional en ámbitos cotidianos más allá de la lógica formal y las matemáticas, incluyendo la acción, la ética y la jurisprudencia, áreas a las que la lógica formal no podía aplicarse.

Debido a la vertiente comunicativa de la lógica no concluyente, se incorporaron aspectos pragmáticos, acercándola a los estudios de la pragmática del discurso y la lógica conversacional. Los estudios de la argumentación han centrado sus esfuerzos en *restaurar la dialéctica* como un juego dialógico para la resolución de conflictos de opinión, investigando las reglas del debate, los movimientos para evitar falacias y las estrategias ganadoras para lograr la aceptación de las tesis en la interacción comunicativa. La nueva dialéctica, con su renovado enfoque social, normativo y procedimental, busca obtener la aceptación del oponente mediante reglas lógicas y pragmáticas.

Busca establecer una afirmación por derivabilidad de las reglas del juego, pero no causar un aumento de credibilidad en el oponente. Algunos autores actuales de esta perspectiva consideran que la argumentación debe estudiarse como un conjunto de actos lingüísticos y no lingüísticos con función comunicativa regidos por normas pragmáticas, y no como un proceso psicológico o de creencias epistémicas. Esta perspectiva se enseña en instituciones educativas bajo el nombre de pragmadialéctica y pensamiento crítico. Recientemente, se han realizado esfuerzos por unir la nueva dialéctica pragmática con la lógica formal en el proyecto de la dialéctica formal.

En la *rehabilitación de la retórica*, aunque hay interés en su lógica especial y su relación con la pragmática del discurso, aún no hay propuestas que la conecten directamente con los estudios del razonamiento humano ni con la credibilidad. Sin embargo, se admite que la retórica está relacionada con la convicción y la persuasión, produciendo algún tipo de cambio en las credibilidades. Por estas razones, la credibilidad no se mencionará explícitamente con frecuencia en las próximas secciones. En su lugar, se ofrece una aproximación histórica de la evolución de las disciplinas implicadas, para delimitar el terreno de la discusión e introducir parte del aparato conceptual que se utilizará en los próximos capítulos. Entre otros, se introducirán los conceptos de éndoxa, esquema de argumentación y algunos aspectos relativos a los entimemas en varios subapartados, cuya descripción detallada se encuentra en el ap. C.

1. Los orígenes de la discusión: los Analíticos, los Tópicos y la Retórica de Aristóteles

El origen del estudio del razonamiento en la tradición occidental se remonta a los filósofos de la Grecia clásica, sobre todo, a la obra de Aristóteles con su teoría del silogismo. Aristóteles estaba interesado tanto en el razonamiento matemático y filosófico, como en el razonamiento práctico y cotidiano. Se ocupó del primer tipo de razonamiento, esto es, del razonamiento concluyente dirigido a la demostración, en su tratado *Sobre la Interpretación*, y en los *Analíticos Primeros y Posteriores* (Garnham y Oakhill, 1996). A su vez, se ocupó del razonamiento práctico, dirigido a la acción, en la *Ética a Nicómaco*. Y, finalmente, del razonamiento cotidiano, dirigido a la argumentación, en los *Tópicos*, en las *Refutaciones Sofísticas* y en la *Retórica* (Oaksford y Chater, 2007; Vega, 2004; Walton 1999, 2001).

1.1. El razonamiento lógico demostrativo y concluyente de los Analíticos

En los *Analíticos* establece su teoría del silogismo demostrativo y concluyente, principalmente motivada por dar con una teoría normativa acerca de cómo se debería razonar para emitir aserciones racionalmente justificadas. Y esta justificación se encuentra en la aplicación de unas reglas deductivas de inferencia que dotan de *validez* a la *forma* de los argumentos.

Brevemente, un silogismo consiste en un razonamiento en el que intervienen tres términos dispuestos en dos premisas (dos términos en cada premisa), de las que se extrae una conclusión (a

partir del término repetido en las dos premisas). Por ejemplo, extraeríamos la conclusión de que “Sócrates es mortal” a partir de las premisas de que “Sócrates es hombre” y de que “todos los hombres son mortales”, donde los tres términos son “Sócrates”, “hombre” (término repetido) y “mortal”. Se trata, pues, de extraer una información nueva a partir de la combinación de dos informaciones dadas en las premisas. La nueva información, en realidad, ya estaba contenida en las premisas, pero es diferente de la que aportaban por separado. Como veremos, el concepto de silogismo va a ser muy relevante en esta tesis. A nivel teórico, porque los entimemas son silogismos; y a nivel empírico, porque los materiales de los experimentos que se presentarán al final de la tesis consisten en silogismos. Por tanto, lo iremos retomando y perfilando en varios de los próximos capítulos.

Por otro lado, que el silogismo tenga una forma válida consiste en que opere mediante reglas de inferencia válidas (correctas). Estas reglas de inferencia válidas consisten, a su vez, en una serie de operaciones mecánicas que se aplican a los términos del argumento, asegurando la verdad de las conclusiones a partir de la verdad de las premisas. Esto es, que, si las premisas son verdaderas, su aplicación asegura que la conclusión será necesariamente verdadera: los argumentos válidos son *preservadores de verdad*. De este modo, se establecen las condiciones para un razonamiento correcto y unívoco. Junto con el concepto de silogismo, también el de validez jugará un papel importante en la preservación de la credibilidad en esta tesis. En este sistema lógico concluyente, las premisas y conclusiones sólo admiten dos valores de verdad: verdadero y falso; es, por tanto, un sistema lógico bivalente. Esto es así en general, aunque es pertinente matizar que en los propios *Analíticos* ya se encuentran ideas de silogismos con premisas dentro del campo de lo posible, y de lo probable (no considerándose, por tanto, ni verdaderas ni falsas).

La teoría silogística desarrollada en los *Analíticos* fue la primera en presentar una formulación específica del método axiomático como organizador de sus enunciados y garante de la verdad de sus teoremas. Además, fue también la primera en formalizar los razonamientos empleando letras (p.ej. A, B, C) como variables para representar términos (p.ej. mayor o predicado, medio, menor o sujeto), cuyo contenido se correspondería finalmente con representaciones de objetos del mundo (p.ej. mortal, hombre, Sócrates). Retomando el ejemplo anterior, extraíamos que C es A a partir de que C es B, y de que Todo B es A. Y esta sería la *forma o figura* de este argumento. Pero este sistema aún guardaba una relación muy estrecha con el lenguaje natural. Así, esta lógica aún no es la que conocemos como lógica formal, sino aquella que le cedería el paso (Garnham y Oakhill, 1996).

1.2. El razonamiento cotidiano y la lógica no-concluyente de los Tópicos y la Retórica

En la vida práctica razonamos sobre contenidos cuya naturaleza nos es imprecisa. En este sentido, Aristóteles adopta una perspectiva dialéctica y retórica en su aproximación al razonamiento cotidiano, considerándolo como un razonamiento sobre lo probable, basado, no ya en verdades, y ni siquiera en certezas, sino en verosimilitudes y signos, donde lo que se afirma en las premisas puede estar sometido a excepciones y a la incertidumbre que nos produce un mundo cambiante.

En este abordaje, Aristóteles desarrolla en los *Analíticos Primeros* y ciertos aspectos en los *Tópicos*, pero, sobre todo, en la *Retórica* su teoría del *entimema*. Concibe los entimemas como un tipo específico de silogismo no necesariamente concluyente. Este tipo de silogismo está constituido por premisas cuyo *contenido* es considerado por una *audiencia* dada como algo meramente *probable* (debido a que está sujeto a excepciones), no como una verdad absoluta. Pero, desde luego, esta consideración de algo como más o menos probable va a depender del *conocimiento* sobre el asunto que tenga la audiencia, y de las opiniones comúnmente establecidas en una comunidad epistémica dada. En cualquier caso, los principios deductivos operan aquí de forma semejante a como lo hacían en los silogismos concluyentes. Así, los entimemas también son preservadores: partiendo de premisas consideradas, en general, como meramente probables, se llega a conclusiones que, consecuentemente, serán también sólo probables (en el ap. C se matiza esta afirmación). Pese a seguir respetándose la corrección formal, la sensibilidad de la audiencia al contenido es lo que pasa a ocupar el primer plano en este tipo de razonamientos.

Esta es una de las características principales de los entimemas. Y los entimemas son el instrumento principal de la retórica clásica. Por tanto, vemos que el objetivo original de la retórica tradicional no era tanto el de servir como un género ornamental y estilístico en la comunicación, sino el de proveer de argumentos razonables (verosímiles) por los cuales la audiencia debería sostener racionalmente determinadas creencias y opiniones en áreas donde el conocimiento es inexacto (Oaksford y Chater, 2007), típica, aunque no exclusivamente, en áreas como la ética, la política y la jurisprudencia. Así pues, la retórica era considerada como la “disciplina técnica que estudia los medios de convicción o persuasión a través del discurso” (Vega, 2012) ofreciendo razones aceptables para tomar decisiones bajo condiciones de incertidumbre (Zarefsky, 1995). En este sentido, en Aristóteles, el razonamiento cotidiano se entiende como un razonamiento probabilístico (Racionero, 1990), pero abordado aún desde una perspectiva cualitativa (Johnson-Laird, 1994).

Según muchos autores, en estos trabajos está el germen de la teoría de la argumentación actual (e.g. Perelman y Olbrechts Tyteca, 1958; Toulmin, 2002; Van Eemeren y Grotendorst, 2011; Vega, 2004, 2012; Walton, 2021), así como de la lógica informal (e.g. Groarke, 2017, 2022; Vega, 2011), del razonamiento plausible (Rescher, 1976; Walton, 2001) y de la lógica del razonamiento rebatible, revocable o derrotable¹ (Bench-Capon y Dunne, 2007; Koons, 2017). En un sentido similar, también se podría decir que en ellos está el germen del modelo sobre el sistema de creencias que se propone en esta tesis, así como los conceptos fundamentales para construir los materiales de los experimentos destinados a demostrar partes de dicho modelo.

¹ El término original del que parten estas traducciones es el de “*defeasible*” (Hart, 1949), que ha sido traducido indistintamente por derrotable, rebatible o revocable.

2. Desarrollo de la lógica concluyente después de Aristóteles

2.1. La Edad Moderna: el racionalismo y una nueva forma de entender la lógica

Desde Aristóteles hasta la Edad Moderna, se observa una gradual exclusión del razonamiento cotidiano en estudios lógicos, enfocándose más en la validez formal y la deducibilidad como principios rectores. En el siglo XVII, figuras del racionalismo como Pascal, Descartes y Leibniz buscaron desarrollar un lenguaje universal, eliminando las ambigüedades de los lenguajes naturales y creando sistemas lógicos formales, concisos y claros, aunque sin grandes avances significativos, según lo comentado por Kant en 1787.

En el siglo XIX, George Boole, De Morgan y otros pusieron las bases de la lógica matemática al integrar formalismos matemáticos con la lógica. Boole desarrolló una lógica binaria algebraica en 1847 y posteriormente aplicó la teoría de la probabilidad a esta lógica en su obra *Las Leyes del Pensamiento* (1854). En todo caso, este último trabajo, adoptó una perspectiva más sociológica y psicológica que estrictamente matemática (Burris, 2018). Gottlob Frege (1879, 1884) luego elevó la abstracción de la lógica con un lenguaje formal que fundamentó el cálculo proposicional y la lógica de predicados.

Avanzando al siglo XX, el intento de Frege por fundamentar matemáticamente la lógica culmina en una serie de trabajos desarrollados por Russell y Whitehead en los *Principia Mathematica* en torno a 1910, aunque un problema con uno de sus axiomas eliminó esta posibilidad. El último intento de desarrollar un programa completo de formalización fue llevado a cabo por Hilbert en la década de 1920. Pero los trabajos de Gödel (1931), Church (1936) y Turing (1936) demostraron que no es posible probar los fundamentos axiomáticos de la aritmética ni asegurar la consistencia de un sistema matemático completo utilizando únicamente sus propios axiomas. Como se puede observar, aquí la lógica tiene ya poco que ver con el estudio del funcionamiento del razonamiento humano. En la segunda mitad del siglo XX, la lógica formal se aplicó extensamente en las ciencias de la computación. Esto incluyó el desarrollo de lenguajes formales, que consisten en símbolos y reglas especificadas para operaciones formales. Estos son esenciales en la computación y la inteligencia artificial actual, donde las operaciones algorítmicas y las reglas de inferencia son clave para procesar y generar nuevas fórmulas o cadenas de símbolos. En este orden de cosas, la naturaleza de la computación se entiende como el ejercicio de operaciones formales sobre símbolos. Estas operaciones, que se ejecutan en una serie de pasos sucesivos en el tiempo llamados algoritmos, son independientes de las posibles propiedades semánticas que puedan tener o no los símbolos. Es decir, las reglas son ciegas al contenido.

2.2. La lógica y el problema de la incertidumbre en la Inteligencia Artificial

Alan Turing definió operativamente la inteligencia como un sistema computacional más, iniciando el interés en la intersección entre lógica y cognición humana y estableciendo las bases de la Inteligencia Artificial (IA). Inicialmente, la IA adoptó una posición logicista, utilizando lenguajes formales y reglas deductivas, con modelos cognitivos, donde destacan los trabajos de Newell y

Simon (1956). Sin embargo, enfrentó problemas técnicos y teóricos, como la extensión del tiempo requerido para los procesos de búsqueda combinatoria y las dificultades en especificar las condiciones necesarias para la validez deductiva en contextos del mundo real.

Durante los años 70, emergieron modelos de procesamiento distribuido en paralelo como alternativa a la serialidad de los modelos simbólicos para paliar los problemas de la extensión temporal. Por otra parte, se comenzó a "empaquetar" conocimientos de ámbitos específicos del mundo real en esquemas y guiones para sistemas expertos (e.g., Schank y Abelson, 1975), aunque estos resultaron rígidos frente a situaciones cotidianas de incertidumbre. McCarthy y Hayes (1969) exploraron el razonamiento "derrotable", identificado por Aristóteles (Koons, 2017), crucial para la teoría de la argumentación en IA. Se desarrollaron sistemas como las lógicas modales, deónticas (von Wright, 1951), dialógicas (Lorenzen, 1955; Lorenzen y Lorenz, 1978), difusas (Zadeh, 1965), por defecto (Reiter, 1980) y no monótonas (Ginsberg, 1987), así como la lógica plausibilística (Pollock, 1987; Rescher, 1976; Walton, Tindale y Gordon, 2014), Teoría de Revisión de Creencias (Alchourrón, Gärdenfors y Makinson, 1985) y las redes de herencia semántica (Horty, Thomason y Touretzky, 1990), tratando de formalizar este tipo de razonamiento sujeto a incertidumbre.

Paralelamente, se adoptaron modelos probabilísticos como las Redes Bayesianas, desarrolladas por Judea Pearl (1988; Pearl y Russell, 2000), que permiten gran flexibilidad al incorporar inferencias bidireccionales, se convirtieron en un marco general para representar conocimiento bajo incertidumbre. Estos avances han llevado a la IA a moverse más allá de la lógica formal clásica, buscando formalizar las inferencias que se producen en condiciones de incertidumbre y con contenido rebatible, acercándose a los tipos de razonamientos tradicionalmente estudiados por la dialéctica y la retórica.

El modelo sobre el sistema de creencias que se expondrá en esta tesis adoptará también una perspectiva probabilística pero subjetivista para definir la medición de la credibilidad y delimitar el tipo de reglas y cálculos a los que se atiene la dinámica de credibilidades.

3. Desarrollo del razonamiento cotidiano y la lógica no-concluyente

Pero ¿cómo evolucionó el estudio del razonamiento cotidiano durante todos esos siglos hasta el momento presente?

Como vimos, el estudio del razonamiento cotidiano en la antigüedad clásica, tanto de Grecia como de Roma, siempre estuvo ligado a los intereses prácticos de la argumentación dialéctica y retórica, de la convicción y la persuasión; en definitiva, siempre estuvo ligado a los procesos de *cambio de credibilidades*. En esta línea, en el presente trabajo defiende que los procesos del razonamiento cotidiano son los mecanismos por los que opera el cambio de credibilidades. Esto hace necesario detenernos algo más en el análisis de la evolución del razonamiento cotidiano. Y hablar de la evolución del estudio del razonamiento cotidiano implica hablar del devenir de la

dialéctica y la retórica desde la antigüedad hasta su resurgimiento en nuestros días, lo que, a su vez, permitirá explicar algunas de sus características y ciertos conceptos principales que van a jugar un rol relevante en esta tesis. Eso es lo que me dispongo a hacer en las páginas que vienen, prestando especial atención al tratamiento recibido en la antigüedad clásica y, sobre todo, en la época contemporánea a partir de la segunda mitad del siglo XX. En cualquier caso, no pretende ser esta una revisión exhaustiva, y soy consciente de que no podré hacer justicia de todas las perspectivas y todos los autores que han abordado estas cuestiones.

3.1. Dialéctica y Retórica en la Grecia Clásica: sentando los cimientos de la argumentación como paradigma del razonamiento cotidiano

Iniciamos esta sección con una cita que expresa la relevancia de los fundamentos históricos para una comprensión adecuada de la actual teoría de la argumentación:

Al igual que la investigación en muchas otras disciplinas, el estudio de la argumentación se remonta a la antigüedad clásica. A diferencia de lo que sucede en la mayoría de aquéllas, sin embargo, el conocimiento de la literatura antigua sigue siendo una condición necesaria en la teoría de la argumentación [...] Algunas concepciones teóricas formuladas por autores clásicos, como Aristóteles y Cicerón, pertenecen todavía al núcleo central de la argumentación. Son una parte integral de los fundamentos de las herramientas hermenéuticas y críticas que están actualmente disponibles para el análisis y la evaluación del discurso y los textos argumentativos.

Van Eemern y Grotendorst, 2003, p. 51

3.1.1. Dialéctica

Originalmente, en la Grecia clásica, la argumentación dialéctica era concebida como una herramienta necesaria en el desempeño filosófico de hallar la verdad de forma razonada a través de un diálogo constituido por preguntas y respuestas. En este ejercicio, eran necesarios un proponente que contestaba y un oponente que preguntaba buscando errores en los argumentos del adversario. El proceder socrático es un claro ejemplo de ello. Era, pues, la técnica del razonamiento dialógico que regulaba las conversaciones y debates que tenían por objeto la verdad de una cuestión. Así, para Platón, la dialéctica era el método filosófico por excelencia, dado que, a través de la contrastación y el correcto razonar, se hacía posible pasar del conocimiento sobre las apariencias del mundo sensible, a la adquisición del verdadero conocimiento que sólo el intelecto podía captar. Por tanto, para Platón, el valor de la dialéctica no estaba tanto en que expresaba la forma en la que los humanos razonan, sino como un método de alcanzar la verdad que busca el filósofo.

Y durante cierto tiempo, también para Aristóteles fue el método filosófico más adecuado. En los *Tópicos*, Aristóteles muestra cómo argumentar a favor o en contra de una aserción dada en un debate dialéctico. Pero, al hacerlo, vio que el dinamismo implícito en todo proceso dialéctico se hace posible debido a que sus premisas no tienen el estatus de verdades primeras y acabadas. En

la dialéctica, no se parte de axiomas de los cuales se derivan teoremas. A raíz de esto, Aristóteles acaba considerando que la dialéctica es insuficiente para llegar al conocimiento concluyente ya que no permite hacer demostraciones.

Al no partir de unas verdades para llegar a otras, la dialéctica se constituye ahora como un método para la *selección y justificación* de *hipótesis*. Y tales hipótesis consisten en opiniones, en el sentido de que “una tesis de la que no se sabe todavía si se cumple en todos los contextos posibles constituye efectivamente una opinión, un enunciado de validez subjetiva” (Racionero, 1990, p. 33). Así pues, la selección del contenido de las premisas para un debate dialéctico se realiza a partir de las opiniones generalmente aceptadas como verdaderas para la mayoría, o para los más sabios y reputados (es decir, *tà éndoxa*, que se traduce como ‘opiniones plausibles’). De este modo, el criterio de verificación de la argumentación dialéctica (y retórica) estará menos en el plano de lo real que en el de las opiniones: será el sistema comunitario de opiniones, el cuerpo de creencias explícitas de una comunidad epistémica particular. Así, “la dialéctica puede *probar* la credibilidad de una tesis, confrontándola con el sistema de las opiniones comunes”² (Racionero, 1990, p. 35). Teniendo en cuenta que el objeto de la opinión común es lo plausible, y lo plausible parte de la base de lo probable³, el razonamiento dialéctico se sitúa en el marco del cálculo de probabilidades (por oposición al cálculo de verdades). Esta condición de plausibilidad y probabilidad en lugar de verdad es una diferencia crucial entre el razonamiento dialéctico y el analítico concluyente.

Así, a pesar de considerar que la dialéctica no es óptima para la adquisición de conocimiento, Aristóteles la tiene por respetable y útil en varios sentidos. Uno de ellos es el de que muestra cómo podemos razonar sobre campos en los que un conocimiento exacto no es posible. Por ello, el terreno natural de aplicación del razonamiento dialéctico será generalmente el de la praxis humana, el de la ética, el de la política y el de la jurisprudencia. Si bien esto último es compartido también por el razonamiento retórico, la dialéctica es útil aun en otro sentido que le es más propio, a saber, como técnica en orden a ejercitarse para el debate crítico en el examen de cualquier materia meramente probable que sea objeto de controversia. Pero advierte el autor del peligro de que pueda derivar en un pseudo saber en el que se tome lo aparente como real: no hay que confundir lo verosímil, con lo aparentemente verosímil.

3.1.2. Retórica

En cuanto a la argumentación retórica, a ésta se le aplica gran parte de lo dicho ya para la dialéctica, pero con un interés mayor en su vertiente comunicativa. La retórica griega fue concebida como una herramienta necesaria en una sociedad prominentemente oral, donde los ciudadanos debían defender sus posiciones ante los tribunales, o debían exhortar a tomar decisiones deliberativas de carácter político, económico o militar, y donde se elogiaban o censuraban cualidades morales de determinados personajes o instituciones públicas.

² O, en otras palabras, fija la credibilidad otorgada al contenido de la proposición *q*, dado el sistema de creencias *p*.

³ Ver sec. 3, cap. 2, pt. I donde se expone la relación entre lo probable y lo plausible en la argumentación.

De este modo, y a diferencia de la dialéctica, el estudio de este tipo de razonamiento cotidiano nunca tuvo como finalidad la de ser un medio para acceder a la verdad (lo que fue duramente criticado por Platón), sino que tenía una finalidad de carácter social, claramente comunicativa y persuasiva. Teniendo en cuenta dicho carácter, el razonamiento retórico quedaba circunscrito a tres géneros de discurso: el discurso forense o judicial, razonando sobre lo que probablemente pudo o no suceder en un pasado hipotético; el discurso deliberativo, referido a la toma de decisiones en base a predicciones acerca de lo que probablemente pueda o no pasar en un futuro hipotético; y el discurso epidíctico, referido a la evaluación ética de un sujeto en base a las opiniones comúnmente establecidas acerca de lo aprobable o reprobable. Tanto los tiempos hipotéticos mencionados, como el contenido y la situación donde se deben aplicar los diferentes discursos, así como el abanico de opiniones comúnmente establecidas en una sociedad hacían imprescindible la toma en consideración del contexto espacial y temporal de los discursos para su efectividad retórica. Estas referencias al *contenido* y al *contexto* son algunas de las diferencias principales del tipo de razonamiento cotidiano abordado por la retórica respecto del razonamiento analítico, que trabajaba con verdades formales descontextualizadas y de interpretación unívoca.

Así pues, por su finalidad comunicativa, los razonamientos retóricos se expresaban en discursos dirigidos a una audiencia particular. Por tanto, la *audiencia* también conforma una parte esencial del contexto comunicativo. Por ello, en todas las obras clásicas de retórica se dedica una atención especial al estudio de las características de la audiencia. Aristóteles puso de manifiesto que los géneros del discurso van ligados a los tipos de audiencias. Así pues, había tres géneros de discursos porque había tres tipos de audiencias a las cuales mover o inclinar mediante la argumentación, pero de haber más tipos de audiencias, habría más tipos de discursos; no hay un límite preestablecido.

En este punto se puede ya observar que las disciplinas dialécticas y retóricas buscaban producir cambios en la aceptabilidad o credibilidad que un oyente o una audiencia otorgase a un contenido dado. La retórica, además, también buscaba producir cambios en la toma de decisiones relativas a acciones y cambios en las actitudes respecto a determinados objetos. Los cambios mencionados no debían ser producidos por cualquier cosa, sino por cierto tipo de razonamientos semejantes a los deductivos, pero falibles, cuyos contenidos versaran sobre aquello que es o no probable, sobre lo que puede o no ser. Finalmente, estos razonamientos debían ser expresados mediante discursos comunicativos dirigidos a un interlocutor o a una audiencia concreta, en una situación concreta. De este modo, la aceptabilidad y el cambio de credibilidades quedaban directamente unidos al razonamiento cotidiano y a la pragmática del discurso⁴.

⁴ Si bien, aún no en el sentido que tiene hoy la *pragmática* como disciplina, surgida de la lingüística y la filosofía del lenguaje a partir de principios de la década de los 60 del s. XX.

3.2. Evolución de la dialéctica y la retórica desde la Grecia Helenística hasta la Modernidad

Desde la Grecia Helenística hasta el mundo romano, la dialéctica mantuvo una función esencial dentro de la educación filosófica, usada para discernir la verdad y detectar errores en el debate, una práctica que resurgió con el neoplatonismo. Destacados filósofos como Cicerón y Quintiliano continuaron esta tradición, enfatizando su importancia en la lógica filosófica. En contraste, la retórica se transformó significativamente. Originalmente diseñada para influir en opiniones sobre temas éticos y políticos mediante discursos bien estructurados, evolucionó hacia un arte más estilístico y literario. Esta evolución se intensificó en el periodo romano, distanciando la retórica del análisis lógico y acercándola a la poética (Barthes, 1975).

Durante la Edad Media, la dialéctica y la retórica, junto con la gramática, formaron el trívium del sistema educativo medieval, en contraposición al cuadrivium, de carácter más matemático. La dialéctica siguió siendo una herramienta clave para el razonamiento y la argumentación, empleada para descubrir la verdad, detectar sofismas y acceder al reino de las cosas creíbles, de los *credibilia* (Ferrater Mora, 2005). Sin embargo, la retórica se redujo a un ejercicio escolástico enfocado en la redacción y predicación de sermones. En el Renacimiento, la retórica se revitalizó como herramienta de elocuencia política, pero sin recuperar su antigua relevancia lógica y de razonamiento sobre las cosas probables.

Hacia el final de la Edad Media y el comienzo de la Modernidad, la retórica se había consolidado como una disciplina centrada en la elocuencia y el estilo, alejada de su función original de razonamiento práctico. Este enfoque perduró hasta bien entrado el período moderno, marcando un largo período de estancamiento antes de su revitalización en el siglo XX.

3.3. Transformaciones de la dialéctica y la retórica desde la Modernidad hasta el siglo XIX

Durante la Edad Moderna, los avances en filosofía, matemáticas y ciencias marcaron un distanciamiento de la escolástica medieval hacia una ciencia más determinista y mecanicista. Figuras como Descartes, Fermat, Leibniz y Newton revolucionaron la geometría analítica, que pasaría a tener un rol indispensable en nuevos los métodos de experimentación empírica y medición, contribuyendo a los primeros desarrollos en teoría de la probabilidad matemática.

En esta época, las matemáticas empezaron a integrarse con la lógica, la cual adquirió también un rol central en la nueva ciencia. Sin embargo, la dialéctica y la retórica, disciplinas arraigadas en el razonamiento basado en opiniones y lo probable, no se integraron bien en este nuevo paradigma. Salvo notables excepciones como los trabajos de la lógica de Port Royal, o el tratamiento dado por Jacob Bernoulli a la argumentación, la dialéctica, la retórica y el estudio del razonamiento cotidiano fueron ignorados. Dominada por el racionalismo, la primera mitad del siglo XVII valoró conocimientos derivados de formas intelectuales inmutables, pues en ello residía la verdad. Así, la aplicación de las matemáticas a la naturaleza permitía encontrar la inmutabilidad tras el cambio; sin embargo, el conocimiento proporcionado por los sentidos solo deriva en apariencia y opinión.

Descartes, Bacon, Ramus, Leibniz y más tarde Kant, vieron la dialéctica como un ejercicio inútil, reduciéndola a lo que Kant describió como una "lógica de la apariencia" (Kant, 1781, pp. 99-100; B86-A62). La retórica, por su parte, fue despreciada como meramente demagógica y estilística, criticada por su orientación hacia la adulación más que hacia la verdad. A lo largo del tiempo, se redujo a un conjunto de técnicas expositivas y ornamentales, entrando en un largo período de decadencia.

En el siglo XIX, bajo la influencia de Fichte y Hegel, la dialéctica se transformó y adoptó una interpretación empírica sobre el desarrollo histórico, una perspectiva que fue ampliada por Marx, Engels y el socialismo soviético. Es decir, se empleaba el término, pero con pocos vínculos con su significado tradicional relativo al diálogo racional. Por su parte, la retórica continuó su descenso, relegada a un arte superficial, desconectada de su función práctica en el razonamiento y la persuasión (Ferrater Mora, 2005; Morata Garavelli, 1997).

4. La restauración del s. XX: Perelman, Toulmin y la Teoría de la Argumentación

De este modo es como entran al siglo XX. Pero en la primera mitad de dicho siglo, se gestó un renovado interés por la argumentación, la dialéctica y la retórica, impulsado por cambios políticos, económicos y tecnológicos en los países occidentales. Según González Bedoya (1989), la democracia y la expansión de los medios de comunicación masiva incrementaron la necesidad de estudiar técnicas de comunicación y estrategias persuasivas. Además, el auge del capitalismo y la globalización fomentaron la demanda de conocimientos en persuasión provenientes de la dialéctica y la retórica.

Este resurgimiento también fue motivado por el colapso del proyecto logicista en la lógica formal a mediados de siglo, que llevó a algunos filósofos a cuestionar la adecuación de la lógica formal como una herramienta válida para el conocimiento. Por otra parte, se volvió a plantear su relación con la racionalidad, y con el razonamiento humano. En estas coordenadas, algunos filósofos se plantearon que, o en la vida cotidiana los humanos no se comportan de manera racional, o deberá cambiar la concepción de la racionalidad. Esto último es lo que tratarán de hacer dos filósofos procedentes, al principio, de la lógica formal: Chaim Perelman y Stephen Toulmin, considerados hoy los artífices del renacimiento de los estudios sobre argumentación.

4.1. Perelman y La Nueva Retórica

En un principio, Perelman trata de aplicar los cánones de la lógica formal a la teoría de la jurisprudencia. En ese intento, se da cuenta de que es una empresa, cuando menos, artificiosa. Según él, ni la jurisprudencia, ni las normas morales, ni las políticas pueden ser adecuadamente formalizadas debido a la práctica imposibilidad de partir de premisas necesarias y universales. Deberán ser abordadas desde otra lógica. El paradigma de tal lógica, según Perelman, es la *Retórica* de Aristóteles. De este modo, Perelman redescubre la dialéctica y retórica aristotélicas,

entendiéndolas de nuevo como lógicas de lo preferible, o razonamientos sobre lo probable dirigidos al cambio de credibilidades, reintroduciendo de nuevo el interés filosófico por esta materia que había ido menguando desde la antigüedad.

En 1958, Perelman y Olbrechth-Tyteca, publican el *Tratado de la Argumentación: La Nueva Retórica*. En esta obra, los autores se centran en el estudio de las estructuras y técnicas argumentativas (en los llamados *esquemas de argumentación*⁵), en el estudio de los medios de prueba para obtener la *adhesión*, en el mecanismo del pensamiento persuasivo y en el estudio de las audiencias. En sus propias palabras, “el objeto de esta teoría es el estudio de las técnicas discursivas que permiten provocar o aumentar la adhesión de las personas a las tesis presentadas para su asentimiento. Lo que caracteriza esta adhesión es la variabilidad de su intensidad.” (Perelman y Olbrechth-Tyteca, 1958/2006, p. 34).

Llevan a cabo este trabajo analizando los medios que emplean habitualmente las ciencias humanas y sociales o el derecho para probar sus tesis. Así, de forma complementaria a la demostración científica o lógico-matemática, hablaríamos de la *justificación* como el quehacer propio de la argumentación. Esto es lo que va a permitir aplicar la razón a la ética, a la justicia y a la acción. Del mismo modo en que pasan de la demostración a la justificación, pasarán también de la dicotomía ‘verdadero – falso’ a hablar de *grados de adhesión de intensidad variable* de una audiencia respecto de una tesis. Por otro lado, su preocupación explícita por el estudio del razonamiento cotidiano queda patente cuando dice que “nuestras preocupaciones son más las de un lógico deseoso de comprender el mecanismo del pensamiento que las de un maestro de oratoria preocupado por formar a procuradores” (p. 37). Finalmente, para estos autores:

La teoría de la argumentación que pretende, gracias al discurso, influir de modo eficaz en las personas, hubiera podido estudiarse como una rama de la psicología. Naturalmente, si los argumentos no son apremiantes, si no deben convencer necesariamente sino que poseen cierta fuerza, la cual puede variar según los auditorios, entonces ¿acaso se la puede juzgar por el efecto producido? El estudio de la argumentación se convierte así en uno de los objetos de la psicología experimental. (p. 41).

Esta tesis pretende dar respuesta a alguna de estas inquietudes.

4.2. Toulmin y Los Usos de la Argumentación

Por su parte, Stephen Toulmin también parte de una reacción al formalismo logicista. Uno de sus objetivos principales en varias de sus obras es “criticar el supuesto, asumido por la mayoría de filósofos anglosajones, de que todo argumento significativo puede expresarse en términos formales [entendido] como una deducción estrictamente concluyente según la geometría euclidiana.” (2002, p. 9). Al igual que Perelman, Toulmin carga aquí contra “la tradición platónica que, dos milenios más tarde, retomó e impulsó René Descartes” (*ibíd.*).

⁵ Este concepto se desarrolla ligeramente en la sec. 5.1., cap. 1, pt. I, y ampliamente en el ap. C.

Algunas de las primeras obras de Toulmin podrían enmarcarse mejor dentro del campo de la filosofía de la ciencia, donde hace los primeros cuestionamientos de calado a ciertos aspectos de la llamada “concepción heredada” de perspectiva (en parte) cartesiana (Echevarría, 1999), abogando en contra de la expresión de las teorías científicas en un lenguaje universal común, y en contra de su reducción axiomática.

4.2.1. El Modelo de Toulmin

En 1958 Toulmin publica su libro *Los Usos de la Argumentación*, con la intención de que fuera un trabajo sobre la epistemología del siglo XX, y no sobre teoría retórica o lógica informal. A pesar de ello, el título ya auguraba su destino. Desde el supuesto de que la lógica y la razón no son propiedad exclusiva del filósofo o del matemático, Toulmin se interesa por el uso cotidiano que hacen de ellas las personas en general. Considera que la práctica lógica que lleva a cabo el razonamiento cotidiano consiste en evaluar la fuerza de los argumentos, es decir, hasta qué punto nos encontramos ante un buen argumento expresado en el lenguaje natural. Para ello, se evalúa la corrección de los enunciados, la solidez de los cimientos en los que se basa y la firmeza del respaldo que podemos darle al conector inferencial. En esto consiste la aplicación de la lógica en los asuntos prácticos. Según él, todo esto estaba muy presente ya en las obras de Aristóteles, pero al no haber recibido ningún tratamiento posterior, no ha habido tampoco ningún progreso en más de dos milenios. Por ello, Toulmin reivindica un retorno del estudio de la lógica y la racionalidad al campo de la praxis de donde nació.

En ese trabajo, Toulmin presenta un análisis de los contextos argumentativos, así como de la estructura de los argumentos, detallando la función de cada uno de sus elementos (dato, garante y conclusión). Concibe los argumentos como una unión de afirmaciones que permiten hacer una inferencia en un contexto dado, sin las constricciones del silogismo analítico concluyente de premisas verdaderas. La Figura 1 (abajo) presenta el núcleo de su modelo, cuyos elementos serán explicados muy brevemente.

El modelo de argumentación desarrollado por Stephen Toulmin ofrece una forma flexible y práctica de analizar la estructura lógica de los argumentos tal como se presentan en el discurso cotidiano y profesional, en contraste con los modelos más rígidos y formalistas de la lógica tradicional.

Sus elementos son los siguientes:

Dato (Data): Son los *hechos* o la *evidencia* presentados para apoyar una afirmación. Los datos son la base sobre la cual se construye el argumento.

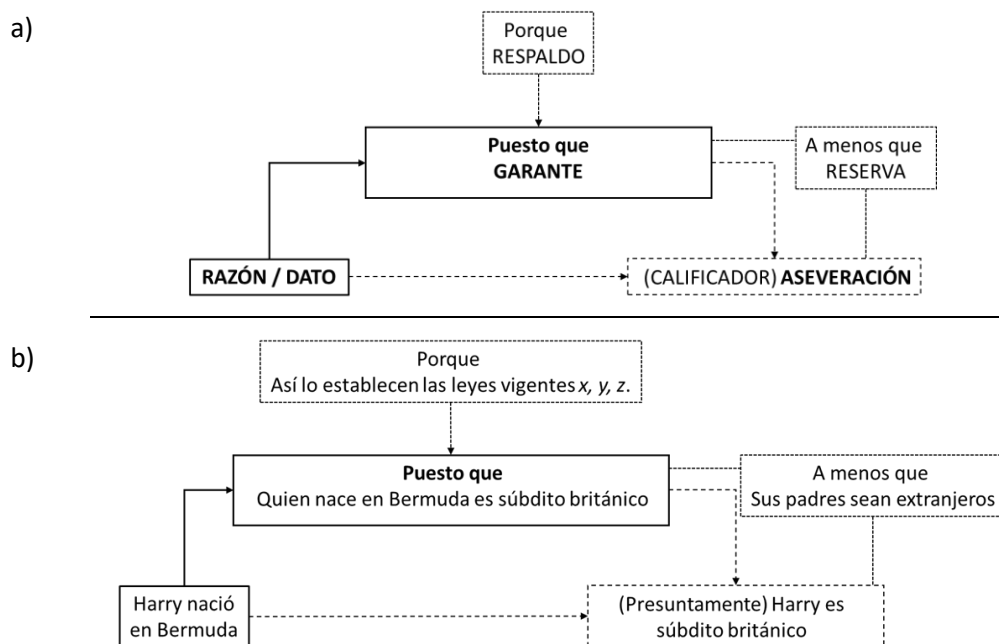
Garante (Warrant): Es la *conexión* entre los datos y la afirmación; es un principio general e hipotético que justifica el paso de los datos a la aseveración. La garantía explica por qué los datos respaldan la afirmación y es a menudo implícita y asumida como entendida.

Respaldo (Backing): Proporciona *soporte adicional al garante*, especialmente cuando la conexión entre los datos y la afirmación no es obvia. El respaldo puede consistir en teorías, datos adicionales o fuentes autorizadas que refuerzan el garante.

Reserva (Rebuttal): Reconoce las condiciones o circunstancias en las cuales el garante no sería válido, no funcionaría, y la conclusión podría ser incorrecta. La reserva es esencial para delinear los límites de la conclusión y mostrar en qué situaciones el argumento no aplicaría. En palabras del autor, son “circunstancias en las que la autoridad general de la garantía ha de dejarse a un lado” (Toulmin, 1958; p. 137). Estas posibles situaciones dan lugar al empleo de los calificadores.

Calificadores (Qualifiers): son “una referencia explícita al grado de fuerza que nuestras razones cofieren a nuestra aseveración en virtud de nuestra garantía” (Toulmin, 1958/2007; p. 137). Estos elementos también se presentan, muchas veces, de manera implícita.

Figura 1
Esquema del Modelo de Toulmin (1958)



Nota: en la parte superior de la figura (a) encontramos el esquema del modelo de Toulmin, con la nomenclatura traducida al castellano de los elementos que lo componen (tomada parcialmente de Bermejo-Luque, 2012). En la parte inferior (b), lo encontramos ejemplificado con contenidos de un argumento cotidiano.

El modelo de Toulmin es valorado por su capacidad para capturar la naturaleza dinámica y contextual de la argumentación real, lo que permite una comprensión más adecuada que la lógica formal sobre cómo se formulan y se defienden las afirmaciones en la vida cotidiana y en contextos profesionales. Esta perspectiva ha tenido una influencia considerable en diversas disciplinas, incluyendo la retórica, la comunicación y la filosofía de la ciencia.

4.2.2. Relevancia del Modelo de Toulmin para las propuestas de esta tesis

Esta tesis se inspira en varios aspectos del modelo de Toulmin para desarrollar un sistema de creencias que analiza la estructura y la dinámica del argumento. Es importante destacar la relevancia de este modelo, resaltando similitudes y diferencias en nuestra aproximación.

Uno de los aspectos adoptados del modelo de Toulmin es la integración de respaldos y reservas en el argumento, que justifican o refutan ideas, respectivamente, influyendo así en el grado de aceptación de una afirmación. Como veremos esta idea subyace a las relaciones de justificación y refutación que discurren en el sistema de creencias propuesto aquí. Por otro lado, en ambas propuestas se contempla la idea de que un dato proporciona una razón para justificar una idea, mediante el apoyo de una regla de relación general e hipotética que muestra la relevancia de estos datos para la conclusión. Sin embargo, el modelo de esta tesis amplía este concepto al incorporar relaciones tópicas, esquemas cognitivos y de argumentación. Además, tanto en el modelo de Toulmin como en el presente, las conclusiones obtenidas no se presentan como absolutamente verdaderas o falsas, sino que reciben un grado de justificación graduado. Ambos enfoques matizan la naturaleza binaria de la certeza en los argumentos y reflejan una comprensión más sofisticada de la justificación.

No obstante, existen diferencias significativas entre ambos modelos. En ambos modelos, las reglas hipotéticas adoptan una estructura condicional y actúan como puentes inferenciales que conectan los datos con las conclusiones. Pero en el modelo de Toulmin, los garantes se formulan como "Si Datos, entonces Conclusión" y, según Bermejo-Luque (2012), los garantes no son premisas del argumento. En contraste, en este modelo esas reglas hipotéticas se presentan como premisas del argumento que implementan una relación tópica justificativa, por ejemplo, del tipo "Si Causa, entonces Efecto", como se desarrollará más adelante (y en profundidad en el ap. C). En cualquier caso, en los experimentos que se presentarán en esta tesis, las premisas mayores de los argumentos serán denominadas "premisas garantes". Pese a ser conscientes de que no se trata exactamente del mismo concepto que los garantes de Toulmin, sí que es cierto que *garantizan* el paso inferencial.

Otra distinción importante radica en cómo se determina la fuerza de la justificación en el modelo de Toulmin, que depende de la existencia de reservas aplicables al garante. En nuestro modelo, se añaden reservas o refutadores aplicables también a la razón o dato mismo. Por lo tanto, aunque este modelo de credibilidad toma prestada la nomenclatura de Toulmin y se inspira en algunos de sus aspectos, no constituye una aplicación psicológica directa de su modelo.

4.2.3. Contraste entre silogismos analíticos y sustanciales en la Teoría Argumentativa de Toulmin

En el silogismo analítico, toda la información está ya contenida en las premisas, unas premisas cuantificadas (mediante el cuantificador universal *todos*, el existencial *algún*, y sus negaciones *ningún* y *algunos no*). La conclusión se obtiene a partir de una reordenación mecánica de los elementos de las premisas en relación con los cuantificadores; y el argumento será válido y preservador de verdad en función de estos aspectos formales. Para Toulmin, la mayoría de los

argumentos cotidianos no son de este tipo. Más bien, en ocasiones sirven para predecir el futuro en función de la experiencia pasada, o para inferir cómo se sienten otras personas basándonos en cosas que dijeron, o en sus gestos, o para justificar posicionamientos éticos, estéticos o políticos. Y en ninguno de esos casos se puede considerar que las conclusiones sean una mera reformulación del enunciado inicial, algo ya contenido en las premisas⁶.

Al silogismo ‘analítico’, Toulmin contrapone el silogismo al que llama ‘sustancial’, donde las claves de la validez residen en el *contenido*. Según el autor, la formalización de ese contenido, muchas veces, no permitiría derivar mecánicamente la conclusión reordenando sus elementos, y, sin embargo, seguiría siendo un argumento perfectamente válido (véase Toulmin, 1958/2007; p. 159 en adelante). Lo importante en estos argumentos es la *relevancia* de los hechos aducidos en apoyo de la conclusión, el *grado* en el que los datos disponibles *justifican* lo que se afirma en la conclusión.

Por otro lado, la lógica formal exige la cuantificación de los enunciados en los silogismos. Sin embargo, esto es algo que raramente ocurre en el lenguaje natural. Y cuando ocurre, las menos de las veces es mediante el universal, el existencial y sus correspondientes negaciones, sino con términos como *la mayoría*, *unos pocos*, *nadie*, etc., o con otros más cercanos a los operadores modales como *posiblemente*, *probablemente*, *muy poco probable*, etc. Así pues, la mayoría de las veces decimos simplemente “A es B”, y no “Todo A es B”, cosa que permite cierta indeterminación. La inclusión de los cuantificadores los haría inaplicables a la mayoría de los asuntos cotidianos, y tendría consecuencias harto indeseables para los terrenos de la acción, de lo ético, y de la jurisprudencia. Como afirma el autor, “las garantías de alcance más general, cuando se invocan en los argumentos éticos, pueden presentar excepciones en circunstancias inusuales, de suerte que como mucho pueden autorizar sólo conclusiones que no pasan de ser presunciones” (p. 158).

4.2.4. La jurisprudencia como modelo para la argumentación cotidiana según Toulmin

Dado pues que la racionalidad y la argumentación cotidianas se rigen por un criterio de justificación que admite refutadores, entonces su modelo no puede ser el de las matemáticas. ¿Cuál debe ser entonces? Después de analizar las posibilidades del modelo psicológico, sociológico y técnico-preceptivo, Toulmin llega a la conclusión de que éste debe ser el modelo de la jurisprudencia, donde uno debe dar y pedir razones, evaluando el *caso* que se aduce en favor o en contra de una tesis. Un buen argumento pasa a ser el que, admitiendo la posibilidad de excepciones a lo que afirma, resiste los embates de los refutadores, debiendo someterse al “Tribunal de la Razón” (p. 25).

En efecto, y a diferencia de Perelman, pese a que la forma en la que los humanos razonan cotidianamente es relevante para Toulmin, considera que transformar la lógica de la argumentación cotidiana en una rama de la psicología la hace demasiado subjetiva y la ata demasiado a los hábitos particulares de inferencia de la gente real, con sus diferencias individuales. Considera, de hecho,

⁶ Esta es otra diferencia respecto de nuestra propuesta: los entimemas, en tanto que argumentos deductivos, cumplirán con una validez de tipo probabilístico, y con el criterio de analiticidad.

que no hay razón para que las palabras mentalistas figuren en los libros de lógica, dado que se puede explicar todo en términos de las proposiciones afirmadas y de los hechos aducidos en su apoyo. Y es que, recordemos, la intención de Toulmin no era presentar un trabajo sobre credibilidades, persuasión o comunicación, sino sobre epistemología.

4.2.5. El legado de la obra de Toulmin

En el terreno epistemológico de mitades del siglo XX, su libro fue o bien ignorado, o bien duramente criticado por algunos filósofos analíticos anglosajones como Strawson o Alexander, quien lo llamó “el libro antilógico de Toulmin”. Sin embargo, fue ampliamente adoptado por los estudiosos de la comunicación en Estados Unidos, quienes acuñaron la expresión de “el modelo de Toulmin”.

Así pues, como podemos ver, tanto Perelman como Toulmin parten de una reacción al logicismo formalista, señalando que impone un marco extremadamente limitante, eliminando así cualquier valor práctico que pudiera tener la lógica en la vida cotidiana, convirtiéndola en algo inaplicable para campos como la ética, la jurisprudencia y la acción. Para el abordaje de estos campos, adoptan un modelo de razonamiento informal, falible, plausible, que atiende al contexto y al contenido, apartado así del ideal matemático formal. Desde luego, esto supuso el punto de partida para las disciplinas posteriores que conformarían los estudios sobre argumentación.

En cualquier caso, la lógica formal continuó avanzando su investigación en estas cuestiones. Claros ejemplos han sido los avances de la lógica deóntica que parten de von Wright (1951), o la lógica dialógica de Lorenzen y Lorenz (1978). Esta última particularmente exitosa en la formalización dialéctica, inspirando varios modelos en teoría de la argumentación, como los de Barth y Krabbe (1982), Walton y Krabbe (1995), de la pragma-dialéctica (van Eemeren y Grootendorst, 1984, 1992), así como de razonamiento legal implementados en inteligencias artificiales (Bench-Capon, 1998; Prakken, 2005).

4.3. Después de los 50 del s. XX: consolidación y desarrollo de los estudios de argumentación

Desde la década de 1950, los estudios de argumentación han ganado importancia, destacando las contribuciones de Perelman y Toulmin, quienes criticaron la aplicación de la lógica formal al lenguaje natural y propusieron visiones más pragmáticas y retóricas. Esta perspectiva fue enriquecida por el clima social de los años 60 en Estados Unidos, marcado por protestas y la demanda de una educación relevante a los problemas cotidianos, lo que fomentó el interés en el discurso efectivo y llevó al desarrollo de disciplinas como el Pensamiento Crítico y la Lógica Informal.

4.3.1. Pensamiento crítico

El Pensamiento Crítico, con raíces en la Grecia clásica y formalizado en el concepto de "pensamiento reflexivo" de John Dewey (1910/1933), ha evolucionado como un movimiento educativo centrado en desarrollar habilidades discursivas y críticas en los estudiantes. Este movimiento se reforzó con la adopción de metas pedagógicas en las escuelas progresistas estadounidenses en los años treinta

y recibió un impulso significativo en la segunda mitad del siglo XX. Incluido formalmente en la educación de California desde 1980, el Pensamiento Crítico se ha establecido firmemente en la enseñanza y evaluación de habilidades analíticas y metacognitivas.

Esta evolución ha sido apoyada por numerosos estudios dentro de la psicología cognitiva y del discurso, destacando las investigaciones de Kuhn (1991), West, Toplak, y Stanovich (2008), y Goldman (2018). Hoy, el Pensamiento Crítico sigue siendo un área vital tanto en el ámbito educativo práctico como en la investigación, subrayando su relevancia persistente en la formación de ciudadanos capaces de razonar y argumentar efectivamente en situaciones de la vida real.

4.3.2. Lógica informal

La Lógica Informal emergió a finales de los años 60 como un movimiento académico liderado por R. Johnson y A. Blair, que se distanció de la enseñanza de la lógica formal simbólica al abordar el lenguaje natural y el razonamiento discursivo utilizado en contextos variados como el político, publicitario, y legal, entre otros. Inspirados en las obras de Toulmin, Perelman y la lógica no matemática de Kahane (1971), la Lógica Informal se propone mejorar la enseñanza del razonamiento crítico a través de una perspectiva normativa que prioriza la relevancia, suficiencia y aceptabilidad de las premisas más que su estructura formal.

Este punto de vista considera la argumentación como un acto comunicativo destinado a justificar una tesis de manera que sea aceptable para el oyente, utilizando inferencias que, si bien no son necesariamente deductivas, son legítimas (Blair y Johnson, 2000). Además, este campo se ha dedicado al estudio de falacias tanto formales como conversacionales y ha expandido su análisis a tipos de inferencias como las inductivas, presuntivas, abductivas y analógicas.

El trabajo de la lógica informal también implica un tratamiento pragmático de los argumentos para adaptar el análisis, la construcción y la evaluación de los mismos a la práctica discursiva real, integrando teorías pragmáticas de figuras como Austin (1962) y Grice (1969, 1975). Este punto de vista fue fundamentado y expandido en la obra de Robert Fogelin, *Understanding Arguments. An Introduction to Informal Logic* (1978), que establece un puente entre la teoría de los actos de habla y la argumentación informal. La lógica informal sigue siendo una disciplina vibrante que busca adaptar los principios lógicos a las necesidades de la argumentación real y cotidiana.

4.3.3. Argumentación dialéctica y pragma-dialéctica contemporáneas

Desde 1958, ha habido una revitalización de la dialéctica y retórica clásicas, culminando en desarrollos significativos en la teoría de la argumentación. Algunos autores contemporáneos consideran que los argumentos corrientes son dialécticos, refutables y sometidos a excepciones, que se pueden modelar como respuestas a preguntas. Ese diálogo de respuestas y preguntas constituye un texto argumentativo (Freeman, 1991). En el proceso de construcción de buenos argumentos, el hablante debe tener en cuenta la anticipación de posibles objeciones y los posibles argumentos alternativos (Blair y Johnson, 1987).

Esto ha llevado a buscar una integración de esta perspectiva con ciertos enfoques formales como la *lógica dialógica* de Lorenzen y Lorenz (1978), procedente de la lógica matemática, consistente en probar la validez de una tesis mediante mecanismos dialógicos puramente sintácticos. También ha habido una aproximación a la *dialéctica formal* de Hamblin (1970), para quien “la descripción de casos reales debe orientarse a señalar características formalizables, mientras que los sistemas formales deben orientarse a arrojar luz sobre los fenómenos descriptivos reales” (p. 256). Para el autor, el modelo argumentativo es el de las reglas que gobiernan los intercambios en la discusión filosófica. En estos contextos se pueden dar muchas falacias, pero la mayoría no serán formales, sino conversacionales. La validez no es lo relevante en estas lógicas formales, sino la satisfacción de los compromisos que adquiere el hablante al afirmar una aserción y al aceptar unas reglas de inferencia y no otras. En este sentido, la formalidad de la bondad argumentativa le viene dada por la necesidad de que el conjunto de compromisos adoptados por cada parte de la discusión constituya un sistema consistente.

La teoría *pragma-dialéctica* de van Eemeren y Grootendorst (1984, 1992) ha evolucionado a partir de estas bases, diferenciándose por su orientación pragmática. Prioriza la cooperación, la relevancia y la honestidad en las discusiones, buscando resolver diferencias de opinión a través de procedimientos dialógicos que evitan falacias y promueven una discusión crítica efectiva. Inspirada también por teorías de actos de habla y lógica conversacional, la pragma-dialéctica intenta tanto resolver conflictos como también asegurar que el discurso se maneje de manera que contribuya significativamente a la comunicación efectiva en diversos contextos.

4.3.4. Argumentación retórica contemporánea

La argumentación retórica contemporánea ha experimentado una significativa rehabilitación y consolidación como un componente esencial en el estudio de la argumentación, reconociendo que tanto las perspectivas lógicas como dialécticas están profundamente imbuidas de conceptos retóricos clásicos. Bermejo-Luque (2006) destaca que la argumentación sirve principalmente para persuadir, subrayando su naturaleza como fenómeno comunicativo que busca la aceptación de una tesis por parte del oyente, lo que inevitablemente lleva a la persuasión o convicción si es exitosa.

La perspectiva retórica se centra en la audiencia, buscando entender y utilizar sus emociones, valores, y creencias para construir argumentos efectivos. Esta orientación hacia la audiencia es crucial porque establece las premisas aceptadas como el punto de partida para la argumentación (Perelman y Olbrechts-Tyteca, 1958). Además, la evolución de la retórica dentro de los estudios de la argumentación ha sido sistemática, analizando cómo las características de la audiencia, el mensaje y el emisor se interrelacionan para formar argumentos convincentes. Kopperschmidt (1977) argumenta que la retórica ocupa un lugar privilegiado en este campo, dado su enfoque en cómo la representación adecuada de la audiencia y la estructuración lógica del argumento pueden facilitar la aceptación de las conclusiones. Este resurgimiento de la retórica está acompañado por un reconocimiento de sus propiedades lógicas, ya identificadas por autores como Gohlke (1943) y Solmsen (1924), y más recientemente, por Russo (1962), Grimaldi (1972, 1980) y Sprute (1977,

1982), quienes destacan que la *Retórica* de Aristóteles es fundamentalmente una contribución a la lógica especial de la argumentación. El interés contemporáneo en la retórica también incluye una revalorización de sus aspectos lingüísticos y preceptivos llevada a cabo por Barthes (1974) y Ricoeur (1975), y una integración de la dialéctica con la ética y la política como sostienen Johnstone (1980) y Arnhart (1981).

Así pues, la argumentación retórica es vista no solo como una herramienta para la persuasión sino también como un marco integral para entender y aplicar la lógica y la dialéctica en el discurso contemporáneo, lo que es esencial para resolver conflictos sociales según autores como Reboul (1990; Reboul y Johnstone, 1988) y Kienpointner (2017).

5. Rehabilitación de conceptos clásicos: los tópicos, el éndoxa y el entimema

En esta revitalización de la retórica, no solamente se ha puesto de manifiesto el valor ineludible que tiene esta disciplina para los estudios de argumentación, sino que también se ha retomado el interés por el estudio de viejos conceptos específicos que estaban presentes en la retórica clásica, presentados ahora a la luz de los avances acaecidos en los ámbitos de la lógica, la pragmática y el discurso. Estos conceptos serán relevantes en nuestro modelo teórico.

5.1. Retomando el concepto de ‘tópicos’ para los ‘esquemas de argumentación’

Brevemente, los *esquemas de argumentación* son formas estereotipadas de argumentación que se encuentran habitualmente en el discurso cotidiano, y están constituidas por reglas generales de relación entre objetos del mundo con las que la mente opera fácilmente al conectar ideas. Estas reglas están gobernadas por criterios como la causalidad, la contrariedad, la semejanza y muchos más. Los criterios que gobiernan estas reglas de relación reciben el nombre de *tópicos*. En base a estas relaciones, los hablantes pueden construir pruebas y justificaciones en la confección de sus argumentos. En efecto, el establecimiento de estas relaciones permite la transición de propiedades de algo conocido y/o aceptado hacia algo desconocido y/o no aceptado. De este modo, el tópico va dirigido a formar suposiciones creíbles acerca de nuevas relaciones.

Por ejemplo, mediante el esquema de argumentación que opera sobre el tópico ‘de la semejanza’ podemos establecer que, si x tiene la propiedad α , e y se parece a x , entonces, probablemente y tiene también la propiedad α . Como se puede observar, estas reglas son *esquemáticas*, en el sentido de que el criterio que rige al tópico relaciona variables (x , y , α , etc.) que deben ser rellenadas por los objetos, eventos y propiedades del mundo que convenga en cada caso. Son como celdas o habitáculos lógicos vacíos regidos por un criterio de conexión que permite la transición de propiedades de un objeto a otro (Racionero, 1990). Por poner un ejemplo muy simple, aún con el criterio de semejanza, si de un “kea” sólo sabemos que es un pájaro parecido a un loro, entonces podremos suponer que tiene el pico curvo y grueso. Este esquema de la semejanza (como todos los demás) es cuasi formal, se puede aplicar a cualquier tema de discusión, lo mismo a

cuestiones de ética que de física (ingenua) siempre y cuando se circunscriban al ámbito de las creencias en el razonamiento cotidiano. Esto es, al terreno del conocimiento incierto. Son reglas de relación falibles, rebatibles (de hecho, el pico real del kea no es grueso, sino fino). Guardan cierta forma de implicación, pero es una implicación con cierto componente sustancial, no puramente formal, por versar sobre los objetos del mundo, y por estar basada en los conocimientos de los usuarios del lenguaje. Esto es, tienen un valor probabilístico y derrotable.

En la recuperación de estos conceptos para la teoría de la argumentación, empezando con el desarrollo de Perelman y Olbrechts-Tyteca (1958), son destacables los trabajos de Kienpointner (1986), Warnick (2000), Braet (2004, 2005), Rubinelli (2006), Rigotti (2009), Macagno, Walton y Reed (2017) y, en cierto sentido, Tindale (2007). Todos ellos han articulado el concepto aristotélico de los *lugares comunes*, o *tópicos* (tópica mayor y menor) como precursores inmediatos de los, hoy llamados, *esquemas de argumentación*. Este concepto se desarrolla ampliamente en el ap. C dedicado a los entimemas.

5.2. Retomando el concepto de éndoxa

El concepto clásico de 'éndoxa' está recibiendo renovada atención como marco de conocimiento común en comunidades epistémicas específicas, crucial para inferencias y presuposiciones. Tardini (2005) lo define como un marco de conocimiento compartido, mientras que Walton (2021; Walton y Macagno, 2006) lo equipara al '*common knowledge*' (Minsky, 1974; Schank y Abelson, 1977) usado en inteligencia artificial para el razonamiento de sentido común.

Rescher (1976; 2013) emplea el concepto de 'éndoxa' para desarrollar un sistema de razonamiento plausibilístico para inferencias rebatibles derivadas de datos parcialmente inconsistentes. Aristóteles define lo plausible como lo que es aceptado por todos o la mayoría, especialmente entre los expertos. Para Rescher, la propuesta de Aristóteles ofrece una gradación acerca de la fiabilidad de la fuente comunicativa sobre la cual poder valorar la plausibilidad de las proposiciones. De este modo, en el sistema de Rescher se emplean valores ordinales cualitativos para establecer comparaciones de plausibilidad sin recurrir a una métrica cuantitativa (Gaviria y Jiménez-Leal, 2014). De manera similar, Vega (1998) enfatiza el carácter gradual y comparativo del éndoxa, definiéndolo como lo que es generalmente aceptado o aceptable en una sociedad. Argumenta que las opiniones se valoran según un consenso, estableciendo que las más aceptadas poseen mayor plausibilidad y autoridad. Vega también subraya la función del éndoxa en la selección de premisas para construir argumentos dialécticos y retóricos, resaltando su importancia como punto de partida consensuado en la argumentación. Esto lo posiciona como una guía valiosa para orientar la investigación contemporánea sobre la argumentación plausible.

5.3. Retomando el concepto de entimema

Recientemente se ha redescubierto y restaurado el concepto clásico de entimema. Como se mencionó anteriormente (sección 1.2.), originalmente, el entimema era un tipo especial de silogismo que operaba retóricamente con premisas evaluadas por una audiencia como más o menos probables en función de sus propios conocimientos. A diferencia de los silogismos concluyentes, las

premisas del entimema no consisten en axiomas, teoremas o verdades plenas. Consisten en contenidos del mundo real, variable y subjetivamente incierto y, por lo tanto, están sujetas a excepciones. Al funcionar como premisas de un silogismo, su grado de aceptación influye directamente en la aceptación de la conclusión, lo cual es uno de los aspectos fundamentales de esta tesis.

Por otro lado, también es cierto que Aristóteles introduce la idea de que, las más de las veces, el entimema es un silogismo “a partir de pocas premisas, incluso menos de las que consta el silogismo de la primera figura. Porque si alguna de estas premisas es bien conocida, no hace falta enunciarla: el propio oyente la suple” (*Ret. I.2, 1357a, 17-19*). Pero, tal como señala Grimaldi (1972), esta característica sólo tiene la intención de simplificar el argumento para facilitar su comprensión evitando la sobrecarga de información redundante. La eficacia persuasiva no deriva esencialmente de su brevedad, sino de su base en probabilidades y signos reconocidos por la audiencia (Racionero, 1990). A pesar de esto, históricamente, se ha consolidado la noción de que el entimema es un silogismo truncado, una interpretación que comenzó con la retórica latina y se extendió a través de la escolástica hasta la modernidad, donde no se reconocía la lógica de lo probable o plausible.

Desde el último cuarto del siglo XX, la percepción de los entimemas ha comenzado a cambiar, como lo demuestran las aportaciones de Grimaldi y Racionero. Ellos argumentan que los entimemas funcionan como cualquier silogismo y que su supuesta incompletitud no es una característica definitoria. Además, Burnyeat (1994) introduce la idea del entimema como un *argumento rebatible*, un concepto que muchos teóricos de la argumentación han desarrollado posteriormente. Según Vega (2004), considerar los entimemas meramente como silogismos incompletos es trivializar su naturaleza. Lo distintivo de los entimemas es que *se basan en creencias del auditorio* para ser comunicativos y persuasivos. Por ese motivo, sus premisas suelen estar sujetas a la incertidumbre informativa de los participantes. Este tipo de argumentos no se ajustan estrictamente a la lógica deductiva o inductiva. Operan bajo unos criterios de *deducción laxa* (Dyck, 2002), que descansan sobre un tópico implementado en un esquema de argumentación, llegando a conclusiones tentativas. Son, en definitiva, argumentos plausibles y rebatibles. Vega también relaciona el concepto original del entimema con teorías modernas como las lógicas no-monótonas, las lógicas por defecto, las plausibilistas y la inteligencia artificial. Sin embargo, es sorprendente que no se utilicen los conceptos de probabilidad subjetiva, especialmente considerando que el concepto de entimema se originó en discusiones sobre el razonamiento probable.

Walton y Macagno (2009) consideran que la interpretación tradicional de los entimemas como silogismos incompletos va siendo cada vez menos plausible a medida que los esquemas de argumentación rebatibles que se están estudiando en la actualidad devienen herramientas cada vez más ampliamente aceptadas e indispensables para la argumentación lógica. Para estos autores, el entimema es, esencialmente, un tipo de razonamiento plausible basado en generalizaciones falibles (esquemas de argumentación), un razonamiento en condiciones de incertidumbre. Como se avanzó hace unas páginas, el concepto del entimema se encuentra plenamente desarrollado en el ap. C. En

cualquier caso, estas nociones desarrolladas a lo largo de este capítulo pueden ser suficientes para entender su función en esta tesis.

CAPÍTULO 2

Concepciones de la Probabilidad

Los objetivos de este capítulo son mostrar cómo la probabilidad se relaciona con la argumentación retórica y dialéctica, el razonamiento cotidiano y, especialmente, con la credibilidad. Para ello, adoptaré una perspectiva histórica que permita entender las distintas interpretaciones que este concepto ha recibido y cómo se han desarrollado.

Actualmente, existen múltiples interpretaciones del concepto de probabilidad, como la interpretación clásica, la lógico-inductiva, la objetiva-frecuentista, la de propensiones o la interpretación subjetivista. Desde una perspectiva conceptual, la probabilidad se reduce a dos tipos básicos: la probabilidad aleatoria y la probabilidad epistemológica (Santos del Cerro, 2002). La probabilidad aleatoria se aplica a eventos del mundo, como los juegos de azar o la meteorología, donde existe un universo finito de posibilidades definidas, pero con resultados que son o parecen indeterminados. La probabilidad epistemológica, por otro lado, se refiere al grado de confianza subjetivo sobre la ocurrencia de una situación que nos representamos, o la veracidad que otorgamos a una proposición, siempre en relación con el conocimiento disponible. Esta noción, más psicológica que fáctica, es la relevante en esta tesis. Ambas nociones de probabilidad tratan sobre la contingencia, la incertidumbre y la falta de control sobre los eventos. Aunque el azar y lo probable son conceptos antiguos, la teoría del cálculo de probabilidades no surgió hasta la segunda mitad del siglo XVII.

A continuación, se realizará un recorrido histórico de las nociones de probabilidad para entender cómo han derivado en estas interpretaciones y analizar su relación con la argumentación y las creencias.

1. Azar y determinismo en la antigüedad pre-clásica

Desde hace unos 40.000 años, la humanidad ha utilizado herramientas como el hueso astrágalo para juegos de azar y artes adivinatorias, evidenciando un interés temprano en el azar y la incertidumbre, vinculados a creencias mitológicas y a intentos de explicación del mundo (Mateos-Aparicio, 2002). En la antigua Grecia, la noción de un destino azaroso pero inevitable permeaba fábulas y mitos, reflejando una convivencia entre el azar y la necesidad como componentes del orden universal (De Echandía, 1995). Sin embargo, filósofos como Leucipo y Demócrito propusieron que todo en el universo, incluido el pensamiento humano, está determinado por leyes mecánicas naturales, y lo que consideramos suerte es simplemente el resultado de un entramado complejo de causas necesarias, demasiado numerosas para ser comprendidas completamente por el ser humano (Russell, 1945/2005). Criticaban la idea de que el orden pueda surgir del caos, argumentando que cada efecto tiene una causa específica y que lo que parece casualidad no es más que un desconocimiento de las verdaderas causas: “los hombres han inventado la imagen de la *Tyché* (suerte) como excusa para su propia irreflexión” (Demócrito, DK 68 A 1, 19).

2. Perspectiva Aristotélica sobre la causalidad, incertidumbre y probabilidad

Aristóteles comparte con los atomistas la noción de que el azar no es causa de nada, sino que los eventos aparentemente azarosos surgen de un entramado de causas específicas que producen unos efectos determinados por ellas (*Fís.* II 5, 197a). Al contrario de los atomistas, que atribuyen la diversidad de fenómenos a diferencias cuantitativas entre los átomos, Aristóteles adopta un punto de vista cualitativo, argumentando que cada entidad se define por características intrínsecas que determinan su comportamiento natural. Por ejemplo, un peral produce peras, no olivas, debido a su constitución específica, que tiene la potencia de producir peras. Este proceso de cambio, de lo que una entidad es en potencia a lo que *puede ser* en acto, subraya que lo *posible* tiene una realidad ontológica en su sistema, por el cual una entidad puede adquirir múltiples formas mutuamente excluyentes. Los cambios derivados de la naturaleza de los objetos pueden ocurrir siempre y de forma necesaria, o de forma contingente la mayoría de las veces. Luego, en la realidad habrá cosas que no ocurren ni siempre ni la mayoría de las veces: lo accidental.

Al igual que los atomistas, Aristóteles también explora el concepto del azar desde una perspectiva epistemológica. Distingue entre eventos necesarios, que ocurren invariablemente debido a causas suficientes, y eventos contingentes, que no son ni imposibles ni necesarios y pueden ocurrir de varias maneras (*Met.* V 5, 1051a37; VI 2, 1027a10). Los eventos accidentales, que no siguen patrones predecibles ni leyes naturales regulares, surgen de la coincidencia de múltiples cadenas causales independientes (*Fís.* II 4), y son, por tanto, impredecibles. Estos eventos no están gobernados por potencias productoras que permitan predicciones generales, lo que los sitúa fuera del alcance del conocimiento científico sistemático (episteme), ya que cualquier análisis causal es solo aplicable al *caso* específico, es a posteriori y no es generalizable.

El azar se entiende, así, como una categoría que captura nuestra incapacidad para discernir todas las interacciones en eventos complejos, indicando que los eventos *accidentales* son comprensibles a través de sus causas físicas específicas, pero estos análisis son post hoc y no predictivos. Esta visión refleja una comprensión de que nuestra interacción con la realidad es limitada por la incertidumbre y la complejidad inherente del mundo. En este sentido, para Aristóteles, el ámbito de lo probable comprende eventos que, aunque generalmente sabemos que suceden de una manera debido a su naturaleza, no siempre ocurren así debido a factores no previstos por el observador. En su *Retórica* (I 2.6, 1357a35), define lo probable como lo que *se sabe* que ocurre la mayoría de las veces pero que también podría ocurrir de otra manera, considerándolo un juicio subjetivo basado en la observación y las limitaciones del conocimiento humano. La naturaleza epistemológica de lo probable deriva de la intervención de lo accidental, que introduce la incertidumbre y hace que lo probable sea inherentemente incierto.

Franklin (2001) resalta que, aunque no cuantitativo, el tratamiento de Aristóteles del azar y la probabilidad prefigura la teoría matemática de la probabilidad y ha influenciado el pensamiento filosófico y legal sobre el tema durante milenios. Esto subraya cómo los seres humanos categorizan y comprenden la incertidumbre y la probabilidad, enfatizando que nuestra relación con estas nociones está profundamente arraigada en nuestras limitaciones epistémicas.

3. De lo probable a lo plausible como cuerpo de enunciados argumentativos

Dicho esto, se vislumbra ya la relación que guarda la probabilidad con los razonamientos empleados por la dialéctica y la retórica. En efecto, estos razonamientos son acerca de lo probable en tanto que verosímil. De este modo, la dialéctica y la retórica se constituyen en las disciplinas que ofrecen las herramientas propias para realizar operaciones sobre las credibilidades, para modificar credibilidades en el curso de la comunicación cotidiana acerca de asuntos cotidianos. Pero todo ello sin olvidar que las credibilidades dependen ontológicamente de los sucesos del mundo real, ofreciendo así “el mínimo respaldo objetivo necesario para que la aceptación del objeto no sea un acto de ciega e irracional adhesión” (Candel-San Martín, 1982/2000). En este sentido, Aristóteles considera que “corresponde a una misma facultad reconocer lo verdadero y lo verosímil y, por lo demás, los hombres tienden por naturaleza de un modo suficiente a la verdad” (*Ret.* I 1.4, 1355a15). Y supuesto esto, en orden a la construcción de argumentos dialécticos y retóricos, el valor de lo verosímil residirá en que se constituya en premisas *plausibles* (*éndoxon*), donde lo plausible es “lo que *parece* bien a todos, o a la mayoría, o a los sabios, y, entre estos últimos, a todos, o a la mayoría, o a los más conocidos y reputados” (*Tóp.* I 1, 100b23). Lo plausible es, por tanto, el conjunto de opiniones conocidas y aceptadas por una comunidad epistémica concreta. Esto posibilita que lo plausible sirva a modo de base de datos donde ir a buscar y seleccionar enunciados que sirvan como premisas en los argumentos retóricos y dialécticos. De esta manera se introduce el contexto de las creencias aceptadas por la audiencia en tales argumentos.

Así pues, y a modo de corolario, “no es preciso exigir que los razonamientos de todos los problemas sean igualmente plausibles y convincentes: pues ya por naturaleza se da que, de las cosas investigadas, algunas son más fáciles y otras más difíciles, de modo que, si se prueba a partir de las cosas más plausibles que cabe, se ha discutido bien” (*Top.* VIII 11, 161b34), considerando como más difíciles las cuestiones referentes a la acción, la ética y la jurisprudencia por ser disciplinas menos exactas.

3.1. Límites y fundamentos de la acción retórica

En la medida en que no cabe deliberar sobre cualquier cosa, sino sólo sobre lo que puede llegar a suceder, la función retórica de lo posible es, en palabras de Racionero (1990), la de servir como un *mecanismo de control* lógico⁷ y ontológico que delimita el margen de maniobra en el que se van a mover las operaciones de credibilidad. Así, será susceptible de modificación la credibilidad acerca de aquello que se halle contenido en el marco de la posibilidad contingente. De este modo, tanto lo imposible como lo necesario actúan de límites naturales para estas operaciones en los razonamientos cotidianos.

Y, aun dentro de lo contingente, el límite estaría en la región de la posibilidad al margen de la inteligencia: no se puede ofrecer ninguna razón que haga probable que, ‘si tiro una moneda, saldrá cara y no cruz’⁸. Por todo ello, es en el ámbito de la posibilidad contingente susceptible de razonamiento donde ni la afirmación ni la negación son, en términos absolutos, en nada más verdaderas; e incluso podría ser que una de ambas fuese más probable, pero cabe la posibilidad de que se dé también la otra. Y es que el cambio de credibilidades se dará a partir de las *consideraciones* acerca de lo que es posible que sea y de lo que es posible que no sea, y eso para los tres tiempos de interés: a través del posible, el tiempo pasado y el tiempo futuro se transforman en un tiempo *hipotético*. Así, es el pasado o el futuro posible el que interesa a la retórica, de tal modo que la cuestión acerca de ‘si ha sucedido’, ‘si sucede’ y ‘si sucederá’ dependen de la cuestión de ‘si es posible’. De este modo, dirá Racionero (1990), el marco de lo posible no sólo limita la acción retórica, sino que también sirve de fundamento de posibilidad de sus enunciados; y, en consecuencia, del cambio de credibilidades.

3.2. Mecanismos e instrumentos de los juicios de probabilidad

Si lo posible es el marco, será la *remisión a los hechos* ciertos, aceptados o plausibles, lo que permitirá desencadenar el desplazamiento mental mencionado hacia esos tiempos hipotéticos

⁷ Las cuestiones referidas a los *modos* de ser (imposible, posible y necesario) tienen tanto correlatos lógicos como ontológicos. La *lógica modal* se encarga del análisis en su vertiente lógica. Aristóteles se ocupa de ello en *De Inter.* 9, 12 y 13; y Hintikka (1962), Putnam (1975) y, sobre todo, Kripke (1980) se han ocupado profusamente de este tema. Por otro lado, llama la atención el aumento reciente en el número de trabajos en inteligencia artificial encaminados a la formalización la retórica y la argumentación en términos de la lógica modal de los mundos posibles (e.g. De Baere, 2003; Dung, 1995; Grasso, 2003), así como, en particular, el trabajo realizado por Budzynska y Kacprzak (2008) en la formalización de la Retórica de Aristóteles en estos términos, y de ahí, a la formalización de su teoría de las probabilidades.

⁸ A decir verdad, sí que se puede. Pero la explicación requiere una enumeración muy elevada de variables y de sus interacciones. El ejemplo sólo tiene la función de ilustrar intuitivamente lo que quiere decir el autor.

posibles en aras a juzgar la probabilidad de aquello que está en cuestión. En este sentido, los hechos o datos del mundo equivalen a los signos (o indicios) que señalan (o apuntan hacia) la posible verdad de la existencia de dicha cuestión. La verosimilitud del hecho establecido o aducido sirve, entonces, de base pretendidamente objetiva a partir de la cual evaluar la probabilidad del objeto bajo consideración: ¿cuán probable resulta A, dado B? De esta manera, la remisión a los hechos (B) tendrá un rol fundamental como *fuentes original de la credibilidad*, otorgándole validez a la probabilidad que vaya a serle asignada a la posible ocurrencia del objeto inicialmente controvertido (A). En cualquier caso, esa probabilidad variará tanto en función de la verosimilitud del propio hecho propuesto, como del grado de *relevancia* de la relación que la audiencia pueda establecer entre el hecho y aquello que se pretende probar a partir del hecho.

Las pruebas que se construyen sobre la base de los hechos para juzgar la probabilidad de situaciones en el marco de lo posible son el entimema y el ejemplo. Éstos constituyen los instrumentos lógico-formales propios de la retórica y la dialéctica, representando, el entimema, un razonamiento análogo a la deducción, y el ejemplo, un razonamiento análogo a la inducción. Estos instrumentos conforman las reglas mecánicas por las cuales operan los cómputos de probabilidad que se ponen en marcha en el cambio de credibilidades por vía de la argumentación⁹.

4. Probabilismo, escepticismo y falibilismo en el mundo greco-romano

El probabilismo de Carnéades en el período helénico se fundamenta en la perspectiva de que el conocimiento surge de la observación sensorial, y es inherentemente falible debido a la naturaleza imperfecta de las impresiones sensoriales y cognitivas. Carnéades de Cyrene, reaccionando al dogmatismo de los estoicos, argumenta que las impresiones adquiridas, incluso bajo condiciones óptimas de observación, pueden ser erróneas y no siempre se puede distinguir entre las impresiones correctas y las incorrectas (Franklin, 2001). Esto lleva a la conclusión de que la verdad es esencialmente inaprehensible y que lo máximo a lo que podemos aspirar son opiniones más o menos verosímiles, probables y persuasivas.

En su escepticismo, Carnéades desarrolla un sistema para evaluar la verosimilitud de las opiniones basado en criterios subjetivos que incluyen la claridad del juicio (i.e., que esté bien definido), la adecuación de las condiciones de observación en las que surge el juicio, su resistencia los refutadores siguiendo un proceder dialéctico y la coherencia con el sistema de creencias existente en el individuo o la comunidad. Esta perspectiva permite clasificar las opiniones según su grado de fiabilidad, desde improbables hasta probablemente ciertas, adaptando la confianza en una opinión según su resistencia a evaluaciones críticas. Así pues, es el valor resultante asignado al juicio u opinión lo que permite *adherirse* a tal representación, y no su carácter de supuesta verdad universal sin excepción (Franklin, 2001). Carnéades también propone dos posturas epistemológicas

⁹ Tendremos ocasión de examinar en detalle las características propias de los entimemas debido a su rol fundamental en la presente investigación (ver ap. C).

frente a la incertidumbre: la suspensión del juicio para evitar errores, o la aceptación tentativa de opiniones sujetas a revisión constante ante nuevas evidencias, reforzando la noción de que nuestro conocimiento es meramente probable y no absoluto (*Acad.* 2.32, 110). Esta orientación refleja la adaptabilidad de nuestros conocimientos a la realidad circundante, permitiendo así la acción racional situada dentro de los límites de lo probable.

En el mundo romano, la noción de probabilismo y falibilismo se desarrolló significativamente entre pensadores como Filón de Larisa, Cicerón y Boecio. Filón sostenía que, aunque es posible alcanzar una comprensión verdadera de la realidad, nunca podemos estar seguros de haberlo logrado debido a las limitaciones inherentes de la razón humana. Esto conduce a la adopción de una perspectiva basada en la probabilidad subjetiva, donde el conocimiento es siempre tentativo y sujeto a revisión continua (Long, 1987). Filón argumentaba que nuestras representaciones cognitivas pueden ser más o menos convincentes y persuasivas, pero pueden no coincidir fielmente con la realidad. Por lo tanto, solo podemos asignarles una probabilidad de veracidad, que se evalúa en función de su coherencia con el resto de las representaciones aceptadas y las condiciones bajo las cuales se formaron. En este sentido se dice que Filón sostiene un probabilismo falibilista (Brittain, 2006) que fue formalmente desarrollado por Charles Sanders Peirce mucho tiempo después.

Cicerón, discípulo de Filón, también adoptó esta perspectiva falibilista, reconociendo que, pese a la existencia de la verdad, solo podemos aproximarnos a ella mediante juicios de probabilidad basados en la experiencia y el consenso. Para Cicerón, lo probable significa el *curso habitual de los eventos, de lo que la mayoría de las veces se cumple y de las creencias comunes*. Schuessler (2015) explica que la interpretación de Cicerón sobre la probabilidad se divide en dos enfoques distintos: el primero, traduciendo el término griego *pithanon* (persuasivo) como *probabilis*, refiere a una probabilidad 'endóxica' que se aplica a las acciones aprobadas por la opinión común o experta. El segundo enfoque traduce *eikos* (probable o verosímil) como *verisimilis* (similar a la verdad), que se basa en la frecuencia con que dos eventos ocurren juntos, interpretando uno como indicativo de la presunta ocurrencia del otro. De este modo conecta las acciones con creencias aceptadas y el conocimiento con la interpretación de la frecuencia de eventos. Boecio continuó esta tradición enfatizando que la comprensión humana está limitada por nuestra incapacidad para abarcar la totalidad de la experiencia. Propuso que la probabilidad se determina por el grado de consenso entre expertos o la comunidad en general, haciendo de la probabilidad una guía crucial para la interpretación y la acción basada en el conocimiento parcial y la opinión probable (Schuessler, 2015; Secades, 2005).

Estos pensadores compartían la idea de que, si bien la probabilidad se asimilaba al grado de credibilidad que se le podía asignar a las proposiciones de conocimiento, no tiene un límite superior dado que nunca se llegaba a la certeza sobre cómo son las cosas. Esta orientación falibilista y probabilista enfatiza que, aunque la verdad definitiva es inalcanzable, podemos operar racionalmente dentro de los límites de lo probable y lo verosímil, adoptando un punto de vista práctico y flexible hacia el conocimiento y la acción.

5. Evolución del pensamiento probabilístico: desde la Edad Media hasta el inicio de la Edad Moderna

Durante la Edad Media, los conceptos de probabilidad, heredados de pensadores como Cicerón y Boecio, jugaron un papel crucial en diversas áreas del conocimiento y la práctica cotidiana. Los medievales utilizaban términos latinos como *probabilis*, *credibilis*, y *verisimilis* para evaluar afirmaciones y argumentos dentro de contextos académicos, legales y éticos, donde la certeza absoluta era inalcanzable (Schuessler, 2015).

Reconociendo que el azar no interviene en los procesos naturales, y que la observación directa de la naturaleza ofrece una imagen enmarañada, los medievales admitieron la necesidad de adoptar una visión probabilística. En este contexto epistemológico, la probabilidad se atribuye a las leyes naturales se entendía desde una perspectiva frecuentista de ‘lo que pasa la mayoría de las veces’. Boecio de Dacia la concibe como *el grado de participación condicionada de un predicado (P) en un sujeto (S)*: habiéndose dado S, se dará P a condición de que no se den accidentes inhibidores. S contiene una propiedad que le predispone a tener P (Schuessler, 2015). Por lo que, ‘si S, entonces probablemente P’.

Además, en la vida práctica, el conocimiento probabilístico permitía manejar la incertidumbre y el riesgo, particularmente en las finanzas, la ley y la ética. En el derecho, por ejemplo, se valoraba la coherencia entre testimonios y los indicios para establecer la probabilidad de las presunciones. La dialéctica del interrogatorio cobró gran importancia para determinar las contradicciones y fijar la probabilidad de un conjunto de afirmaciones. Y en la ética, las decisiones se basaban en opiniones probablemente ciertas, pero admitiendo la posibilidad de error. Conciben que la conducta humana está influenciada por el libre albedrío y que, por ello, es intrínsecamente incierta y sujeta a juicios de probabilidad. Los debates escolásticos aceptaban la legitimidad de opiniones opuestas e igualmente probables, siempre que estuvieran respaldadas por diferentes grupos de expertos igualmente creíbles. A partir de ahí, procedían razonando sobre premisas plausibles (i.e., con probabilidad endóxica) para resolver estas cuestiones, y cuyo razonamiento derivaba en conclusiones cuya probabilidad nunca podía rebasar la de sus premisas. Esto último enlaza con algunas consecuencias del modelo presentado en esta tesis, que será contrastado con otros procedentes de una perspectiva plausibilista del razonamiento y la argumentación.

La mayoría de las características de la dialéctica plausibilista medieval, como son la necesidad de coherencia, la sistematización jerárquica del cuerpo de enunciados, el peso relativo de las fuentes de opinión, la posibilidad de coexistencia de opiniones altamente probables pero incompatibles, o la posibilidad de otorgar un peso ordinal para la comparación de opiniones han recibido una atención renovada en los trabajos de Rescher (1976, 1977) y Walton (1992, 2008), teniendo una importancia notable en sus teorías formales del llamado razonamiento plausible.

El probabilismo del siglo XVI marcó una evolución significativa en la teoría de la probabilidad, adaptando las nociones antiguas y medievales a un contexto de creciente complejidad en decisiones morales y jurídicas (Santos del Cerro, 2002). Esta escuela de pensamiento reconocía que, si bien algunos fenómenos naturales podrían entenderse bajo un modelo de determinismo (a veces laxo), la conducta humana, influida por el libre albedrío y variables contextuales, escapa a certezas absolutas. Así, algunos juristas y filósofos, como Luis de Molina, promovieron una visión práctica que valoraba las leyes y decisiones basadas en la probabilidad derivada de la experiencia, el consenso y la autoridad, reconociendo la falibilidad inherente de cualquier conclusión. Esta perspectiva no solo ayudó a manejar la incertidumbre en las decisiones cotidianas, sino que también legitimó la adaptabilidad y revisión continua de las creencias a la luz de nuevas evidencias y circunstancias. El hecho a valorar está incrustado en una compleja maraña casuística de condicionantes irrepetibles y accidentales que lo podrían justificar. Esta concepción derivó en la crítica de que, finalmente, conduce al relativismo.

6. Resumen del concepto pre-moderno de probabilidad

El concepto de probabilidad surge a partir de los conceptos modales de necesidad e imposibilidad, que delimitan la posibilidad contingente. Además, incluye una *noción frecuentista*, definida como la ocurrencia de un evento la mayoría de las veces respecto de un total *conocido*, y una *noción endoxástica* que valora la *aprobación* comunitaria o experta de estas frecuencias o ideas. Así, la probabilidad es entendida como una cualidad *epistémica*, aplicable no directamente a los sucesos, sino a las *representaciones* sobre estos sucesos contenidas en pensamientos o en proposiciones. La probabilidad tiene una función relacional y condicional entre el hecho observado, al que uno puede remitirse, y la presunción subjetiva de otra cosa indicada por la ocurrencia de ese hecho.

Además, se reconoce que algunas representaciones de la realidad son más probables que otras, especialmente las consideradas próximas a lo necesario, las que son ampliamente aceptadas por expertos, y las que son *coherentes* dentro de un *sistema de creencias* establecido. Esta conceptualización de la probabilidad como algo *gradual y ordinal* facilita la *comparación y el cálculo cualitativo* de la igualdad o desigualdad del peso de las razones, aplicando un método de cómputo racional a través de la controversia y la *argumentación dialéctica*. Esta conceptualización es particularmente útil en el manejo de la *incertidumbre* inherente a la toma de decisiones morales, la ética y la jurisprudencia, permitiendo una *adhesión* variable del sujeto al contenido de una representación y proporcionando un camino hacia la verdad (Schuessler, 2015; Gavira y Jiménez-Leal, 2014).

7. Cambio de paradigma en la filosofía de la naturaleza y de la ciencia: de la cualidad a la cantidad

El cambio de la Edad Media al período Moderno marcó un punto de inflexión en la percepción de la naturaleza, influenciado significativamente por el nominalismo británico de Escoto y Ockham y catalizado por transformaciones políticas y religiosas en Europa. Hasta ese momento, se concebía la naturaleza de un modo cualitativo, compuesta de sustancias esenciales con propiedades cambiantes, entendiendo la existencia y el comportamiento de las entidades a través de causas finales y una teleología intrínseca a los objetos.

Sin embargo, en los siglos XVI y XVII, emergió una orientación mecanicista que conceptualizaba la naturaleza como un universo continuo y homogéneo, regido uniformemente por leyes fundamentales y compuesto por partículas indistinguibles más allá de sus atributos cuantitativos. Esta visión desplazó el entendimiento organicista de la naturaleza, rechazando las causas finales a favor de causas externas eficientes y mecánicas, lo que alineó el estudio de la naturaleza con una perspectiva cuantitativa. Este resurgimiento del modelo atomista marcó el inicio de la ciencia moderna.

La consecuencia de esta nueva visión fue un paradigma en el que la realidad natural podría ser entendida y descrita matemáticamente, simplificando y clarificando la comprensión y predicción de fenómenos naturales mediante un método científico sistemático y objetivamente verificable. La ciencia moderna se caracterizada por la búsqueda de leyes universales descriptibles mediante cálculos matemáticos y por una aproximación empírica libre de interpretaciones subjetivas y axiológicas. Con ello, se introdujo una distinción clara entre la antigua probabilidad subjetiva y la nueva probabilidad matemática, marcando un hito en la historia de la ciencia y la filosofía natural.

8. Del probabilismo cualitativo a la geometría del azar

A partir de este momento, suceden dos desarrollos históricos fundamentales: la transición del probabilismo filosófico hacia una visión rigorista en la jurisprudencia, y el origen del cálculo de probabilidades, también conocido como 'geometría del azar'. El probabilismo cualitativo originariamente permitía múltiples interpretaciones de un hecho, basadas en el conocimiento limitado y las opiniones de expertos, dando lugar a una perspectiva cualitativa y subjetiva en la ética y la jurisprudencia. Esta metodología se volvió cada vez más incompatible con las concepciones emergentes de determinismo y mecanicismo científico, que exigían un análisis objetivo y matematizado de la realidad (Hacking, 1995; Gómez Camacho, 2002).

Pascal fue uno de los primeros en criticar la flexibilidad del probabilismo, abogando por una base axiomática para la jurisprudencia que eliminaba la interpretación subjetiva de las intenciones y enfatizaba una aplicación de la ley basada en principios matemáticos y deductivos (Pascal, Cartas Provinciales). Este cambio paradigmático llevó al declive del probabilismo, relegándolo a un mero

'juego retórico' en el discurso intelectual y sentando las bases para el desarrollo de la probabilidad como una ciencia matemática, fundada en principios objetivos como el de indiferencia, que asume la equiprobabilidad de eventos en un conjunto finito cuando no hay razones para priorizar uno sobre otro (Mateos-Aparicio, 2002; Santos del Cerro, 2002).

Paralelamente, el cálculo de probabilidades emergió de la necesidad de entender dispositivos de aleatorización en juegos de azar. Figuras como Cardano y Galileo sentaron las bases de la combinatoria y la teoría de probabilidades durante el Renacimiento. La correspondencia entre Pascal y Fermat sobre el 'problema de los puntos' impulsó el campo formalmente, aplicando métodos matemáticos a decisiones bajo incertidumbre y estableciendo conceptos como la ganancia esperada y la utilidad (García Secades, 2002). Pascal articuló esta concepción como 'geometría del azar':

“Así, juntando el rigor de las demostraciones de la ciencia con la incertidumbre del azar, y conciliando ambas cosas en apariencia contrarias, puede, obteniendo su nombre de las dos, atribuirse con razón el asombroso título de ‘geometría del azar’” (Pascal, 1654; citado en Areán-Álvarez, 2012).

Este desarrollo no solo transformó la comprensión de la probabilidad de una práctica basada en la retórica a una centrada en cuantificación y objetividad, sino que también estableció una unidad de medida homogénea para evaluar eventos, crucial para el avance del análisis probabilístico en diversas disciplinas científicas y prácticas.

9. El desarrollo de la probabilidad epistémica en la Edad Moderna

En la Edad Moderna, el desarrollo de la probabilidad epistémica avanzó significativamente con las aportaciones de Christian Huygens y más tarde, de Gottfried Wilhelm Leibniz y Jacob Bernoulli. Huygens, a través de su obra *De Ratiociniis in Ludo Aleae* de 1657, estableció las bases de la probabilidad matemática y el concepto de esperanza matemática, influyendo profundamente en los trabajos posteriores en el campo. Leibniz se interesó por la aplicación de la probabilidad a la jurisprudencia y otros ámbitos de decisiones cotidianas, buscando una guía matemática para resolver disputas y clarificar la ambigüedad del lenguaje natural. Propuso un sistema donde los grados de probabilidad se expresasen numéricamente para reflejar grados de certeza, aunque no completó este desarrollo.

Jacob Bernoulli, influenciado tanto por Leibniz como por la Lógica de Port Royal, llevó la probabilidad más allá del cálculo en los juegos de azar, considerándola como un atributo de la opinión y la credibilidad. En su tratado *Ars Conjectandi*, Bernoulli exploró cómo los argumentos podrían cuantificarse en términos de probabilidad, abogando por una aplicación práctica de la probabilidad para guiar la toma de decisiones en la vida cotidiana. Su perspectiva no solo incluía el análisis matemático, sino también la argumentación, viendo la probabilidad como inherentemente

vinculada a la credibilidad y la fuerza de los argumentos (Girón, 1994). Este marco teórico proporcionó las bases para aplicaciones modernas de la probabilidad en campos tan variados como la inteligencia artificial y la teoría de la evidencia (Shafer, 1992).

9.1. De la inferencia probabilística deductiva a la inductiva: el problema de la probabilidad inversa

Antes de Bernoulli, la probabilidad se estructuraba en torno al principio de uniformidad e indiferencia, permitiendo inferencias probabilísticas *deductivas*, que iban de las causas a los efectos, y operaban bajo la premisa de que todas las posibles ocurrencias de un evento eran equiprobables, como en el caso de los juegos de azar. Esta orientación asumía una comprensión completa y precisa de todas las posibilidades involucradas (entendidas como causas), lo que facilitaba la aplicación de la regla de la razón de casos favorables entre posibles¹⁰. Este tipo de inferencia, de naturaleza deductiva, partía de causas conocidas para predecir efectos específicos, considerando que las condiciones iniciales conocidas y el principio de indiferencia eran las causas que determinaban de manera fija la probabilidad de un evento (Landro y González, 2013). Que el dado tenga seis lados junto con el principio de indiferencia es la causa de que la probabilidad de obtener un cuatro sea de 1/6. Esta es la conocida como probabilidad teórica o *a priori* de inferencia directa.

Sin embargo, la aplicabilidad limitada del principio de indiferencia a situaciones más allá de modelos idealizados como los dados y las monedas llevó al reconocimiento de sus limitaciones para capturar la complejidad de situaciones más generales o concretas y cotidianas, donde no siempre es posible definir o conocer todas las posibilidades. Esto es, donde hay falta de información. Esto marcó una transición hacia una perspectiva más empírica e inductiva, donde la inferencia se mueve de efectos observados a causas no observadas, ampliando el alcance de la probabilidad a situaciones menos controladas y más dependientes de la evidencia empírica acumulada (el caso paradigmático es el de conocer la composición de una urna de bolas de colores – como causa – a partir de las sucesivas extracciones – como efectos suyos).

Bernoulli fue pionero en este nuevo enfoque cuantitativo inverso, tratando de calcular la probabilidad de una proposición sobre una clase de eventos (causa hipotética) basándose en la observación repetida de estos eventos bajo condiciones consistentes. Esto se conceptualizó como probabilidad *a posteriori*, basada en datos observados y distinta de la probabilidad *a priori*, que se deduce sin observación empírica. Este enfoque trató de cuantificar la incertidumbre basada en la acumulación de evidencia empírica y trató de calcular la probabilidad de causas desconocidas a partir de efectos conocidos, como en el ejemplo mencionado de la composición de una urna (Landro y González, 2013). En este sentido, el incremento de la experiencia por observación disminuye el error de la inferencia y produce la generación de expectativas sobre el comportamiento futuro del

¹⁰ Más tarde, será conocida como “regla de Laplace”: a sacar un número cualquiera en un dado le correspondería una probabilidad de 1/6, puesto que es caras favorables / caras posibles que tiene el dado.

fenómeno¹¹. De este modo, llegamos a la expresión matemática de la racionalidad: en función de lo visto, esto es lo que debería creer un agente racional¹².

El problema con este método es que requería un número extremadamente alto de observaciones para alcanzar una certeza aceptable, como ilustró Bernoulli en sus cálculos. A pesar de la promesa de esta orientación para manejar la incertidumbre a partir de datos empíricos, las limitaciones prácticas de requerir un gran volumen de datos para realizar inferencias fiables fueron notables (Gómez-Villegas, 2001). Bernoulli y, posteriormente, De Moivre (1733) avanzaron en la formalización matemática de estas inferencias inductivas, reduciendo el número necesario de observaciones y aumentando la precisión de las estimaciones de probabilidad a partir de datos empíricos. El esfuerzo de estos matemáticos no solo amplió la aplicación de la probabilidad más allá de los contextos teóricos y a priori, sino que también estableció una base para la estadística inferencial moderna, marcando un paso significativo en la comprensión y el manejo de la incertidumbre en diversos campos del conocimiento y la decisión (Landro y González, 2013; Madrid Casado, 2012).

9.2. El Teorema de Bayes: la Probabilidad Subjetiva como Solución al Problema de la Inferencia Inductiva

Veamos ahora cómo el teorema de Bayes resuelve los desafíos planteados por la inferencia inductiva, conectando las probabilidades directas e inversas, y tratando las creencias y el conocimiento de los sujetos tanto a priori como a posteriori.

Tradicionalmente, las inferencias probabilísticas deductivas se basaban en el principio de uniformidad. Este principio permitía calcular la probabilidad de futuros efectos a partir de causas conocidas, como los lados de un dado. Se empleaba un método que distribuía la probabilidad total de manera uniforme entre todos los posibles resultados, gracias al principio de indiferencia. Este enfoque era aplicable en contextos controlados como los juegos de azar. Sin embargo, este método presentaba la limitación de no poder justificar adecuadamente que el comportamiento pasado de un evento observado pudiera servir como predictor del comportamiento futuro. Esto era

¹¹ Estamos haciendo referencia al llamado Teorema Áureo, o Teorema de Bernoulli, o Ley de los Grandes Números. Consiste en que la “certidumbre moral” acerca de una proposición podría aproximarse tanto como se quisiera aumentando el número de observaciones (Girón, 1994). Esto es, que a medida en que aumenta el número de observaciones *empíricas* del experimento aleatorio, la *frecuencia relativa* de un suceso tiende a aproximarse a un número fijo que se corresponde con la *probabilidad teórica* del suceso (Madrid Casado, 2012). También se le conoce como ley de la *estabilidad de la frecuencia*. Es decir, que cuantas más veces tiremos una moneda, la frecuencia relativa del suceso “cara” se irá acercando cada vez más al 50%. Y ese valor es el que le correspondería a nivel teórico dividiendo *favorables entre posibles*, siempre y cuando la *ley de la indiferencia* fuese correcta. Los trabajos en esta línea continuaron con Nicolás Bernoulli y De Moivre, tratando de saber cuántos ensayos son necesarios para poder decir “ya está, ya tenemos la probabilidad del evento”. Gracias a este último sabemos que la distribución binomial puede ser aproximada por la distribución normal.

¹² La versión actual de esta perspectiva es la medida de *calibración* propuesta por van Fraassen (1983) como justificación frecuentista de la probabilidad subjetiva. Brevemente, decimos que una función de credibilidad está realmente calibrada en un mundo posible en cuestión si la credibilidad que el sujeto asigna a una proposición dada coincide con la frecuencia relativa con la que las proposiciones de ese tipo son verdaderas en ese mundo conocido por el sujeto (Pettigrew, 2015).

especialmente evidente en contextos donde las causas no eran completamente conocidas o eran demasiado complejas para ser modeladas de esta manera. Para superar estas limitaciones, el teorema de Bayes ofrece un método para conectar la probabilidad de un efecto dado una causa conocida (probabilidad directa) y la probabilidad de una causa dada la observación de un efecto (probabilidad inversa). Este teorema transforma la comprensión de cómo los datos observados pueden influir en la evaluación de hipótesis teóricas o causas. En su forma más simplificada, el teorema de Bayes se expresa como:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Donde $P(A|B)$ es la probabilidad de la hipótesis A dado el dato B ; $P(B|A)$ es la probabilidad de observar el dato B dado que la hipótesis A es cierta; $P(A)$ es la probabilidad a priori de A ; y $P(B)$ es la probabilidad total de observar B .

Bayes introdujo la noción revolucionaria de que $P(C)$ (i.e., la probabilidad de una causa *antes* de cualquier observación) podría considerarse una variable aleatoria en lugar de una constante. Esto permitió asignar una distribución de probabilidad *a priori* basada en el conocimiento o las creencias del individuo. Esta perspectiva abre la posibilidad de ajustar nuestras creencias en respuesta a nueva información, recalculando la probabilidad de hipótesis o causas en base a datos observados.

Si se asume la aleatoriedad, entonces es lícito aceptar el principio de uniformidad de la naturaleza que implica la equiprobabilidad o ley de la indiferencia. Así pues, si se trata a $P(C)$ como una variable aleatoria, entonces su probabilidad se distribuirá uniformemente entre el resto de posibles causas en un abanico *finito* de éstas. Por ejemplo, si para un efecto E tenemos tres posibles causas C_1 , C_2 y C_3 , entonces la probabilidad *a priori* de cada una de esas causas sería de $1/3$ *antes de la observación de E*. Las múltiples causas posibles son equiprobables *si no sé nada más* acerca de ninguna de ellas. Si no sé nada más, no dispongo de ninguna *razón* para suponer que una debería ser más probable que otra¹³. Pero en caso de disponer de tal razón, entonces es lícito tenerla en cuenta, en lugar de recurrir a la equiprobabilidad. Así pues, para Bayes, la probabilidad *a priori* se *determina subjetivamente* en función de los *conocimientos* o las *creencias* de los individuos. Pero unas personas pueden tener conocimientos o creencias diferentes de otras personas. En este sentido, la subjetividad de $P(C)$ implica su variabilidad. Así pues, como acabamos de ver, desde esta perspectiva, no es una constante desconocida, sino una variable con una distribución de probabilidad conocida *a priori*.

¹³ Es importante señalar aquí, que Bayes está aceptando el principio de indiferencia, no como un principio ontológico que rija el funcionamiento de la naturaleza en la realidad, sino como un principio *epistemológico* que es aconsejable seguir en situaciones de desconocimiento por parte del sujeto individual. Se parte de tratar a la variable en cuestión como aleatoria a modo de premisa inicial, para finalmente llegar a una conclusión que excluye la aleatoriedad en el momento *a posteriori*.

Lo que resulta desconocido y se pretende calcular es la probabilidad de la causa *a posteriori*, después de la observación de un efecto. Por tanto, ahora se tratará de una probabilidad condicionada a la experiencia de observación de los datos. Se corresponde, pues, con $P(C | E)$. Así, la probabilidad $P(C)$, determinada *a priori*, se podría ir corrigiendo a medida que fuésemos observando nuevos datos, entendidos como efectos de la causa postulada. Este teorema permite calcular probabilidades finales a partir de establecer unas probabilidades iniciales y de unos datos obtenidos experimentalmente. Por último, la probabilidad total del efecto, o $P(E)$, consiste en la probabilidad de observar ese efecto si la causa postulada fuese cierta, más la probabilidad de observarlo debido a causas alternativas a la postulada, procedentes del abanico finito total de posibilidades. Es decir, que $P(E) = P(E | C) \times P(C) + P(E | \neg C) \times P(\neg C)$. Y puesto que $P(E)$ incluye probabilidades *a priori* (i.e. $P(C)$ y $P(\neg C)$), y estas pueden variar entre individuos, queda claro que $P(E)$ también puede variar entre individuos.

El Teorema de Bayes ha adquirido gran importancia en diversas disciplinas, especialmente debido a que enlaza la probabilidad de un efecto dada una causa con la probabilidad de una causa dado un efecto observado. Por ejemplo, es fundamental en contextos como la medicina, donde permite evaluar la probabilidad de una enfermedad a partir de síntomas observados, basándose en la probabilidad previamente conocida de esos síntomas dada la enfermedad y la prevalencia de la enfermedad. Este enfoque permite inferencias sobre causas a partir de efectos sin necesitar un número excesivo de observaciones. Se fundamenta en probabilidades subjetivas establecidas *a priori*, que se ajustan conforme se recopila nueva información. Este ajuste es crucial porque las probabilidades *a priori* afectan tanto al numerador como al denominador de la fórmula bayesiana, lo que permite que dos individuos, aun partiendo de la misma evidencia, puedan llegar a conclusiones probabilísticas diferentes *a posteriori*. Esto subraya cómo el teorema facilita la interpretación personal y muestra que las discrepancias sobre los resultados experimentales pueden derivarse de diferencias subjetivas en las probabilidades iniciales asignadas a las causas ($P(C)$). El establecimiento de estas probabilidades *a priori* debe ser racional, significando que la suma de las probabilidades para todas las causas posibles en un conjunto finito debe sumar uno. En contextos de desconocimiento absoluto, esto se logra a través de la equiprobabilidad. Sin embargo, en situaciones donde existen conocimientos o creencias específicas, estas influencias pueden justificar la asignación de probabilidades *a priori* no equiparables, manteniendo la coherencia racional en su asignación.

Richard Price, quien publicó póstumamente la obra de Bayes, le añadió un anexo en el que señaló la importancia que el Teorema de Bayes podría llegar a tener para el desarrollo de una teoría de la inducción capaz de explicar los procesos psicológicos subyacentes al razonamiento cotidiano de sentido común. Concibe la probabilidad *a posteriori* como una medida actualizable del conocimiento del sujeto sobre la veracidad de una proposición en función de la evidencia disponible. Así, las palabras de Price anticipan la relevancia del vínculo entre la probabilidad, el razonamiento y la credibilidad; un vínculo que, con altibajos, se había ido gestando desde la

antigüedad. En efecto, el supuesto de que las probabilidades bayesianas pueden describir adecuadamente los cálculos que realiza el razonamiento humano tiene hoy muchos adeptos.

Hoy en día, la relevancia del Teorema de Bayes tanto para la inferencia estadística y la toma de decisiones basada en evidencia como para la comprensión y modelado del razonamiento humano está fuera de discusión. En psicología, su importancia radica en que permite ajustes de creencias basados en nueva información, y explica cómo las diferencias en las creencias previas de las personas pueden llevar a distintas interpretaciones de la misma evidencia.

10. Evolución del pensamiento probabilístico: de la intuición a la formalización matemática

La perspectiva subjetivista continuó de la mano de Laplace, para quien “los problemas fundamentales de la vida no son más que problemas de probabilidades [...]. En el fondo, la teoría de la probabilidad es solo sentido común reducido a cálculo” (citado en Madrid-Casado, 2012, p. 131), o de De Morgan, quien, en 1838, declaró: “considero que la palabra probabilidad significa el estado mental con respecto a una aserción, o a un evento por venir, o respecto a cualquier otra materia en la que el conocimiento absoluto no existe” (p. 173).

El siguiente gran hito en teoría de la probabilidad vendría ya en el siglo XX, cuando, en 1933 Andréi Kolmogórov representó el cálculo de probabilidades asimilado a la teoría de la medición formulándolo en un lenguaje de la teoría de conjuntos, y propuso los axiomas básicos de la probabilidad. Estos axiomas son los siguientes:

- 1) No existen probabilidades negativas: $0 \leq P(\alpha) \leq 1$
- 2) La probabilidad de una tautología o evento seguro es 1: $P(\Omega) = 1$ (donde Ω es el ‘conjunto universal’ no vacío).
- 3) La probabilidad de la disyunción de los enunciados lógicamente excluyentes es igual a la suma de sus probabilidades respectivas: $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$
- 4) La probabilidad de la conjunción de dos enunciados es igual a la probabilidad del primero, asumiendo que el segundo sea satisfecho, multiplicada por la probabilidad del segundo: $P(\alpha \wedge \beta) = P(\alpha | \beta) \times P(\beta)$

De este modo, vemos que las funciones de probabilidad se dan en un espacio *finito* Ω (conjunto universal, que va del 0 al 1) que está compuesto por subconjuntos (como α o β , llamados *campos* o *álgebras*) cerrados bajo complementación y unión. Por ejemplo, si todo $\Omega = \alpha + \beta$ (unión) y $\Omega = 1$, entonces $\beta = 1 - \alpha$ (complementación). Así, los tres primeros axiomas hacen referencia a los límites del terreno de la probabilidad, y el cuarto a la probabilidad condicionada. Del cuarto axioma se deriva el Teorema de Bayes, teorema que, para muchos autores en psicología, es central para el razonamiento probabilístico que permite la actualización de creencias.

Por su parte, la perspectiva subjetivista continuó avanzando a principios del siglo XX de la mano de Frank P. Ramsey (1929) quien interpretó la probabilidad, ya explícitamente, como *grado de creencia* en una proposición o grado de *adhesión* a la posibilidad de verificación de un suceso (Madrid-Casado, 2012). Esta perspectiva fue formalizada en 1937 por Bruno de Finetti, identificando los grados de incertidumbre como un sistema de apuestas justo para un “agente” que establece la probabilidad a priori de A, las condiciones de B dado A y, finalmente, calcula la probabilidad de A dada la información B. Estos trabajos fueron difundidos por Savage (1954) responsable del resurgimiento de la probabilidad bayesiana, y de que a la perspectiva subjetivista de la probabilidad se le llame directamente interpretación bayesiana.

11. Interpretaciones de probabilidad

La interpretación de la probabilidad es un tema complejo y aún no resuelto que implica adherirse a los axiomas del cálculo de probabilidades, como los propuestos por Kolmogórov. Según Salmon (1966), es crucial que existan métodos fiables para determinar los valores de las probabilidades y que estas tengan un campo de aplicación práctico. Hájek (2019) añade que cualquier interpretación debe explicar la relación entre frecuencias y probabilidades y cómo estas últimas guían las creencias y decisiones racionales de los individuos.

A continuación, se expondrán de manera muy sucinta todas las interpretaciones de probabilidad no subjetiva. Después, como sección aparte, pasaremos a considerar la probabilidad subjetiva, pues es la que se adoptará tanto en el modelo teórico como en los experimentos de esta tesis.

11.1. Interpretaciones no subjetivistas

Interpretación Clásica de la Probabilidad: La interpretación clásica, que data de Pascal, Huygens y se extiende a Keynes (1921) y Jaynes (1968), trata la asignación de probabilidades de manera a priori, en ausencia de evidencia o asumiendo que se distribuye de manera equilibrada. Esta perspectiva considera la probabilidad como una relación de casos favorables entre los posibles, basándose en el principio de equiprobabilidad y enfrentando críticas por su naturaleza apriorística y su desconexión con frecuencias observadas (Keynes, 1921; Batanero, Cañizares y Díaz, 1996). En el mundo cotidiano, puede ser difícil definir cuales y cuántos son los casos posibles; a su vez, estos tienen que ser finitos. Además, la equiprobabilidad se aplica a los casos en los que no disponemos de información alguna, lo que muchas veces no se cumple.

Interpretación Lógica (Inductiva) de la Probabilidad: Desarrollada por figuras como Ramsey y Carnap, esta interpretación ve la probabilidad como una medida del grado de apoyo que una evidencia proporciona a una hipótesis. A diferencia de la deducción, donde la conclusión está implícita en las premisas, la inferencia inductiva introduce información nueva en la conclusión. En la conclusión hay más información de la que había en las premisas: es ampliativa. El objetivo de esta

interpretación es explicar cómo se produce formalmente esta ampliación. Esta visión enfrenta el desafío de precisar qué constituye una relación de inferencia inductiva y cómo las probabilidades lógicas pueden vincularse con las frecuencias observadas, lo que algunos autores consideran un objetivo potencialmente inalcanzable para establecer afirmaciones de interés real, ya sea en el ámbito cotidiano o científico (Hájek, 2019).

Interpretación Frecuentista de la Probabilidad: La interpretación frecuentista asocia directamente la probabilidad con la frecuencia relativa de un evento en una serie de ensayos, considerándola una propiedad objetiva y empíricamente verificable del mundo. Autores como von Mises han intentado axiomatizar esta perspectiva basándose en la estabilización de frecuencias relativas a través de repetidos ensayos. Sin embargo, esta aproximación se complica cuando se enfrenta a eventos únicos o a clases de referencia finitas, planteando el "problema del caso único" y cuestionando la adecuación de identificar probabilidades con frecuencias relativas en todas las circunstancias (Hájek, 2019).

Interpretación de Propensiones de la Probabilidad: Originada como una evolución de la interpretación frecuentista, esta perspectiva considera las probabilidades como "propensiones" o tendencias físicas intrínsecas de objetos y situaciones para generar resultados específicos. Definidas como propiedades objetivas y estables a largo plazo, estas propensiones representan cómo un objeto o configuración, bajo condiciones repetidas, tenderá a producir una secuencia de resultados con frecuencias relativas estables (Díez y Moulines, 1999; Hájek, 2019). A pesar de su base objetiva, las propensiones enfrentan críticas significativas debido a la dificultad de determinar empíricamente estas propiedades y evaluar hipótesis basadas en ellas (Sober, 2000; Hitchcock, 2002).

11.2. Interpretación de la probabilidad subjetiva

La Interpretación Subjetiva de la Probabilidad, que se extiende desde Aristóteles en la antigüedad clásica hasta la modernidad, ha evolucionado pasando de ver lo probable como algo que ocurre frecuentemente a una perspectiva epistemológica acerca de lo que *se sabe* que ocurre la mayoría de las veces. Y de ahí, a una concepción plausibilista o endoxástica, donde lo plausible es lo probable en tanto que aceptado como tal en una comunidad epistémica o por los expertos de esa comunidad. Esta interpretación está estrechamente relacionada con la argumentación y la retórica, y se aplica en la toma de decisiones, la ética y la jurisprudencia, destacando la importancia de la coherencia y la cuantificación de la probabilidad dentro de un sistema.

Con el desarrollo de la matemática moderna, la medición de la probabilidad ha avanzado significativamente. Filósofos como Leibniz y Bernoulli en los siglos XVII y XVIII, respectivamente, intentaron fusionar estos métodos matemáticos con la filosofía de la probabilidad. Leibniz, quien sostenía una interpretación epistemológica de la probabilidad, buscó formalizar lo contingente para el derecho. Bernoulli, por su parte, también considera la probabilidad como grado de certeza, entendiéndola como un atributo de la opinión. Así, creó unas reglas de combinación de probabilidades basadas en argumentos. Además, fue el primero en abordar la probabilidad inversa

o inductiva mediante su teorema de los grandes números como justificación para determinar el grado en el que la observación permite sostener una conjetura. En todo caso, al estar basado en frecuencias relativas, no podía aplicarse a eventos de ocurrencia única, acerca de los cuales también tenemos opinión.

Thomas Bayes en 1763, y posteriormente Pierre-Simon Laplace en 1795, avanzaron en la comprensión de la probabilidad, presentándola como un grado de certeza que se ajusta con nueva evidencia. Esta idea fue reforzada por De Morgan en 1847, quien articuló la probabilidad como un *grado de creencia*. Este desarrollo enfatiza la idea de que la probabilidad es una representación del entendimiento subjetivo del mundo más que un simple reflejo de la realidad física. Esto es, es una medida de nuestro conocimiento e ignorancia a su respecto (Díez y Moulines, 1999). Cuando el sujeto no dispone de *toda* la evidencia relevante para poder afirmar la verdad de una proposición o para la verificación concluyente de un suceso, entonces le asignará un valor entre 0 y 1¹⁴. Este valor expresa la intensidad de su creencia en la proposición u ocurrencia del suceso *dada la evidencia de que dispone*. Así, *todas las probabilidades son condicionadas* (Díez y Moulines, 1999), dependientes de los conocimientos previos del individuo (Oaksford y Chater, 2007). La asignación de probabilidades *a priori* obedece a la ley de la indiferencia si no se tienen conocimientos previos relevantes, mientras que se asignarán valores concretos si se tienen *razones* que lo permitan. Las probabilidades *a posteriori* se ajustan con la entrada de nueva evidencia y su relación con los conocimientos preexistentes. Dado que los conocimientos preexistentes varían entre individuos y comunidades, diferentes personas, aunque racionales, pueden asignar distintos grados de credibilidad a la misma hipótesis ante la misma evidencia.

Esta interpretación flexible es conocida como bayesiana, y ha sido respaldada por figuras como Keynes (1921), Ramsey (1928), de Finetti (1937) y Savage (1954). Estos autores sostienen que los grados de creencia deben seguir los principios matemáticos de la probabilidad para asegurar coherencia y evitar contradicciones, manteniendo la consistencia en el sistema de creencias. Ramsey (1926) asimiló la teoría de la probabilidad a una “lógica de la creencia parcial” que exige que un agente racional sea *lógicamente consistente*. Pese a que distintos agentes racionales puedan diferir en sus grados de creencia, no pueden hacerlo en su modo de combinarlos (Díez y Moulines, 1999).

La teoría de la probabilidad numérica se desarrolló inicialmente en el contexto de los juegos de azar y apuestas. Por este motivo, los grados de creencia se han operativizado tradicionalmente en términos del *comportamiento* de apuestas. Se considera que los grados de creencia consisten en la confianza de un agente racional en la verdad de una proposición: hasta qué punto está dispuesto a depender de ella para sus acciones. De Finetti (1937) propuso que la cantidad que una persona está dispuesta a apostar refleja su nivel de confianza en la ocurrencia de un evento. Por ejemplo, si alguien cree con un 75% de certeza que un restaurante está abierto, debería apostar el 75% de una

¹⁴ Por el contrario, cuando dispone de *toda* esa información, la probabilidad desaparece. Sólo hay conocimiento cierto de las leyes deterministas que rigen la naturaleza.

suma determinada a favor de esta afirmación (Schwitzgebel, 2019). Esta aproximación ha sido útil en áreas como la teoría de la decisión y la economía para manejar la incertidumbre, exigiendo que los grados de creencia sean consistentes con los axiomas de Kolmogórov para evitar inconsistencias, como las que se presentan en la *apuesta holandesa*, donde el apostador termina perdiendo sin importar el resultado del evento (Teira, 2012). Así, un apostador inconsistente será siempre irracional. Sin embargo, el enfoque operacionalista que vincula directamente las creencias con las decisiones de apuestas ha sido criticado por su potencial para no reflejar las verdaderas creencias de los individuos, afectadas por factores personales o situacionales. Eriksson y Hájek (2007) señalan que la cantidad de dinero apostada puede no ser un indicador fiable de la creencia, variando según la situación económica del apostador, lo que subraya que el valor monetario puede no corresponderse siempre con la confianza real en un hecho.

Para evaluar las probabilidades subjetivas, se propone utilizar 'utilidades' en lugar de criterios económicos, adaptando mejor la medida a las circunstancias individuales y reflejando la utilidad real para la persona en lugar de un valor monetario puro. Numerosos autores, incluidos Arnaud y Nicole (1662), De Finetti (1937), Jeffrey (1965), Ramsey (1926), Savage (1954), y Von Neuman y Morgenstern (1944), han establecido relaciones entre probabilidades, utilidades y preferencias. Según Oaksford y Chater (2007), estas deben ajustarse simultáneamente según las elecciones de los agentes, lo que permite inferir sus utilidades y probabilidades subjetivas a partir de sus preferencias. Esta idea, fundamentada en la teoría de la decisión, enfatiza la racionalidad en la toma de decisiones orientadas a maximizar la utilidad esperada, alineándose con los axiomas de la teoría matemática de la probabilidad. No obstante, Eriksson y Hájek (2007) critican este método por asumir conexiones no garantizadas entre creencias y deseos, que en realidad pueden ser conceptual y fácticamente independientes. Ellos defienden que los grados de creencia son conceptos primitivos que no deberían derivarse de comportamientos como las apuestas. Las disposiciones a la conducta son solo una parte del rol funcional de las credibilidades. Este punto de vista instrumentalista, que prioriza la acción sobre el conocimiento, es cuestionado en cuanto a su pertinencia para analizar el sistema cognitivo y para el modo de medir los grados de creencia de manera psicológicamente válida.

Aunque la racionalidad del agente, la conformidad con los axiomas de la probabilidad matemática y la coherencia son fundamentales en la probabilidad subjetiva, este enfoque aún requiere de un criterio similar a la verdad que le permita distinguir las representaciones merecedoras de asignaciones de probabilidad de las que no lo son. Aristóteles ya relacionaba lo verosímil con *lo que comúnmente ocurre o lo aceptado por expertos*, destacando cómo nuestras percepciones de lo probable están influenciadas por lo observado frecuentemente o las opiniones de una comunidad. Ramsey (1928) expuso que la razonabilidad de un grado de creencia u opinión se mide por su correspondencia con las frecuencias observadas. Por ejemplo, una creencia sobre la toxicidad de setas amarillas basada en la observación de que la mayoría de las setas amarillas son venenosas es un ejemplo de cómo las creencias deben calibrarse con las frecuencias relativas para ser consideradas racionales (van Fraassen, 1984). Este método de calibración garantiza que los grados de creencia se ajusten a los axiomas de la probabilidad. Adicionalmente, Gaifman (1988)

resalta cómo las opiniones de expertos pueden influir en la formación de creencias racionales. Sugiere que un individuo podría adoptar las probabilidades asignadas por un experto en función de su confianza en dicho experto. Este proceso, conocido como "asignación de expertos", es una continuación moderna de la práctica clásica de basar las probabilidades en opiniones autorizadas (*éndoxa*) y juega un papel crucial en cómo las personas forman opiniones racionales hoy en día.

En conclusión, la Interpretación Subjetiva de la Probabilidad refleja un profundo desarrollo histórico y filosófico que transforma nuestra comprensión de lo probable. La probabilidad, originalmente concebida desde una perspectiva plenamente epistemológica, se diluyó en una disparidad de interpretaciones, observándose hoy un fuerte resurgimiento de la perspectiva epistemológica. Esta interpretación ha evolucionado a través de las contribuciones de pensadores fundamentales como Leibniz, Bernoulli, Bayes y Laplace, destacando el papel de la probabilidad como un atributo de nuestro entendimiento y conocimiento subjetivo del mundo. En la era contemporánea, la interpretación subjetiva o bayesiana está adquiriendo una relevancia creciente para evaluar creencias y decisiones en un marco que valora la coherencia lógica y la adaptabilidad ante nueva información.

CAPÍTULO 3

Filosofía de las creencias: ontología, clasificación y organización en sistemas de creencias

Este capítulo avanza en la construcción del modelo psicológico de la credibilidad, introduciendo elementos conceptuales clave tras revisar conceptos lógicos y de probabilidad en el capítulo anterior, y su relación con el razonamiento y la argumentación cotidiana. Aquí se exploran propuestas filosóficas para definir y clasificar las creencias según su accesibilidad consciente y su función en los sistemas de justificación de la memoria, preparando el terreno para una discusión más amplia sobre su organización y justificación.

Muchos filósofos describen la creencia como una *actitud proposicional*, en la que el sujeto adopta una postura respecto al estatus de verdad de un contenido. Schwitzgebel (2019) define la actitud proposicional como el estado mental que implica tener una postura u opinión sobre una proposición, expresable en la forma 'S-A-que P', donde 'S' es el sujeto, 'A' la actitud, y 'P' la proposición. Esta perspectiva contempla que un individuo puede tener diversas actitudes hacia una misma proposición, como esperar, desear, temer o creer que P. Stalnaker (2002) amplía el concepto de proposición más allá del lenguaje, abarcando cualquier contenido representacional que pueda expresarse declarativamente. Además, se reconoce que creer no requiere una constante actividad consciente, admitiendo que las creencias pueden ser tanto activas como tácitamente almacenadas, categorizándolas en ocurrentes, disposicionales, explícitas e implícitas.

Finalmente, en este capítulo se aborda cómo se estructuran las creencias en sistemas que interactúan mediante procesos inferenciales, justificando o refutando otras creencias. Se analizan dos estructuras principales: el fundacionalismo, que construye justificaciones desde creencias básicas conectadas a percepciones, y el coherentismo, donde la coherencia interna justifica todas las creencias del sistema. Ambos modelos ofrecen variantes que reflejan la complejidad y dinamismo de los sistemas de creencias.

1. Qué es una creencia: diferentes propuestas

Esta sección explora diversas propuestas sobre la naturaleza de las creencias procedentes de las principales teorías de la filosofía de la mente; en particular: la representacionalista, la disposicionalista y la funcionalista.

1.1. Propuesta Representacionalista

La teoría representacionalista considera que las creencias son estados mentales que contienen representaciones de objetos o situaciones del mundo, siendo intencionales y declarativas (Brentano, 1874; Fodor, 1975; Searle, 1980). Estas representaciones pueden influir causalmente en el comportamiento, pues predisponen al individuo a actuar como si la creencia fuera cierta. Para algunos filósofos representacionalistas, esta es una característica distintiva de las representaciones tipo-creencia. Además, se debate sobre el formato de estas representaciones, que algunos ven como lingüísticamente estructuradas, mientras otros proponen formatos más visuales o espaciales, como escenarios o mapas mentales.

1.1.1. Teoría Representacional de la Mente (TRM)

La Teoría Representacional de la Mente (TRM) de Fodor plantea que las representaciones mentales son estructuras lingüísticas dotadas de sintaxis combinatoria y semántica compositiva. Según Fodor (1975, 1987), estas representaciones se componen de símbolos atómicos que se combinan siguiendo las reglas de un lenguaje formal, resultando en frases o fórmulas bien formadas. El valor de verdad de estas representaciones, esencial para su función como creencias, se deriva de la disposición y características de sus componentes sintácticos, similar a la lógica proposicional usada en ciencias de la computación.

Fodor argumenta que este lenguaje del pensamiento permite la productividad y sistematicidad del pensamiento humano, haciendo posible pensar un número indefinido de proposiciones y manipular esas representaciones para generar nuevas creencias o realizar inferencias lógicas. Un ejemplo de esto es la capacidad de creer que "si Juan corrió, entonces María nadó", y a partir de adquirir nueva información ("Juan corrió"), llegar a creer que "María nadó". Esta última representación está *causada* por operaciones sobre la forma sintáctica de las representaciones mentales implicadas. Este lenguaje del pensamiento es visto por Fodor como una estructura lingüística abstracta e innata, compartida por todos los humanos pero inaccesible a la introspección directa. Las operaciones sobre estas estructuras sintácticas pueden traducirse en acciones concretas y decisiones, así como en la formación de nuevas creencias, gracias a la implementación de procesos computacionales que respetan las restricciones semánticas y de verdad. A pesar de estas capacidades, Fodor reconoció limitaciones en la TRM, especialmente en su capacidad para explicar la integración de creencias en sistemas holísticos y contextuales. Críticos como Rives (2010) y el propio Fodor en otros trabajos (2000, 2008) han señalado que, dado que la TRM solo tiene alcance para las propiedades modulares de las representaciones, no abarca la complejidad de las creencias

como entidades integradas en redes más amplias y dependientes del contexto. En este sentido, algunos críticos argumentan que la estructura subyacente del pensamiento y la creencia puede no ser lingüística en absoluto.

1.1.2. Los Mapas Mentales

Varios filósofos sostienen que las representaciones mentales pueden estructurarse de manera análoga a un mapa, tal como argumentan Camp (2007) y Lewis (1994). Esta estructura permite una alta productividad y sistematicidad en el pensamiento, ya que los elementos del mapa mental pueden ser manipulados, combinados o repetidos para representar múltiples estados de cosas, como invertir la posición relativa de un río y una montaña. Los mapas mentales se caracterizan por su alto nivel de integración: cualquier cambio en un elemento afecta a los adyacentes. Así, los cambios en una creencia deberán afectar a otras. Según Schwitzgebel (2019), esta perspectiva demanda un nivel de integración excesivo entre las representaciones del sistema. Esto le impide reflejar adecuadamente la coexistencia de creencias inconsistentes que se podrían dar en el sistema de creencias humano. El contenido de las creencias requiere de un nivel de especificidad tan elevado que podría impedir el nivel de indeterminación necesario para su flexibilidad y cambio.

En el contexto psicológico, estos modelos subyacen a varias teorías como las "cuasi-fotos" de Kosslyn (1980), los modelos mentales en razonamiento de Johnson-Laird (1983), y los modelos de situación en procesos del discurso de Van Dijk y Kintsch (1983). Además, ciertas interpretaciones vinculan los mapas mentales con la semántica de mundos posibles de la lógica modal, como sugieren Boden (1988) y Johnson-Laird (1994). Según esta perspectiva, el grado de creencia se determinaría por el número de mundos posibles en los que la creencia es verdadera, respecto de un total (i.e., favorables entre posibles). Esto permite abordar el problema de las "medias creencias" y las indeterminaciones en el contenido de las creencias. Más tarde, Johnson-Laird cuestionó esta posibilidad, argumentando que podría ser computacionalmente inviable para la mente humana (Johnson-Laird y Byrne, 2002).

1.1.3. Conexionismo

El conexionismo presenta una visión alternativa de la estructura de las representaciones mentales, desafiando las nociones tradicionales basadas en la manipulación de símbolos y el uso de lenguajes formales lógico-matemáticos. En lugar de adherirse a procesos computacionales seriales, el conexionismo se inspira en la organización y funcionamiento del cerebro, utilizando una red de nodos interconectados por arcos, similares a las neuronas y sus conexiones sinápticas. La cognición, según este modelo, emerge de la propagación simultánea o paralela de la activación a través de estos nodos, donde cada nodo puede tener diferentes niveles de activación determinados por pesos variables.

Dentro de este marco, una representación mental es vista como un patrón de activación distribuido por toda la red, cambiante y adaptable en función de la experiencia del individuo (Ballesteros, 1992). Esta capacidad de adaptación permite cambios conceptuales significativos a partir del aprendizaje, reflejando una evolución de los patrones de activación en respuesta a nuevas

informaciones o experiencias. Aunque el conexionismo no emplea un lenguaje formal o simbólico, implementa un procesamiento computacional y representacional que permite dinamismo y contextualidad en las estructuras mentales. En el contexto de las creencias, el conexionismo no se centra directamente en ellas, pero se vincula indirectamente mediante redes bayesianas que gestionan probabilidades subjetivas (McClelland, 1998; McClelland y Thompson, 2007). Esto proporciona un modelo para comprender la credibilidad y la incertidumbre en las percepciones y juicios humanos.

1.1.4. ¿Y Qué Hay de la Credibilidad en Todo Esto?

La perspectiva representacionalista concibe la creencia como un estado mental que alberga representaciones declarativas, intencionales y con poderes causales (sobre la conducta u otras creencias). Se han centrado en desarrollar teorías detalladas sobre su formato y características como la composicionalidad, productividad y sistematicidad. Esta perspectiva pone más énfasis en la naturaleza de las proposiciones que en el estado mental que las alberga.

Recientes desarrollos en esta perspectiva subrayan que concebir una creencia como un acto de "creer que p" implica adherirse a una norma de verdad, tal como lo argumentan Shah y Velleman (2005) y Engel (2018). Esta norma establece que una creencia es *correcta* solo si el contenido proposicional es verdadero, lo que establece una "dirección de ajuste" del contenido de la creencia hacia la realidad del mundo, en contraposición a actitudes como deseos o intenciones, donde se busca que el mundo se ajuste al contenido de la actitud. Burge (2010) destaca que el propósito fundamental de las creencias es generar representaciones precisas y verídicas de la realidad. Sin embargo, esta perspectiva no ha explorado profundamente el aspecto de la intensidad o fuerza con que se adhiere a la verdad de las representaciones, un aspecto crucial de la credibilidad o el grado de creencia. Ciertamente es que la posesión de la representación es binaria, o se tiene o no se tiene; lo que es gradual es la fuerza de la adhesión a la verdad de la representación. Y es ese carácter de grado lo característico de la credibilidad o grado de creencia, muchas veces asimilado a la probabilidad subjetiva.

1.2. Propuesta Disposicionalista

La propuesta disposicionalista, respaldada por pensadores como Braithwaite (1933), Marcus (1990) y Ryle (1949), plantea que la creencia es esencialmente una actitud proposicional cuya importancia no reside en las representaciones mentales que implica, sino en las disposiciones conductuales que provoca. Según esta visión, una creencia se define más por la tendencia a actuar de cierta manera bajo condiciones específicas que por el contenido mental representativo que la sostiene. La disposicionalidad se manifiesta en conductas como asentir a una proposición si se cumple que el individuo se encuentra en un estado óptimo de consciencia y sinceridad, o mostrar sorpresa al descubrir que lo creído no es cierto. Eriksson y Hájek (2007) y Schwitzgebel (2019) exploran estas disposiciones, proponiendo que tales tendencias pueden describir una creencia sin necesidad de invocar representaciones mentales. Sin embargo, esta perspectiva enfrenta críticas, especialmente en cuanto a su capacidad para explicar diferencias en comportamientos derivados de la misma

creencia debido a factores como deseos o creencias adicionales, lo que Chisholm (1957) señala como una posible confusión entre las causas de la conducta y sus efectos.

La medición tradicional del grado de creencia se ha basado en la disposición a la conducta de apuestas, según de Finetti (1937) y Ramsey (1926), donde se apuesta una cantidad representativa de la convicción en la verdad de una proposición. Sin embargo, esta metodología ha enfrentado críticas por su simplificación y por vincular directamente las creencias con utilidades y deseos. Las apuestas están ligadas a las utilidades, y éstas a los deseos. Y la dirección de ajuste mente-mundo de los deseos es la contraria a la de las creencias. Eriksson y Hájek (2007) argumentan que las creencias y las disposiciones a actuar basadas en ellas no deberían considerarse equivalentes. Las creencias buscan representar el mundo y son ajustadas cuando hay discrepancias entre lo creído y la realidad, mientras que los deseos representan cómo queremos que el mundo sea, sin necesidad de ajustarse a la realidad para ser considerados correctos. Por otro lado, estos autores también señalan que las disposiciones a la conducta se explican en función de las creencias y no al revés: “el orden de explicación o de predicción no es: ‘estas son tus disposiciones o preferencias para apostar; por tanto, estas son tus credibilidades o deseabilidades’. Sino más bien: ‘estas son tus credibilidades y deseabilidades; por tanto, estas serán tus disposiciones o preferencias para apostar’” (p. 208). Además, las creencias influyen en más que solo decisiones de apuestas; también afectan estados emocionales y psicológicos, demostrando una relación más compleja entre creencias, deseabilidades y comportamiento.

1.3. Propuesta Funcionalista

La propuesta funcionalista define las creencias no por su estructura interna, sino por su función dentro del sistema cognitivo. Según teóricos como Armstrong (1968), Block (1978), y Fodor (1975), la identidad de las creencias está determinada por sus relaciones causales con estímulos sensoriales, otros estados mentales y conductas. Schwitzgebel (2019) clasifica estas relaciones en cuatro tipos: percepción (observar p , causa la creencia p), razonamiento teórico (la creencia en p y en $p \rightarrow q$ causará la creencia en q), razonamiento práctico (si *deseo* M y *creo* que mediante A se llega a M , entonces realizo la *conducta* A) y aserción (si *creo* p , sinceramente diré p), donde las creencias funcionan tanto como causas como efectos dentro de procesos cognitivos.

Esta propuesta sugiere que las creencias son valiosas por su capacidad de dirigir la atención, facilitar inferencias, motivar acciones y propiciar expresiones verbales, reflejando un papel activo en moldear y responder al entorno percibido. Aunque el funcionalismo aún no ha explorado completamente los grados de creencia, su compatibilidad con el disposicionalismo apunta hacia una visión común de que las creencias son representaciones con propiedades causales que predisponen a conductas específicas. Esto resuena con el punto de vista de Ramsey (1926), quien describe las creencias como “una propiedad causal... que podemos expresar como el punto hasta el que estamos dispuestos a actuar en consecuencia” (Ramsey, 1926; p. 169), destacando su eficacia causal en nuestras acciones. Además, aunque el funcionalismo y el representacionalismo se centran en diferentes aspectos de las creencias, ambos reconocen su naturaleza representacional y causal,

sugiriendo que una creencia efectiva es aquella que emerge de y conduce a interacciones significativas con el mundo, balanceando su origen perceptivo y sus consecuencias conductuales.

1.4. Conclusión

Las propuestas representacionalista, disposicionalista y funcionalista consideran las creencias como actitudes proposicionales, pero difieren en cómo conceptualizan el papel de las representaciones y los componentes actitudinales. El representacionalismo enfatiza la importancia de poseer una representación adecuada para que esta funcione efectivamente como una creencia. Esta representación puede ser lingüística, como en la Teoría Representacional de la Mente (TRM), o no lingüística, como en los modelos basados en mapas mentales o patrones de activación neural, destacando su flexibilidad y actualización continua basada en la experiencia. Por otro lado, el disposicionalismo minimiza la importancia de las representaciones en la definición de creencias, concentrándose en lugar de ello en las disposiciones conductuales que las creencias provocan. Según esta perspectiva, las creencias se manifiestan como propensiones a actuar de ciertas maneras bajo condiciones específicas, lo que también influye en cómo se entienden los grados de creencia, particularmente en términos de disposiciones a apostar. El funcionalismo, en cambio, define las creencias no por sus representaciones internas sino por la función que desempeñan dentro del sistema cognitivo. Esta orientación considera que las creencias son importantes por las causas que las generan y los efectos que producen, situándolas en un entramado causal que afecta tanto la cognición como la conducta.

Aunque estas teorías comparten ciertos aspectos, también heredan críticas similares, especialmente en cuanto a la reducción de los grados de creencia a meras disposiciones conductuales. En última instancia, la comprensión más profunda de las creencias podría residir en la evaluación de probabilidades, vinculando así el componente actitudinal de las creencias con la manera en que se relacionan las representaciones con la subjetividad del individuo. Esta visión emergente apunta a una interpretación más matizada de la credibilidad y los grados de creencia más allá de simples disposiciones a la conducta.

2. Clasificación de las creencias

Las creencias se pueden clasificar en varios tipos según diferentes criterios. Una clasificación básica las divide en *ocurrentes* y *disposicionales*. Las creencias ocurrentes son aquellas activas en la conciencia en un momento específico, generalmente debido a estímulos ambientales. Por ejemplo, ver una película puede activar la creencia de que "el kung fu chino es un arte marcial más completo que el boxeo inglés". Fuera de ese momento de activación, dicha creencia permanece disposicional, almacenada en la memoria y disponible cuando sea necesaria. Esta categoría de creencias disposicionales no debe confundirse con la perspectiva disposicionalista sobre las creencias, aunque comparten terminología (Schwitzgebel, 2019).

Además, las creencias se clasifican en *explícitas* e *implícitas*. Las creencias explícitas son aquellas para las que existe una representación almacenada en la memoria. En contraste, las creencias implícitas son derivaciones lógicas de creencias explícitas que no están representadas en la memoria de forma individual. Por ejemplo, si se cree explícitamente que hay siete días en la semana, implícitamente se cree que hay más de seis y menos de ocho días. Del mismo modo, sabiendo lo que es un elefante y lo que es una catedral gótica, uno también puede tener la creencia implícita de que ‘un elefante cabe en una catedral gótica’, aunque nunca antes haya pensado en ello. Aunque se puedan considerar implícitas, en realidad estas creencias no tienen representaciones concretas en la mente hasta que se hacen explícitas a través de algún tipo de operación cognitiva. Schwitzgebel (2019) critica a los representacionistas por su dificultad para explicar las creencias implícitas, argumentando que no es práctico ni plausible asumir que cada creencia individual está almacenada de manera representacional en la mente. Algunos filósofos, como Dennett (2017/1978) y Fodor (1987), sugieren que las creencias implícitas pueden ser el resultado de patrones emergentes del "hardware" y "software" cognitivo, sin representaciones individuales asociadas. Estas creencias se vuelven explícitas cuando el contenido relevante se deriva de algo ya representado.

Finalmente, la distinción entre creencias ocurrentes y disposicionales también se aplica a las creencias explícitas e implícitas. Mientras que las explícitas pueden ser tanto ocurrentes (activas) como disposicionales (latentes), las implícitas, al no tener representaciones hasta ser derivadas, se consideran siempre ocurrentes; antes de ser explícitas, no existen como tales, sino solo en potencia.

3. Transmisión de justificación en los sistemas de creencias

En los sistemas de creencias y conocimiento, todas las creencias —ya sean ocurrentes, disposicionales, explícitas o implícitas— dependen de la memoria para su existencia y funcionamiento. Estas creencias, en gran medida, tienen la función de representar lo que consideramos tentativamente verdadero o falso, y su corrección se rige por un criterio normativo de verdad, especialmente cuando se expresan verbalmente. Bonjour (1985) sugiere que el sistema cognitivo aspira a conocer la verdad y utiliza la *justificación* como medio esencial para aproximarse a ella, definiendo la justificación como “*grosso modo*, algo como una *razón* o *garante* de algún tipo que cumple con algún estándar apropiado” (p. 5-6). Sin embargo, estar justificadas no garantiza que las creencias sean correctas.

Filosóficamente, existen dos perspectivas sobre el origen de la justificación de las creencias. Una sostiene que la justificación viene del mundo externo, debido al carácter representacional de las creencias. La otra argumenta que las creencias se autojustifican a través de su coherencia interna. Ambas perspectivas coinciden en que, frecuentemente, la justificación de una creencia se basa en otras creencias presentes en la memoria, integradas en sistemas de creencias conectados, y que la transmisión de la justificación puede depender de los procesos de memoria y razonamiento.

Este capítulo continúa explorando dos orientaciones principales sobre los procesos de justificación en los sistemas de creencias: el fundacionalismo y el coherentismo, ambos reconociendo que los sistemas de creencias operan dentro de la memoria.

3.1. Fundacionalismo

El fundacionalismo, también conocido como fundamentismo, postula que la memoria actúa no solo como un almacén de creencias sino también como un mecanismo crucial de justificación. En esta concepción, las creencias racionales se justifican mediante el apoyo de otras creencias previamente justificadas (creencias derivadas) o a través de experiencias directas del mundo externo o de intuiciones axiomáticas de la razón (creencias fundacionales). Asumiendo que el sustento de muchas creencias es *inferencial* (Fumerton, 2016), el fundacionalismo enfatiza la necesidad de detener el proceso de justificación en una fuente original, de modo que obtengan su justificación de una manera *no inferencial*, evitando una regresión infinita o circular de la justificación. Estas serían las creencias básicas, que constituyen el sustento primario de todo el proceso de justificación.

3.1.1. Argumentos a favor y en contra de la hipótesis fundacionalista

Aristóteles, considerado el primer fundacionalista (Hasan y Fumerton, 2016; Poston, 2008), argumenta en sus *Analíticos Segundos* (L. I, 3) que hay conocimientos que se fundamentan directamente en la evidencia sin necesidad de demostración previa, deteniendo así la regresión epistémica. Esta perspectiva evita tanto la regresión infinita como la justificación circular, ambas vistas como problemáticas. La primera nunca concluye por lo que la justificación nunca llega; y la segunda es racionalmente inaceptable: cualquier creencia está automáticamente justificada, pues es su propia razón. Pero existen algunos autores que, rechazando el fundacionalismo, defienden la regresión infinita o la circularidad. La corriente *infinitista*, representada por Aikin (2011) y Klein (1999, 2007), contrapone que la justificación de una creencia puede depender de una cadena infinita de razones no repetitivas, apelando a la capacidad humana de justificar indefinidamente. Por otro lado, algunos *coherentistas*, como Lehrer (1990; 2003), aceptan la justificación circular si el círculo es suficientemente amplio y que la justificación para cualquier afirmación, tarde o temprano, regresará a la afirmación original. Otros coherentistas, inspirados por Bonjour (1985), buscan superar la regresión apelando a un sistema holístico donde las creencias se justifican mutuamente.

Por otra parte, el fundacionalismo también tiene a su favor el argumento de los juicios sobre casos particulares (Poston, 2008), donde ciertas creencias perceptivas directas o conceptualmente evidentes se consideran autojustificadas. Estas incluyen percepciones sensoriales inmediatas y creencias aparentemente obvias, que no requieren demostración externa: que el blanco no es negro, que la parte es inferior al todo, que algo es rojo porque me parece rojo, o la creencia de que “tengo manos” (Pryor, 2000). El fundacionalismo ha sido criticado por considerarse que las creencias básicas son arbitrarias al carecer de razones que las apoyen (Pollock y Cruz, 1999). Pero los fundacionalistas consideran que una creencia puede estar justificada por la sola experiencia, sin la necesidad de que tal experiencia sea una razón previamente integrada en el sistema.

3.1.2. Versiones del fundacionalismo: fuerte, moderado y débil

El *fundacionalismo fuerte*, también conocido como clásico y de corte cartesiano, enfrenta críticas significativas. Postula que, similar a un sistema axiomático, el sistema de creencias debe fundamentarse en unas pocas creencias que alcanzan el máximo nivel de credibilidad. Estas creencias deben ser infalibles (cuya falsedad sea muy improbable), indubitables e incorregibles, es decir, que no pueden ser alteradas por nueva información. Basándose en estos criterios, el sistema puede desarrollar otras creencias a modo de teoremas y mantener la coherencia y consistencia interna. Apelando al cálculo de probabilidades, Lewis (1952) argumenta que la probabilidad, de tener cabida dentro de este sistema, debe basarse en certezas, ya que, sin un fundamento cierto, ninguna creencia probable podría sostenerse. Si en la regresión de las justificaciones no se llega a encontrar una certeza absoluta, se continúa buscando hasta obtenerla. De lo contrario, las conclusiones no serían racionalmente sostenibles. Además, señala que una cadena extensa de justificaciones tiende a acumular incertidumbre, reduciendo la probabilidad de que las conclusiones finales se consideren ciertas.

El fundacionalismo fuerte se estructura por creencias básicas consistentes en proposiciones lógicas, matemáticas y estados de conciencia individual, como las certezas de tipo cartesiano como "pienso" (luego existo). Estas creencias se consideran inmediatas, infalibles, indubitables e incorregibles, proporcionando una base sólida de seguridad epistémica. Sin embargo, Poston (2008) argumenta que son inadecuadas para comprender el conocimiento cotidiano en un mundo material y variable. Esta perspectiva, al centrarse en principios lógicos que son a priori y desconectados de la realidad empírica, tiene dificultades para vincularse con el mundo externo, perdiendo así lo que era una de sus principales virtudes epistemológicas. Además, la constante variabilidad del mundo y la inexactitud en su observación raramente permiten alcanzar una certeza absoluta, lo que muestra una limitación significativa de esta aproximación al asumir que el sistema de creencias humano funciona como un sistema lógico-matemático cerrado.

Las limitaciones del fundacionalismo fuerte llevaron al desarrollo del *fundacionalismo moderado* (Huemer, 2001; Pryor, 2000). Este enfoque acepta la existencia de creencias básicas sobre el mundo. Considera que estas creencias, si bien pueden parecer ciertas inicialmente, no tienen por qué ser infalibles o indubitables. Esta orientación permite una conexión más realista con el mundo al reconocer que la relación entre la realidad y su representación mental es falible, pudiendo no ser completamente precisa. Así, la *apariencia* dota de justificación a las creencias empíricas básicas, sin necesidad de ser completamente certeras (Huemer, 2013). Esto permite que las creencias básicas sean revisables y ajustables con nueva información. Por ejemplo, en el caso de la ilusión perceptiva de Müller-Lyer, inicialmente se puede justificar la percepción de que dos líneas son de longitudes diferentes debido a su apariencia. Sin embargo, si posteriormente se descubre mediante medición que ambas líneas son iguales, esta nueva información corrige la creencia inicial. Formalmente, si al sujeto s le parece que p , entonces, en ausencia de derrotadores (i.e. contraejemplos o contraargumentos), s tiene algún grado de justificación para creer que p (Cullison, 2010; Huemer, 2007; Tucker, 2010). Para que estas creencias sirvan de base para justificar otras creencias (i.e.,

creencias derivadas), deben presentar un estado epistémico meramente positivo, aunque Van Cleve (2005) argumenta que algunas deben poseer alta credibilidad para ser fiables. Esta versión del fundacionalismo es más flexible que las anteriores. No exige que las creencias básicas tengan certeza absoluta. Además, amplía significativamente el espectro de creencias básicas permitidas, incluyendo creencias perceptivas falibles derivadas de la experiencia, en lugar de limitarse únicamente a certezas lógicas, matemáticas o fenomenológicas de tipo cartesiano.

El *fundacionalismo débil* reduce aún más los requisitos de certeza para las creencias básicas, argumentando que estas solo poseen un grado de justificación inicial muy bajo, insuficiente para cumplir con los estándares requeridos para el conocimiento y para ser consideradas como premisas válidas para otras creencias. Según Bonjour (1985, p. 28), estas creencias básicas son meramente "inicialmente creíbles" y no completamente justificadas. Debido a esta insuficiencia inicial, es crucial la coherencia entre estas creencias inicialmente creíbles para *amplificar* su nivel de garantía justificativa hasta el punto necesario para considerarse conocimiento. Posteriormente, estas pueden funcionar como premisas para otras creencias. La experiencia juega un papel fundamental al proporcionar los primeros indicios que apoyan una creencia o su introducción en el sistema. Sin embargo, esta creencia debe ser confirmada según su coherencia o conexión con el resto del sistema de creencias. Si encaja adecuadamente, su justificación se fortalece; si no, o si resulta contradictoria, debe ser descartada. Lewis (1946) plantea que toda información recordada se considera creíble *prima facie*, ya que normalmente almacenamos en la memoria solo aquello que percibimos como al menos mínimamente verdadero, con una probabilidad superior a 0.5. Argumenta que la *congruencia* entre las creencias dentro de un sistema aumenta su credibilidad. Un elemento se considera congruente con el sistema cuando su probabilidad a priori, evaluada en aislado, es menor que su probabilidad a posteriori, condicionada al resto de elementos del sistema ($P(A) < P(A | B)$). Cualquier información que no concuerde con el conocimiento existente en la memoria se considera menos probable, con una probabilidad menor a 0.5. Por lo tanto, un sistema de creencias coherente puede reforzar la justificación de sus componentes, siempre y cuando ya posean una credibilidad inicial suficiente.

En esta exposición, se observa cómo las diversas formas de fundacionalismo se diferencian por el nivel de certeza requerido para las creencias básicas, evolucionando de un enfoque lógico a uno probabilístico. Los tipos principales de creencias básicas incluyen proposiciones lógicas y matemáticas, así como creencias autoevidentes sobre experiencias inmediatas, y creencias perceptivas que reflejan la realidad experiencial.

3.1.3. Naturaleza de las relaciones inferenciales entre creencias

Todas las creencias no básicas (i.e., derivadas) provienen de las básicas mediante inferencia. Existen varias propuestas sobre la naturaleza de estas inferencias:

Deductivismo: Esta propuesta apela a la lógica deductiva, donde a partir de premisas ciertas se desarrollan argumentos deductivamente válidos que interconectan creencias, formando un sistema

cerrado por consecuencia lógica. Este método requiere un fundacionalismo fuerte; y sin este fundamento, el enfoque deductivista no es viable.

Inductivismo: Contrasta con el deductivismo al basarse en la experiencia. Las creencias básicas empíricas, siendo particulares, fundamentan creencias más generales sobre el mundo. Por ejemplo, la generalización de que “todos los cuervos son negros” surge no de observaciones completas, sino de inducir patrones a partir de muestras limitadas. Estas creencias generales, al no ser directamente observables, se clasifican como derivadas y son producto de inferencias inductivas.

Según Poston (2008) e Iranzo (2012), la inferencia inductiva estándar es de tipo enumerativo. Este proceso implica listar instancias relevantes de una categoría, como cuervos, y verificar si poseen una propiedad específica, como ser negros. Si una muestra representativa no muestra contraejemplos, es razonable concluir que todas las instancias relevantes poseen dicha propiedad, por ejemplo, “todos los cuervos son negros”. Aunque se base en observaciones verificables, la inducción es inherentemente falible, sujeta a revisión y corrección, lo cual se ajusta al fundacionalismo moderado. A diferencia de la deducción, la inducción amplía el conocimiento más allá de los axiomas fundamentales al generar conclusiones expansivas. Sin embargo, no siempre está claro cuándo una muestra se considera suficientemente representativa a nivel psicológico.

Otro método es la inferencia *abductiva*, o inferencia a la mejor explicación. Esta técnica justifica una proposición si constituye la explicación más convincente para un conjunto de datos. Por ejemplo, si observamos la desaparición de queso y escuchamos ruidos en la pared, podríamos concluir que “hay ratones”, siendo esta la explicación más sencilla y amplia que se ajusta a los datos observados, aunque es susceptible de revisión si aparece nueva información (Poston, 2008). Esta forma de inferencia es particularmente adecuada para conectar observaciones dispersas bajo una hipótesis coherente. La inferencia abductiva se alinea efectivamente con el fundacionalismo débil, permitiendo que un conjunto de observaciones individuales, aunque insuficientes por sí solas, justifiquen colectivamente una hipótesis coherente. Un ejemplo claro es cuando un policía recoge testimonios independientes y poco fiables sobre un hombre rubio de mediana edad en una escena del crimen (Lewis, 1946). Si todos los testimonios coinciden, la explicación más plausible y que mejor se ajusta a los datos es que la presencia del hombre es real, proporcionando así una justificación sólida para esta afirmación.

3.2. Coherentismo

El coherentismo describe un modelo en el cual las creencias se justifican mediante su coherencia con un sistema integrado de creencias. En esta línea de pensamiento, la justificación de una creencia no se da de manera aislada sino como parte de un entramado holístico, donde cada creencia respalda a otra dentro de una red compleja de apoyos mutuos. Esta red se asemeja a la estructura de una tela o a las dovelas de un arco románico, donde cada componente sustenta y es sustentado por los demás, formando un todo cohesivo *simétrico* y estable. Según Davidson (1986), en el coherentismo, todas las creencias son igualmente fundamentales, ya que no existen unas creencias básicas y otras derivadas; todas se sostienen mutuamente. La coherencia de un sistema no solo se

mide por la consistencia lógica entre las creencias, sino también por su capacidad para formar explicaciones cohesivas y probabilísticamente consistentes. Este modelo permite la formación de subsistemas temáticos dentro de una estructura más amplia, cada uno sosteniendo su propia coherencia interna. En el coherentismo, se sostiene que un sistema de creencias coherente debe incluir al menos dos creencias, aunque no hay un límite máximo definido. No es obligatorio que todas las creencias de un individuo estén interconectadas, lo que evita que una incoherencia aislada comprometa la integridad del sistema completo. Esto permite la existencia de subsistemas temáticos independientes que no influyen entre sí, evitando así la propagación de errores a través del sistema general de creencias. Además, se acepta que una creencia puede pertenecer a varios subsistemas, obteniendo justificación a través de múltiples conexiones.

Las variaciones dentro del coherentismo reflejan diferentes perspectivas sobre el papel de la coherencia en la justificación. Algunas versiones, conocidas como *coherentismo de suficiencia*, argumentan que la mera coherencia de un sistema confiere justificación a todas sus creencias, sin necesidad de que cada creencia esté individualmente justificada. Otras, como el *coherentismo de necesidad*, ven la coherencia como un requisito esencial, aunque no necesariamente suficiente, para la justificación de un sistema de creencias. Estas distinciones enfatizan cómo la coherencia funciona como fundamento para establecer la validez de las creencias dentro del sistema.

3.2.1. Argumentos a favor y en contra de la hipótesis coherentista

El coherentismo argumenta que las bases cartesianas y empiristas para justificar creencias son demasiado restrictivas y falibles, respectivamente, impidiendo así una justificación efectiva del conocimiento. Los coherentistas sostienen que solo una creencia puede justificar otra, y que la experiencia sensorial por sí sola no provee razones suficientes para la justificación. Según esta visión, para estar justificado en creer algo, como la presencia de una taza, se requiere también *ser consciente* de estar en condiciones perceptivas óptimas y que los sentidos funcionen correctamente (Lehrer, 1990; Murphy, 2006). Estas *creencias de fondo* proporcionan la estructura de apoyo necesaria para la *justificación* de creencias perceptivas. En lugar de apoyarse en una regresión infinita de razones, el coherentismo favorece un modelo de justificación circular (Alston, 1993; Lehrer, 1997), considerando que la circularidad no es un problema siempre que el círculo sea lo suficientemente grande. Otros abogan por un sistema holístico y simétrico (BonJour, 1985), contraponiéndolo a uno lineal y asimétrico que considere que las creencias están individualmente justificadas. En el sistema holístico, todo el sistema está justificado, y una nueva representación lo estará en la simple medida en que sea coherente con el sistema. Todas las creencias son igualmente fundamentales, el sistema entero se justifica si es coherente y una creencia se justifica por todas las creencias del sistema simultáneamente. Esta estructura permite que el sistema sea considerado como un todo integrado, donde cada creencia se justifica dentro del contexto del sistema completo.

Los argumentos del coherentismo han sido criticados desde varias perspectivas. Primero, se cuestiona la necesidad de ser consciente de las creencias de fondo justificadas para validar otras creencias, como un niño que cree ver una taza sin reflexionar sobre las condiciones perceptivas o el

funcionamiento de sus sentidos. Segundo, la estructura circular de justificación ha sido criticada porque si una creencia debe justificarse mediante una serie de creencias que finalmente vuelve a la original, entonces la creencia inicial estaba auto-justificada (A justifica a A), perpetuando un ciclo de autovalidación.

BonJour (1984) argumenta contra la necesidad de justificar individualmente cada creencia dentro del sistema. Utiliza el argumento de Lewis (1946) sobre la concordancia entre testigos independientes como analogía, sugiriendo que la coherencia entre múltiples fuentes de información independientes, aunque sean poco fiables individualmente, puede incrementar la probabilidad de que la información reportada sea verdadera. Esto se puede aplicar a la percepción, donde la coherencia entre diferentes sentidos (p.ej., entre lo visto y lo oído) refuerza la credibilidad de un contenido. De este modo, uno tendría una creencia perceptiva justificada. BonJour (1985) extiende este concepto, proponiendo que la coherencia en sí misma es suficiente para justificar una creencia sin necesidad de una credibilidad previa (*desde cero*). Dice que ese mismo ejemplo muestra cómo no es necesario siquiera que haya una mínima credibilidad inicial. Para el sujeto que juzga, la mera congruencia en p entre testimonios en tal situación de independencia le lleva a creer razonablemente que eso no puede haber ocurrido por casualidad, que tiene que haber una razón. Aun desconociendo esa razón o las razones de cada individuo para afirmar p , la mera congruencia en p es una razón en sí misma para creer p . Con la coherencia basta. Sin embargo, los estudios de Huemer (1997, 2011) y Olsson (2002) han mostrado que la coherencia sola no afecta la probabilidad de la veracidad de los informes sin una credibilidad inicial mínima. Van Cleve (2011) también ha demostrado que, aunque la coherencia puede aumentar la justificación de creencias preexistentes, no puede crear justificación de la nada. Ante estos hallazgos, otros teóricos como Thagard (2000, 2005) sugieren que quizás la probabilidad no sea la mejor manera de evaluar la coherencia entre creencias.

El coherentismo también ha sido criticado por priorizar la coherencia sobre la verdad, lo que lleva a la "objeción del aislamiento": las creencias no necesitan reflejar la realidad para ser justificadas, solo necesitan ser coherentes entre sí. Esta crítica sobre la desconexión legal con el mundo destaca la importancia de la experiencia y la percepción en la justificación de las creencias perceptivas. BonJour (1985) defiende que las percepciones generan *creencias espontáneas* que se justifican *si se cree* que esas creencias de primer orden tienden a ser ciertas. Lehrer (1990) propone que el sistema de creencias incluye nociones sobre condiciones adecuadas para formar creencias fiables. Esto permite tanto justificar creencias correctas como mantener creencias erróneas. En el caso extremo, esto último es lo que sucede en personas con trastornos delirantes, donde las percepciones y experiencias cotidianas se integran en un delirio coherente. La mayoría de las críticas al coherentismo sostienen que la justificación debe aproximarse a la verdad. La simple coherencia entre creencias no garantiza su veracidad o corrección y, por tanto, no es suficiente para justificarlas plenamente. Aunque las teorías coherentistas suelen enfatizar la importancia de las creencias basadas en la experiencia directa para evitar que el sistema de creencias se aisle de la realidad,

ejemplos como los delirios indican que incluso esto puede ser insuficiente para asegurar que las creencias reflejen la realidad de manera fiable y adaptativa.

3.2.2. Versiones de coherentismo

En función de su estructura interna: El coherentismo define la justificación de las creencias basándose en la coherencia interna de un sistema de creencias. Según Ewing (1934), un sistema coherente debe ser lógicamente consistente, sin contradicciones. Sin embargo, esto no basta: un sistema formado por las creencias de “mañana es martes”, “los lobos son mamíferos” y “soy una taza” es consistente, pero no coherente. Además, las creencias deben estar lógicamente conectadas, formando un conjunto cerrado por consecuencia lógica. No obstante, Olsson (2017) ilustra sistemas coherentes que no cumplen esta conexión lógica estricta. Por ejemplo, veamos un sistema formado por las proposiciones: $A \equiv$ ‘Juan estaba en la escena del crimen en el momento del robo’, $B \equiv$ ‘Juan posee una pistola del tipo usado por el ladrón’ y $C \equiv$ ‘Juan depositó una gran suma de dinero en su cuenta bancaria al día siguiente’. Aunque no se derivan lógicamente unas de otras, son coherentes más allá de la mera compatibilidad.

Lewis (1946) propone que un sistema es coherente si cada creencia está respaldada por las demás a través de relaciones probabilísticas, donde la probabilidad de una proposición aumenta al asumir la veracidad de otras (i.e., $P(p) < P(p \mid q)$). Así, en el ejemplo anterior, aunque A no se deduce de B y C, su probabilidad aumenta si B y C son ciertas (i.e., $P(A) < P(A \mid B \text{ y } C)$). Pero, ¿en función de qué se produce ese aumento? Bonjour (1985) amplía este concepto argumentando que la coherencia también depende de la *consistencia probabilística* y la *densidad de conexiones* entre creencias. Olsson (2017) sugiere medir la coherencia por la *densidad inferencial*, evaluada por la relación entre conexiones y el total de creencias, ofreciendo una visión cuantitativa de la coherencia en un sistema de creencias. Pero la cuestión sigue siendo ¿qué es aquello que permite o produce esas relaciones?

En función de la aceptación en el sistema: En el coherentismo, la información entrante es evaluada por la mente para determinar su incorporación a la red cognitiva personal. Frente a múltiples sistemas con coherencia similar, se necesitan criterios para decidir cuál aceptar. Nicholas Rescher (1973) propone seleccionar los candidatos entrantes con mayor consistencia interna y, luego, elegir el más plausible (con mayor apoyo endoxástico). De este modo, ante dos sistemas igual de coherentes, uno será más creíble que el otro. Keith Lehrer (1990, 2003) sugiere que una proposición está justificada si encaja bien con el *sistema de aceptación* del individuo, constituido por proposiciones previamente aceptadas por el sujeto. Este proceso está relacionado con la validación y monitorización epistémica que se discutirá más adelante (sec. 1.5. cap. 1, pt. II). Según Lehrer, la coherencia no es una propiedad global de todo el sistema, sino del sistema local de aceptación. Esto conecta con supuestos de nuestro modelo que se desarrollarán más adelante: la mente busca optimizar la consistencia entre las creencias ocurrentes activas en la memoria de trabajo en un momento dado, es decir, en el sistema local de creencias.

Por otra parte, ante un conjunto de proposiciones aparentemente verdaderas pero conflictivas entre sí, Paul Thagard (2000) propone que la elección de las razonablemente aceptables debe estar basada en criterios de coherencia explicativa entre una proposición en cuestión y el resto de las proposiciones del sistema. Según Thagard, la mente crea coherencia generando hipótesis que vinculan unas creencias con otras, permitiendo explicar nuevos datos o incluso otras hipótesis. Esto implica que la coherencia y aceptación de una proposición dependen de encontrar otras proposiciones en el sistema que le sirvan a modo de hipótesis explicativas. Y, según Thagard, las relaciones explicativas son deductivas. Así pues, si dos proposiciones entrantes armonizan bien entre sí, la aceptación o rechazo de una de ellas afecta simultáneamente a ambas. En cambio, si hay incompatibilidades, como inconsistencias lógicas, solo una de las proposiciones conflictivas puede ser aceptada, siempre que se encuentre una hipótesis que la vincule adecuadamente con el resto del sistema. Thagard también señala que, en ocasiones, se requieren múltiples hipótesis para explicar un fenómeno, lo que puede reducir la coherencia del subsistema a medida que aumenta el número de hipótesis si éstas no están perfectamente conectadas. Por otro lado, sostiene que las proposiciones que describen observaciones directas deben tener prioridad sobre otras creencias en caso de conflicto, ya que estas gozan de una aceptabilidad inherente independiente de la coherencia del sistema. Esta perspectiva acerca sus supuestos a un tipo de fundacionalismo débil, en el que la coherencia explicativa y la verdad están conectadas a través de la descripción de los mecanismos causales que rigen las relaciones entre los objetos del mundo empírico, según argumenta en su trabajo de 2009.

3.2.3. Naturaleza de las relaciones inferenciales entre creencias

Hay un acuerdo general entre los filósofos en que las creencias de un sistema se relacionan de algún modo las unas con las otras. Tal y como se indicó más arriba, tradicionalmente, se han propuesto tres tipos de relaciones: las lógicas o deductivas, probabilísticas o inductivas, y las relaciones explicativas.

3.2.3.1. Propuesto lógico-deductiva

En la teoría lógico-deductiva, se establece que un sistema de creencias coherente debe, como mínimo, garantizar la consistencia lógica. A nivel semántico, la consistencia se cumple si existe una interpretación en la que todas las proposiciones del sistema sean simultáneamente verdaderas; y a nivel sintáctico, cuando no hay contradicción entre ninguna de las proposiciones del sistema. Sin embargo, simplemente evitar contradicciones no es suficiente para afirmar que un sistema es coherente, ya que también debe existir una derivabilidad deductiva donde las creencias están conectadas a través del principio de cierre. Este principio sostiene que, si se está justificado para creer en un conjunto de proposiciones P , y también se cree que P implica lógicamente otra proposición Q , entonces se justifica la creencia en Q . No obstante, mantener la consistencia y el cierre deductivo implica un grado de conocimiento casi omnisciente de todas las creencias y sus implicaciones lógicas, algo prácticamente imposible de alcanzar para cualquier persona (Cherniak, 1984; Murphy, 2006). Esto sugiere que, aunque el coherentismo valora la consistencia, no exige

necesariamente la consciencia plena de esta por parte del individuo; es suficiente que el sistema, en sí, sea consistente.

El requisito de que un sistema de creencias se fundamente exclusivamente en la consistencia lógica y el cierre deductivo plantea problemas importantes. Al exigir un conocimiento completo de todas las implicaciones de cada creencia, crean un sistema monótono que carece de la flexibilidad necesaria para adaptarse a nueva información o revisar creencias existentes. Tal sistema sería incapaz de aprender o evolucionar, puesto que cualquier dato nuevo contradictorio podría colapsar el sistema entero, una situación que no corresponde a cómo las personas realmente modifican sus creencias. Murphy (2006) apunta que, además, es posible mantener creencias justificadas que son lógicamente inconsistentes sin conocimiento de esa inconsistencia. Esto llevó a Bonjour (1989) a rechazar la consistencia lógica como un requisito para la coherencia del sistema de creencias. En consecuencia, se hace evidente que la coherencia de un sistema no se asegura simplemente con la consistencia lógica. Se propuso entonces que, además, debían estar conectadas por relaciones deductivas como las de consecuencia o cierre lógico. Pero también vimos que pueden darse conjuntos de creencias coherentes, pero no conectadas de manera deductiva. Esto queda ejemplificado con el conjunto de creencias sobre el robo (visto más arriba). Esto indica que la coherencia podría requerir de relaciones adicionales que no se basen únicamente en la lógica deductiva. Es posible que la consistencia lógica no sea necesaria, pero que sí lo sea cierto tipo de consistencia. Después de todo, un sistema que se reduce al absurdo es completamente inútil.

3.2.3.2. Propuesta probabilístico-inductiva

La propuesta probabilístico-inductiva introduce la consistencia probabilística como alternativa a la rigidez de la lógica deductiva, enfatizando una perspectiva de grados de creencia más flexible y adaptado a la incertidumbre. En este marco, la verdad de una proposición no se evalúa en términos absolutos sino como una probabilidad en una escala de 0 a 1, permitiendo que las creencias se ajusten en función de evidencias y contexto sin demandar una certeza absoluta. Por ejemplo, la creencia en la probabilidad de lluvia basada en la observación de nubarrones puede ser de 0.7, reflejando una inclinación hacia la consideración de verdad de esa proposición pero con ciertas reservas, reconociendo la posibilidad de que no sea así (0.3). Esto mostraría el juego entre certeza e incertidumbre que se produce en las creencias, que siempre están condicionadas por la evidencia disponible (p.ej., observaciones o conocimientos). Este modelo de creencias también se adapta al contexto circundante, como en la percepción de que alguien es alto o bajo basándose en la estatura media de los de mi alrededor: no diré lo mismo de los españoles si soy de Guatemala que si soy de Suecia. La probabilidad permite una gestión más dinámica de las contradicciones y la variabilidad en la aceptación de creencias.

La consistencia probabilística en la teoría de la coherencia va más allá de ajustar contradicciones aparentes y se utiliza para modelar relaciones de apoyo mutuo entre creencias mediante el cálculo de probabilidades. A diferencia de la deducción directa de una *proposición* a partir de otra, esta propuesta permite inferir *credibilidades* de unas creencias a otras, permitiendo así la justificación, revisión y actualización de creencias basándose en nuevas evidencias. Por

ejemplo, una creencia puede cambiar con el tiempo a medida que se acumulan pruebas (modelo bayesiano), ilustrando cómo una creencia en un tiempo momento t_1 ($Pt_1(x) = r$) puede evolucionar en un tiempo t_2 tras la consideración de nueva evidencia ($Pt_2(x | e) = s$). Esta idea también es apreciada por los fundacionalistas moderados.

3.2.3.3. *Propuesta de las relaciones explicativas*

Esta propuesta sugiere que para que un conjunto de creencias sea coherente, esas relaciones deductivas e inductivas que se dan en el sistema deben ser, además, explicativas. Estas relaciones no solo conectan las creencias entre sí, sino que también proporcionan una justificación comprensible de cómo una creencia puede apoyar a otra. Por ejemplo, en el ejemplo del caso de robo, la coherencia entre las creencias que implican a 'Juan', se logra al formular una hipótesis explicativa que conecta diversas evidencias, como su presencia en la escena y la posesión de objetos relacionados con el crimen y el ingreso bancario. Es una inferencia a la mejor explicación. Por otro lado, un conjunto de creencias dispares como "mañana es martes", "los lobos son mamíferos" y "soy una taza" no constituye un sistema coherente ya que no existe una explicación subyacente que relacione estas afirmaciones de manera significativa. Esto impide que tales creencias formen un sistema coherente de justificación.

En el marco de las relaciones explicativas, siempre se requieren dos componentes: el *explanans*, que es la explicación o causa, y el *explanandum*, que es el fenómeno que se explica. Las relaciones explicativas suelen abordar preguntas de "¿por qué?", proporcionando una comprensión causal de los fenómenos en cuestión. Para justificar coherentemente una creencia, es necesario mostrar cómo ciertos hechos o causas hacen que el fenómeno explicado sea esperable o comprensible. Hempel (1967) definió la explicación ideal como la deducción de un hecho particular a partir de leyes generales y condiciones específicas, un proceso llamado subsunción. Sin embargo, su modelo ha sido criticado por admitir inferencias que no siempre son intuitivamente aceptables como explicaciones y omitir otras que sí podrían ser válidas. Hempel argumentó que una explicación debe incluir al menos una ley general y un hecho particular como condición antecedente, pero esta estructura a menudo resulta demasiado rígida para abarcar la diversidad y complejidad de las explicaciones válidas en contextos reales.

La transición de esta versión deductiva a la inductiva o probabilística en las relaciones explicativas es legítima en tanto que la explicación también responde a la pregunta de cuán probable es el *explanandum* dado el *explanans*. A diferencia de la deducción, donde el *explanandum* es completamente esperable, la inducción solo asegura que el es probable. En este contexto, un hecho particular (el *explanandum*), se respalda mediante una ley probabilística general (por ejemplo, $P(A | B)$) y hechos particulares como condiciones antecedentes. Un ejemplo claro es la justificación de la creencia C de que "Luís tiene cáncer de pulmón" basada en la propiedad A de "desarrollar cáncer de pulmón" y la condición B de "haber fumado muy larga e intensamente". La explicación, entonces, toma la forma $P(A | B) \& B$, lo que lleva a concluir C , siempre que se conozca al menos uno de los elementos del *explanans*. Este modelo de justificación considera el contexto cognitivo del sujeto, permitiendo que, si posee cualquier par de estas tres creencias (A , B o C), pueda

inferir la tercera para mantener la coherencia del conjunto. Según Olsson (2017), algunos consideran que cuando las creencias se comportan así, se hacen más propensas a ser correctas, aumentando su justificación. En estos casos, las relaciones explicativas se interpretan como relaciones inductivas probabilizadoras en última instancia. Esta interpretación va más allá de la simple lógica formal, incorporando elementos de probabilidad y contexto para evaluar la justificación de un sistema de creencias.

Pero la asunción coherentista de que las relaciones explicativas son simétricas, donde A y B se explican y justifican mutuamente, es muy controvertida. Más aún cuando esas relaciones explicativas involucran causalidad. Si A es la causa de B, B no lo puede ser de A. Las relaciones causales son inherentemente asimétricas (Díez y Moulines, 1999). En general, se considera que una buena explicación para un conjunto de creencias o proposiciones debe tener un algo poder predictivo, proporcionar información causal sobre su ocurrencia, ser simple y, a su vez, ampliamente conectiva dentro del sistema. Debe proporcionar un alto grado de integración con el menor número de supuestos, maximizando su alcance explicativo.

Se puede observar que el coherentismo tiene ciertas similitudes con el fundacionalismo, ya que ambos marcos teóricos utilizan relaciones lógico-deductivas, probabilístico-inductivas y explicativas para estructurar y justificar sistemas de creencias, aunque con distinto énfasis en la coherencia y la fundamentación de las relaciones entre creencias.

CAPÍTULO 4

Fundamentos de la psicología cognitiva para el estudio de la credibilidad

A lo largo de las páginas que siguen se expondrán los principales componentes psicológicos implicados en el estudio de la dinámica de credibilidades, objeto de esta tesis. Este capítulo enlaza con los anteriores, desarrollando la evolución del abordaje logicista y formalista en los estudios de la cognición humana, en contraposición a una perspectiva que considera que los cómputos y procesos mentales son sensibles al contexto y al contenido de las representaciones mentales.

En este sentido, este capítulo constituye un recorrido por el devenir conceptual de la disciplina, mostrando las razones que inicialmente llevaron a considerar que el pensamiento consistía en una lógica formal encarnada, algo incompatible con una psicología de las credibilidades. Asimismo, se intentarán mostrar las razones que llevaron a un abandono paulatino de esa perspectiva, poniendo el énfasis en el razonamiento cotidiano, hasta llegar a lo que hoy se considera 'el Nuevo Paradigma' probabilístico del razonamiento. En esta exposición, la relación entre la argumentación y el razonamiento cotidiano, en virtud de la investigación empírica sobre los refutadores o rebatidores, adquiere una relevancia principal.

1. Contexto cultural en los albores de la psicología del pensamiento

Los inicios de la psicología del razonamiento estuvieron profundamente influenciados por las teorías de la lógica formal. Helmholtz en 1866 postulaba que el conocimiento humano operaba de manera silogística, realizando inferencias inconscientes para formular conclusiones a partir de sensaciones actuales y conocimientos previos almacenados (Rodríguez y Rodríguez, 2000). Sin embargo, estos supuestos no se derivaban directamente de los avances en lógica formal ni se puede decir que él se estuviera refiriendo propiamente a ningún tipo de razonamiento. Generalmente, se reconoce a Jean

Piaget como el primer psicólogo que combinó formalmente la lógica con el estudio del razonamiento, estableciéndose como una figura fundamental en la historia de la psicología del razonamiento según varios autores como Garnham y Oakhill (1996), Holyoak (2002), y Johnson-Laird (2008).

Después de un largo periodo en el que la lógica y el estudio del razonamiento se mantuvieron separados, George Boole, en el siglo XIX, reunió de nuevo ambos campos al argumentar que la lógica formal establece las normas del razonamiento humano, tratando la lógica como un estudio de las leyes esenciales de las operaciones mentales involucradas en el pensamiento. Sin embargo, hacia finales de ese siglo y principios del XX, Frege y Russell lideraron una nueva separación entre la lógica y la psicología (Bonatti, 1998). Frege enfatizó que la lógica se ocupa de principios universales, mientras que la psicología trata aspectos subjetivos y personales. Russell veía la lógica principalmente como una herramienta para la investigación ontológica, apartada del subjetivismo psicológico. En Estados Unidos, el auge del neopositivismo del siglo XX llevó la psicología hacia el conductismo, centrado en aspectos observables y mensurables de la conducta, excluyendo la consideración de procesos mentales internos. Esta orientación conductista prevaleció en EE. UU. hasta los años 1960, definiendo la conducta humana como una respuesta directa a estímulos externos, sin intervención de procesos mentales internos.

2. Relaciones entre la lógica y el pensamiento en la obra de Piaget

En contraste con la tendencia estadounidense, Jean Piaget en Europa defendió la idea de que las reglas lógicas son fundamentales para el razonamiento humano, centrándose en cómo se desarrolla cognitivamente el niño. Según Hanna (2006), los trabajos iniciales de Piaget exploraron la lógica formal clásica, fundamentando su teoría del desarrollo mental en las ideas de Boole y los "Principia Mathematica" de Russell y Whitehead. Piaget describió cómo la mente evoluciona desde un estado sensoriomotor, centrado en el presente inmediato, hasta alcanzar las "operaciones formales" a los 12 años, donde el individuo puede realizar razonamientos lógicos con conceptos abstractos y no presentes en la experiencia directa inmediata. Estas fases reflejan un aumento en la complejidad del "aparato lógico" de la mente, que al final se equipara con la capacidad lógica de un adulto, considerado como un "motor lógico perfecto".

A pesar de su aproximación a una "lógica mental", Piaget no delineó completamente las estructuras lógicas específicas o las reglas de inferencia utilizadas, un aspecto que Bonatti (1998) señala como una limitación en su obra. Sin embargo, las "operaciones formales" de Piaget se comparan con las teorías del lenguaje del pensamiento propuestas por lingüistas como Chomsky y filósofos como Fodor (Hanna, 2006), sugiriendo similitudes entre los modelos de desarrollo cognitivo y teorías lingüísticas.

3. La teoría de la demostración y de la deducción natural

George Boole y David Hilbert jugaron papeles fundamentales en la conceptualización de las leyes del pensamiento como procesos matemáticos y lógicos. Boole trató el razonamiento humano como una composición de símbolos según reglas matemáticas, viendo los símbolos como entidades arbitrarias manipuladas bajo leyes comunes a la lógica y la matemática. Siguiendo esta línea, Hilbert en la década de 1920 enfatizó que el pensamiento operaba mediante transformaciones simbólicas guiadas por reglas sintácticas, lo que consideraba esencial para entender las operaciones del entendimiento humano.

Sin embargo, Hilbert reconoció que su sistema podía ser demasiado abstracto para modelar directamente el funcionamiento mental. Por ello, Gerhard Gentzen en 1934 intentó hacer la teoría más accesible eliminando los axiomas y enfocándose en mejorar las reglas de inferencia para reflejar más fielmente el razonamiento humano real. Esto llevó al desarrollo del sistema de deducción natural, que, aunque no era un modelo psicológico *per se*, estableció un puente crucial entre la lógica formal y la psicología del razonamiento, influyendo en las teorías posteriores sobre la Lógica Mental propuestas por teóricos como Braine (1978) y Rips (1994).

4. Turing y las máquinas que piensan

En la década de 1940, Alan Turing avanzó en las intuiciones de Hilbert al desarrollar un modelo teórico que mostraba cómo una máquina podía realizar funciones anteriormente consideradas exclusivas de la mente. Este modelo teórico, que operaba según la forma de los símbolos implementados, sugirió que la mente podría conceptualizarse como un procesador de información, funcionando con algún tipo de lógica sobre distintos sustratos físicos.

Aunque Turing no especificó qué tipo de lógica empleaban los humanos, propuso que diferentes soportes físicos, configurados adecuadamente, podían replicar respuestas humanas mediante procedimientos mecánicos (en referencia al conocido 'test de Turing'). Más tarde, esta idea de que el sistema de cómputo es relativamente independiente del sustrato físico sobre el que se implementa contribuyó a la conceptualización de la mente en términos de *realizabilidad múltiple*. Las ideas de Turing fueron fundamentales para el desarrollo posterior de la teoría de la computabilidad, las ciencias de la computación y la inteligencia artificial, así como para la revolución cognitiva, que se apoyó en la metáfora del ordenador y las gramáticas chomskianas que derivaron en la Teoría Computacional de la Mente. Esta revolución subrayó que los procesos de pensamiento podrían ser independientes del sustrato físico que los ejecuta.

5. Ciencias de la computación y la metáfora del ordenador

La metáfora del ordenador en las ciencias de la computación describe a los ordenadores como sistemas que manipulan símbolos formalmente, siguiendo reglas sintácticas sin considerar el significado o valor de verdad (Cantwell-Smith, 2003). Esto implica que pueden realizar inferencias y llegar a conclusiones válidas siguiendo reglas lógicas sin comprender el contenido semántico. La "condición de formalidad," planteada por Jerry Fodor (1975), establece que un sistema puede manipular símbolos únicamente según su estructura, sin involucrarse con su significado. Esta metáfora fue fundamental para entender cómo un mecanismo físico puede procesar información de forma lógica, emitiendo respuestas observables sin perder la coherencia semántica.

Este enfoque permitió un nuevo modelo para la psicología, alejándose del conductismo al centrarse en las reglas formales que gobiernan los procesos mentales internos. Los sistemas de reglas de producción y redes semánticas adoptaron estos principios para replicar la metáfora del ordenador en las representaciones e inferencias, demostrando cómo la manipulación formal de símbolos puede generar respuestas coherentes. Sin embargo, esta perspectiva formalista puede tener limitaciones al intentar capturar la riqueza y complejidad de las inferencias humanas.

6. Las aportaciones de las gramáticas de Chomsky a la psicología cognitiva

Las obras de Noam Chomsky de finales de la década de 1950 influyeron enormemente en la psicología cognitiva y fueron cruciales para superar el conductismo y vincular el formalismo logicista con esta disciplina. En *Estructuras Sintácticas* (1957) y *A Review of B. F. Skinner's Verbal Behavior* (1959), Chomsky argumentó que el lenguaje no se aprende simplemente asociando palabras adyacentes. En cambio, requiere operaciones complejas sobre estructuras abstractas presentes en la mente de los hablantes. Chomsky mostró cómo la estructura sintáctica del lenguaje podía representarse a través de un sistema de lógica formal deductiva, al que denominó gramática formal. Este sistema, similar a un sistema axiomático, contiene estructuras primitivas y reglas derivacionales que permiten generar un número infinito de oraciones a partir de un conjunto finito de componentes. Según esta gramática universal generativa, las propiedades comunes de todas las lenguas reflejan una facultad específica de la mente: la facultad del lenguaje.

La facultad del lenguaje, según Chomsky, tiene una estructura separada de la razón y determina las características de las lenguas humanas. Por lo tanto, las propiedades de los idiomas existentes reflejan esta facultad y proporcionan información sobre la mente humana. No obstante, la estructura del lenguaje no es equivalente a la totalidad de la mente, solo es un aspecto particular de su funcionamiento. Estas ideas mostraron que el estudio del lenguaje podía beneficiarse del análisis formalista, sugiriendo la posibilidad de explicar las estructuras mentales mediante modelos lógicos rigurosos.

7. Teoría Computacional de la Mente (TCM)

Las ciencias de la computación y la lingüística de Chomsky estuvieron influenciadas por el racionalismo, el logicismo y el formalismo de la teoría de la demostración. Estas disciplinas proporcionaron los cimientos para la Teoría Computacional de la Mente (TCM), que se convirtió en el pilar de la psicología y las ciencias cognitivas (Holyoak, 2002). Según sus defensores, "no sería descabellado describir la Ciencia Cognitiva Clásica como un intento extendido de aplicar los métodos de la teoría de la demostración a la modelización del pensamiento" (Fodor y Pylyshyn, 1988, pp 29-30).

7.1. Concepto general

La TCM propone que la cognición humana se basa en procedimientos mentales que procesan representaciones abstractas intencionales, siguiendo las reglas de un lenguaje del pensamiento. Estas representaciones, que incluyen conceptos, palabras, imágenes, olores, todo, se estructuran en un formato proposicional, lo que permite a la mente operar mediante secuencias guiadas por la sintaxis en lugar de la semántica (Cela-Conde y Marty, 1998). De acuerdo con la TCM, el pensamiento se representa en un lenguaje del pensamiento, y los procesos mentales siguen reglas derivacionales que, basadas en la sintaxis, generan fórmulas bien formadas (Horst, 2002). Este mecanismo asegura la coherencia semántica sin que el sistema tenga que comprender el significado de los símbolos. Esto es similar a un silogismo: un sistema físico puede llegar a una conclusión válida sin entender el significado de "hombre" o "mortal," solo aplicando reglas formales.

Jerry Fodor argumenta que esta propiedad permite explicar la cognición de manera mecanicista, sin necesidad de una conciencia que interprete el contenido, evitando el problema del homúnculo. Así, el pensamiento automático puede producir respuestas racionales sin necesidad de entender el significado. Para Pujadas-Torres (2002), las diferencias de contenido en las actitudes proposicionales deben reflejarse en las representaciones, de modo que influyan en la conducta a través de diferencias funcionales y causales en su manipulación sintáctica. De esta manera, se preserva la influencia de las actitudes en el comportamiento, sin que la comprensión del contenido sea un requisito explícito.

7.2. Lenguaje del pensamiento y composicionalidad

La TCM sostiene que el pensamiento organiza símbolos siguiendo reglas sintácticas, una idea influenciada por la teoría de la demostración de Hilbert y la lingüística de Noam Chomsky. Chomsky demostró que un conjunto finito de unidades léxicas, estructuras sintácticas y reglas transformacionales puede generar infinitas oraciones gramaticalmente correctas, otorgando al lenguaje una semántica composicional que permite expresar significados complejos (Krifka, 2003). La TCM introduce el concepto de *lenguaje del pensamiento* (o mentalés), que establece un paralelismo entre el lenguaje natural y el pensamiento. Los defensores argumentan que la representación mental es composicional, al igual que la lingüística, donde el significado de una oración se deriva de las palabras y su organización. De manera similar, los pensamientos se

construyen a partir de conceptos y su disposición en el mentalés. Fodor (1981) sostiene que el significado y la intencionalidad de los estados mentales (i.e., según la TCM, de las actitudes proposicionales) proceden del significado e intencionalidad de los símbolos que constituyen el lenguaje del pensamiento. Este sistema es altamente generativo, permitiendo la composición de todos los pensamientos posibles, similar a cómo las reglas gramaticales generan infinitas oraciones.

Sterelny (2002) argumenta que el lenguaje del pensamiento permite que las personas formen opiniones sobre el mundo, creen nuevas representaciones, viajen mentalmente en tiempo o espacio, y contemplen escenarios contrafactuales que influyan en la toma de decisiones, lo que a su vez permite explicar la acción racional.

8. Críticas dirigidas al formalismo y al logicismo de la teoría computacional de la mente

La TCM ha sido elogiada y criticada a partes iguales. Las críticas, en su mayoría, apuntan a la insuficiencia de la computación para explicar el pensamiento, subrayando la importancia de los contenidos semánticos para una explicación de la mente cabal desde la psicología científica.

8.1. No hay conocimiento sin contenido

Pylyshyn (1973) identifica que el modelo computacional de la mente genera conocimiento mediante la relación sintáctica de elementos proposicionales, pero no garantiza su verdad, la cual es esencial para el conocimiento. Según Pujadas-Torres (2002), Si la referencia y la verdad son nociones semánticas, y el sistema no tiene acceso a los contenidos semánticos, no se puede saber si produce conocimiento, ni se puede hablar, en rigor, de una psicología cognitiva computacional del conocimiento.

Searle (1980, 1990) sostiene que una semántica basada en la intencionalidad de los símbolos sin un intérprete es una explicación circular. Argumenta que las representaciones en ordenadores no son simbólicas ni sintácticas por sí mismas, sino que estas propiedades dependen de los usuarios. El *argumento de la habitación china* de Searle concluye que la computación no es suficiente para explicar el pensamiento ni el conocimiento, ya que no tiene acceso a los contenidos de las ideas. El producto final de la cognición humana consiste en saber o entender, y los ordenadores no entienden nada.

8.2. La semántica composicional ignora el contexto

Otra crítica importante hacia la TCM es que las operaciones formales de la computación son independientes del contexto. Los procesos formales de la composicionalidad generan una interpretación local o modular de la semántica, donde el significado de una proposición p depende solo de p (sus componentes y estructura sintáctica), ignorando el contexto en el que aparece. Además, se supone que los constituyentes mantendrán su significado en distintas representaciones, lo cual no está garantizado. Los significados dependen tanto de las unidades de análisis como de su

interacción con los contextos lingüísticos, cognitivos, físicos y sociales (Burge, 1986; Krifka, 2002; Putnam, 1975; Van Dijk, 1977, 2008). Esta interacción contextual resulta crucial para modular la credibilidad, como se destacará en la propuesta teórica de esta tesis (sec. 1.3., cap. 1, pt. I). Según Pujadas-Torres (2002), una teoría del contenido debe considerar esta interacción contextual.

La TCM, al ser una teoría lógico-matemática, afirma que, aunque las representaciones mentales son intencionales porque se refieren a estados del mundo, por lo demás funcionan igual que los números. Sin embargo, el hecho de referirse a un mundo en constante cambio puede marcar la diferencia en cómo se entiende la interacción entre las representaciones mentales y el contexto.

8.3. Formatos representacionales alternativos al proposicional

La TCM sostiene que las operaciones formales del lenguaje del pensamiento requieren que todas las representaciones mentales adopten un formato proposicional común, a-modal e inaccesible para la introspección directa. Según esta teoría, al imaginar un triángulo, nuestro sistema recupera su descripción proposicional, obteniendo las instrucciones para generar la imagen en la memoria de trabajo, siendo el resultado lo que percibimos conscientemente.

No obstante, el debate sobre el formato representacional está lleno de controversias tanto filosóficas (Camp, 2007; Fodor y Pylyshyn, 1981; Rescorla, 2009) como psicológicas (Paivio, 1971; Barsalou, 2008; Kosslyn, 1980; Zwaan, 2004). Durante un tiempo, la discusión se centró en si, además del formato proposicional, podía existir un formato imaginístico (Kosslyn, 1980, 1994). Sin embargo, a partir de las décadas de 1980 y 1990, surgieron modelos alternativos que perfeccionaron esta idea, proponiendo un formato mental basado en modelos mentales multimodales, accesibles a través de procesos de simulación mental con diversos grados de abstracción (Johnson-Laird, 1983; Van Dijk y Kintsch, 1983). Estas propuestas no rechazan la TCM, sino que abogan por formatos no proposicionales en el mentalés, con una visión pluralista para la computación.

Debido a la importancia de la simulación mental, estas representaciones derivaron en lo que ahora se conoce como representaciones corporeizadas (Barsalou, 2008; Glenberg y Kaschak, 2002; Solomon y Barsalou, 2001; Zwaan y Yaxley, 2003; Zwaan, 2004, 2016). Esta perspectiva cuestiona fuertemente la existencia de las representaciones proposicionales, subrayando que hay evidencia experimental que apoya la idea de que almacenamos representaciones multimodales (Conway, 1990, 2002; Rubin, 2006; Schacter, Dobbins y Schnyer, 2004), pero no hay evidencia concluyente que respalde representaciones a-modales (Barsalou, 2008).

8.4. Inverosimilitud de la teoría de la demostración como explicación de la cognición

La crítica desarrollada por Oaksford y Chater (1998) es crucial para entender el cambio en la investigación de la psicología del razonamiento y el paradigma probabilístico emergente. Además, es fundamental para el modelo teórico que se presenta en esta tesis. Fodor y Pylyshyn (Fodor, 1987; Fodor y Pylyshyn, 1988; Pylyshyn, 1984) sostienen que la Teoría Computacional de la Mente (TCM) debe respaldar la psicología popular en su capacidad para explicar la conducta de las personas y

hacerlo de forma científica. La psicología popular atribuye deseos y creencias a los individuos para explicar su comportamiento, asignando así actitudes proposicionales. El comportamiento surge de inferencias formales sobre estas representaciones, las cuales son sensibles a la forma, pero no al contenido, manteniendo la coherencia semántica basada en valores de verdad. Para lograr esto, las inferencias de la TCM deben ser deducciones válidas conforme a la teoría de la demostración, con premisas verdaderas (Oaksford y Chater, 1998). Según Fodor y Pylyshyn, la TCM está basada en la teoría de la demostración, la cual se apoya en la lógica deductiva clásica. Esta lógica es bivalente, ciega a los contenidos, y monótona. La monotonía implica que, si una conclusión se deriva de un conjunto de premisas, siempre se seguirá esa conclusión, sin importar la información adicional, a menos que esta contradiga por completo el conjunto de premisas original.

El contenido de las actitudes proposicionales se relaciona con objetos del mundo, como mesas, nubes, sentimientos y lluvia, así como con las relaciones cotidianas, como las causales, temporales o sociales. Estos contenidos se refieren a un mundo cambiante, del que solo se tiene un conocimiento limitado, incierto y revisable. Lo que parecía cierto ayer puede parecer falso hoy, lo que significa que la información adicional puede invalidar conclusiones previas. Las inferencias hechas sobre estos contenidos parten de premisas con una semántica inestable, que no se limita a valores absolutos de verdadero o falso. En cambio, esta semántica es graduada, oscilando entre ambos extremos. Por ello, las conclusiones siempre pueden ser desafiadas, lo que hace que las inferencias en la psicología popular sean no-monótonas y refutables.

La mente aborda problemas ambiguos en condiciones de incertidumbre, adaptándose a las variaciones contextuales y utilizando el conocimiento disponible para hacer inferencias rápidas sobre objetos y relaciones cotidianas. Dado que las inferencias refutables son comunes, Oaksford y Chater (1998) argumentan que la TCM debe ser capaz de acomodarlas para mantenerse como una teoría psicológica viable. Para sostener los postulados logicistas, la TCM debería incorporar un lenguaje formal compatible con la teoría de la demostración, basado en una lógica no-monótona que pueda capturar la gama de inferencias permitidas por el sentido común, con su semántica graduada.

En el campo de la representación del conocimiento en la inteligencia artificial (IA), se ha intentado abordar las inferencias derrotables mediante el desarrollo de lógicas no-monótonas. Según Oaksford y Chater (1998, 2007), estas lógicas solo cubren un rango de operaciones limitado, y lo hacen a un costo de procesamiento demasiado elevado. Por lo tanto, no alcanzan a abarcar el repertorio de inferencias que los humanos pueden hacer con facilidad. Además, la mayoría de los algoritmos son intratables en tiempo, debido a su dependencia de la resolución de problemas de explosión combinatoria, como la verificación de la consistencia del sistema para cada operación. Así, la cognición no puede reducirse a una lógica formal implementada como la teoría de la demostración.

La complejidad de estos algoritmos implica una crítica a la múltiple realizabilidad de lo mental que postula la TCM. Oaksford y Chater (1998) sostienen que una teoría psicológica debe tener en

cuenta el sustrato biológico donde se implementan los algoritmos para reflejar el tiempo real necesario para los procesos mentales. Los algoritmos deben ser compatibles con la velocidad de transmisión de los impulsos nerviosos. Por lo tanto, el tipo de hardware influye en los algoritmos posibles, ya que algunos no son manejables en tiempo real.

En consecuencia, es improbable que la teoría de la demostración logre capturar el rango de inferencias derrotables del sentido común y, de ser posible, es dudoso que pudiera implementarse en un hardware biológico. Es también cuestionable que puedan desarrollarse algoritmos manejables que instancien la teoría de la demostración para la mayoría de las tareas psicológicas. Estas tareas dependen de la aplicación del conocimiento previo del mundo, el cual complementa y elabora la información disponible. Esto aplica tanto para el razonamiento inductivo y abductivo, como para el condicional. Asimismo, las inferencias implícitas en los procesos de comprensión del discurso, la categorización basada en relaciones de semejanza, estructuras prototípicas y la derrotabilidad del conocimiento semántico en los procesos perceptivos refuerzan la crítica hacia la lógica formal.

8.5. Acerca del alcance de la TCM

Los defensores de la Teoría Computacional de la Mente admiten que esta teoría tiene un alcance limitado. Se centra en un ámbito restringido de la cognición, principalmente en los *sistemas de entrada* que son locales o modulares, encapsulados e independientes. Esto incluye la percepción visual y el procesamiento sintáctico del lenguaje. La TCM no abarca los procesos cognitivos superiores, como el razonamiento, la comprensión del discurso o las creencias. Estos se encuentran en *sistemas centrales* o globales, que, según Fodor, son informacionalmente penetrables, holísticos, interconectados, isotrópicos y quineanos, y según otros autores como Dreyfus (1978), Winograd y Flores (1986), no pueden formalizarse.

Según Fodor (1987), los procesos mentales superiores, debido a sus características, son inaccesibles a la verificación empírica. Cuanto más global e interconectado es un proceso con otros aspectos de la cognición, más difícil resulta entenderlo. Así, para él, la psicología científica debe limitarse a explicar los procesos modulares, dejando de lado cuestiones semánticas y pragmáticas. Cela-Conde y Marty (1990) sugieren que a medida que un proceso se vuelve más global, más difícil es explicarlo con precisión desde la TCM. Pero añaden que eso es un problema de la TCM, no de la ciencia cognitiva. Fodor admite que la TCM no puede explicar plenamente los procesos centrales como el razonamiento, la comprensión del discurso o la formación de creencias. Sin embargo, esto no significa que estos procesos no sean computacionales o que no utilicen el lenguaje del pensamiento. Fodor cree que toda cognición es computacional y se lleva a cabo a través de un tipo de lenguaje del pensamiento que podría ser muy diferente para los procesos no modulares.

9. Consecuencias de la TCM: el programa logicista en el razonamiento

La Teoría Computacional de la Mente se asienta en la comprensión de la mente como una entidad lógico-formal. Esto permitió ver la mente como un sistema que procesa información de forma similar a un ordenador, controlada por un programa de instrucciones simbólicas sin necesidad de ningún componente misterioso. El advenimiento de los ordenadores digitales dotó de plausibilidad a la hipótesis de que el cerebro funciona como un sistema lógico, y el pensamiento se realiza por un lenguaje formal, simbólico y proposicional (Fodor, 1975) o, de forma equivalente, mediante una *lógica mental* (Braine, 1978; Rips, 1983). Esta lógica mental es considerada por autores como Fodor (1975, 2000), Dvitt y Sterenly (1987), y Pylyshyn (1984) como fundamental para explicar la coherencia semántica en el razonamiento y cómo la mente preserva la verdad en sus secuencias de pensamiento.

La lógica mental también permite el desplazamiento mental en tiempo y espacio, la sistematicidad y productividad del pensamiento (Fodor, 1987; Fodor y Pylyshyn, 1988), así como la capacidad de razonar sobre categorías y clases (Fodor, 1981). En general, esta lógica es vista como la única forma posible de comprender las reglas subyacentes del razonamiento (Waskan, 2003), funcionando como una teoría de la competencia que define la capacidad de las personas para razonar correctamente, cuando están en condiciones de evitar errores. El programa logicista en la psicología del razonamiento se consolidó al entender la lógica mental como una versión de la lógica proposicional derivada del sistema de deducción natural de Gentzen. Esta perspectiva se fundamentaba en las tesis de Piaget, la teoría de la demostración de Hilbert y Gentzen, y el lenguaje del pensamiento de Fodor. Los modelos más representativos fueron propuestos por Braine (1978, 1990; Braine y O'Brien, 1991, 1998) y Rips (1983, 1994), quienes asumían que el razonamiento humano opera mediante cálculos lógicos sobre representaciones simbólicas, con un sistema de deducción natural implementado mentalmente.

La psicología cognitiva se centró en descubrir la lógica formal subyacente en la mente y las reglas de inferencia que se usaban. Sin embargo, durante el desarrollo de esta teoría surgieron datos empíricos difíciles de encajar con esta perspectiva. Los experimentos de Peter Wason sobre la *tarea de selección* (Johnson-Laird y Wason, 1970; Wason, 1966; Wason y Johnson-Laird, 1972) revelaron que la gente mostraba un razonamiento sistemáticamente irracional en algunas situaciones, desviándose de las normas de la lógica. Oaksford y Chater (2007) señalaron que estos resultados fueron ignorados o relegados por los teóricos cognitivos, que interpretaron los errores como problemas de interpretación o dificultades para acceder a las reglas lógicas mentales (Braine, Reiser y Rumin, 1984; Rips, 1994). Así, se intentó preservar la noción de que el razonamiento lógico subyacía al pensamiento humano.

Los primeros estudios de Peter Wason influyeron en algunos psicólogos que abandonaron la lógica formal como explicación del razonamiento humano. Originalmente, la tarea de selección utilizaba contenidos puramente formales, como letras y números. Sin embargo, al cambiar a contenidos cotidianos, los participantes mostraron un rendimiento mucho mayor, cometiendo

menos errores, razonando mejor. Este cambio reveló que el razonamiento humano no podía explicarse completamente a través de la lógica formal, que no considera el contenido de las afirmaciones. Estos resultados han sido interpretados como una evidencia de que el sistema cognitivo humano es sensible a los contenidos en sus cómputos de razonamiento y toma de decisiones. Esto dio lugar a modelos alternativos como los esquemas de razonamiento pragmático de Cheng y Holyoak (1985) y la teoría de los Modelos Mentales de Johnson-Laird (1983, 2008).

A continuación, se discutirá la evolución de la psicología del razonamiento en relación con su progresivo abandono del programa logicista y la apertura a la consideración de la gradualidad en la fortaleza de las conclusiones debido al acceso a refutadores en la memoria. Se observará cómo este cambio progresivo de perspectiva hacia el estudio de un razonamiento más cotidiano se relaciona con la argumentación y, finalmente, cómo se adecúa al estudio de las creencias.

10. La lógica mental

Durante los años 70, la perspectiva logicista en psicología se consolidó, y la teoría de la demostración basada en el sistema de deducción natural de Gentzen se estableció como hipótesis estándar para entender la mente en el campo de la psicología del razonamiento. Estas teorías, centradas en la noción de prueba, argumentan que un razonamiento es válido si su conclusión se deriva de las premisas aplicando reglas formales secuenciales, centrándose sólo en la estructura simbólica.

Los primeros modelos psicológicos de la deducción natural fueron propuestos por Henle (1962), Osherson (1975) y Johnson-Laird (1975). Sin embargo, estos se concentraron más en la implementación algorítmica que en la plausibilidad psicológica, sin datos empíricos sólidos. Luego, Braine (1978) y Rips (1983) avanzaron al investigar empíricamente qué reglas formales de inferencia usaban las personas, identificando una lógica mental específica. Aunque su orientación se vincula explícitamente con la deducción natural, Braine y Rips reconocen que la lógica mental puede divergir de la lógica formal tanto teórica como empíricamente.

10.1. Sistema de lógica mental de Braine y colaboradores

El modelo de lógica mental, desarrollado por Braine y colaboradores, propone que el pensamiento humano opera mediante un "lenguaje del pensamiento" lógico e innato (Braine y O'Brien, 1998). Según esta perspectiva, la memoria declarativa está compuesta de proposiciones lingüísticas representadas mediante símbolos discretos, manipulados por la sintaxis de este lenguaje. Así, permite vincular propiedades con entidades de forma lógica, manteniendo características semánticas a través de disyunciones, conjunciones, condicionales y negaciones (O'Brien, Roazzi, Athias, Días, Brandão y Brooks, 2003). El sistema emplea la notación de la lógica de primer orden para traducir el lenguaje natural en conceptos almacenados en la memoria semántica y episódica. Las premisas se abstraen en estructuras formales que se almacenan en la memoria de trabajo,

donde se aplican reglas de inferencia iterativamente para derivar conclusiones. En un último paso, se restaura el contenido de las premisas.

Desde una perspectiva evolucionista, O'Brien et al. (2003) argumentan que el lenguaje del pensamiento no verifica la consistencia de las proposiciones en la memoria declarativa, apartándose así de los preceptos de la lógica. Nuestros ancestros no albergaban muchas proposiciones y no tenían la habilidad de construir pruebas complejas de consistencia, por lo que el conjunto de esquemas de inferencia debe ser reducido. Sostienen que los esquemas de inferencia en la lógica mental deben ser simples y ejecutarse rápidamente, centrados en estructuras como disyunciones, conjunciones, condicionales y negaciones. Distinguen entre razonamientos simples, automáticos y universales (esquemas centrales), y razonamientos complejos que no son universales. Los esquemas centrales, innatos, se aplican automáticamente cuando las premisas están presentes en la memoria de trabajo (Braine et al., 1984; Braine y O'Brien, 1998).

Para validar este modelo, Braine y sus colaboradores realizaron experimentos en los que los participantes evaluaron la veracidad de conclusiones a partir de premisas. Los argumentos con esquemas centrales resultaron en errores mínimos, mientras que los argumentos con esquemas complejos presentaron más errores (Braine et al., 1984). Experimentos de "naming" mostraron que las inferencias se realizan automáticamente durante el procesamiento del discurso, reforzando la validez de los esquemas centrales en el razonamiento humano (Lea, 1995; Lea, O'Brien, Fisch, Noveck, y Braine, 1990).

10.2. Sistema de lógica mental de Rips

El sistema de Lógica Mental desarrollado por Rips (1994), llamado PSYCOP (psicología de la prueba), es un modelo algorítmico diseñado para mantener la verdad en el razonamiento. Basado en el sistema de deducción natural, no incluye axiomas, pero sí reglas de inferencia para la realización de pruebas mentales. Según Rips, al evaluar un argumento, se puede probar la conclusión dada o extraer nuevas conclusiones. En ambos casos, tanto las premisas como las conclusiones iniciales se almacenan en la memoria de trabajo. Las reglas mentales de inferencia deductiva, consideradas correctas por el individuo, conectan las premisas y las conclusiones como puentes formales. Estas reglas se aplican de dos maneras: directamente, desde las premisas hacia la conclusión, o inversamente, desde la conclusión hacia las premisas. El proceso itera, escaneando la memoria de trabajo para encontrar premisas relevantes, introducir suposiciones, o añadir premisas intermedias que llenen huecos en la cadena inferencial. El sistema continúa hasta conectar las premisas con la conclusión, completando así la prueba. Si no se encuentra una conexión, el proceso se repite hasta que no queden más reglas por aplicar. Cuando se generan conclusiones sin tener una conclusión objetivo que permita inferencias inversas, PSYCOP aplica solo reglas directas. Algunas reglas, como el Modus Ponens¹⁵ (MP), se consideran primitivas, por lo que pueden aplicarse directamente en cuanto las proposiciones desencadenantes están presentes en la memoria de trabajo.

¹⁵ La regla MP establece que 'si p , entonces q ; y se da p ; debe inferirse q ' o ' $p \rightarrow q$, y p , $\vdash q$ '.

Rips reconoce las limitaciones del sistema PSYCOP, que puede fallar en obtener conclusiones válidas debido a la falta de reglas mentales apropiadas, restricciones en la memoria de trabajo o la aplicación de reglas incorrectas. A pesar de estos inconvenientes, las personas intentan llevar a cabo una prueba mental al resolver problemas. La identificación de qué reglas pertenecen al sistema se basa en observaciones empíricas. Rips se apoyó en trabajos previos (Braine, Reiser y Rumin, 1984; Johnson-Laird, 1975; Osherson, 1975; Sperber y Wilson, 1986) para realizar experimentos en los que los participantes evaluaban la validez de diferentes argumentos. Descubrió que ciertas reglas, como el Modus Ponens, se aplican sistemáticamente y de forma correcta, mientras que otras, como la introducción de la disyunción, resultan menos fiables. Por ello, algunas reglas se consideran primitivas, mientras que la identificación completa de esta base primitiva es una tarea empírica en curso.

Los sistemas deductivos de Braine, Rips y otros pueden utilizarse para analizar argumentos comunes. Sin embargo, dado que el razonamiento diario suele ser no-monótono y depender del contenido, estos modelos encuentran limitaciones para inferencias no demostrativas y creencias, lo que añade complejidad en la aplicación de la lógica mental a situaciones cotidianas.

10.3. Valoración del modelo de la lógica mental desde la presente perspectiva

A pesar de los avances significativos y las validaciones experimentales de los modelos de lógica mental de Braine y Rips, parece haber limitaciones en su capacidad para manejar el razonamiento bajo incertidumbre y, consecuentemente, el tratamiento de las creencias. Estos modelos, eficaces en contextos controlados, enfrentan limitaciones cuando la información es incompleta o ambigua (e.g. O'Brien, Braine, Connell, Noveck, Fisch, & Fun, 1989), una constante en el razonamiento humano cotidiano. La lógica mental, en su forma actual, no describe adecuadamente cómo las personas evalúan la verosimilitud y la verdad de sus creencias. El razonamiento humano va más allá de la aplicación mecánica de reglas lógicas, involucrando una evaluación continua del contenido y el contexto de las proposiciones, esencial para ajustar creencias ante nuevas informaciones y contextos.

La lógica mental ha sido muy útil para discriminar entre reglas lógicas que se aplican de forma más natural y aquellas que no. Además, ayuda a entender cómo se realizan inferencias simples y automáticas, así como el papel de la memoria de trabajo en estos procesos. Pese a todo, no abordan adecuadamente el manejo de creencias y el razonamiento bajo incertidumbre. Se centran en conclusiones binarias, omitiendo los grados de veracidad esenciales en situaciones reales y dinámicas. Además, tratan las proposiciones de forma aislada, sin considerar cómo el contexto influye en su interpretación, y por lo que a los presentes objetivos respecta, en la interpretación de las creencias. Esta orientación, pese a su complejidad interna, simplifica en exceso el razonamiento, sin capturar cómo las personas adaptan sus creencias a nuevas informaciones, un aspecto crítico en un entorno en constante cambio.

Por ello, es importante expandir estos modelos para incluir una evaluación más holística de cómo las personas procesan información. Integrar conceptos de probabilidad y lógica difusa o

similares podría ofrecer la flexibilidad necesaria para reflejar la dinámica del razonamiento humano y la adaptación de las creencias en contextos más realistas. Esta revisión conceptual podría superar las limitaciones actuales y proporcionar una comprensión más amplia y aplicable de la cognición humana.

11. Teoría de los Modelos Mentales de Johnson-Laird

Al principio, Johnson-Laird también apoyaba la perspectiva sintáctica propia de la teoría de la demostración (Hanna, 2006). Pero modificó su perspectiva sobre el razonamiento humano tras observar los estudios de Wason, que mostraban que el contenido es tan crucial como la estructura formal en los procesos de razonamiento. A diferencia de lo que ocurre con la lógica, la interpretación de las oraciones en la vida cotidiana muchas veces está modulada por el conocimiento, tanto del referente en cuestión como de los conocimientos generales (Johnson-Laird, 2008). Por otro lado, Johnson-Laird también señaló que las personas pueden hacer inferencias sobre información nueva o desconocida. Esto es, también pueden realizar operaciones por la forma del argumento. Esto sugiere la necesidad de un sistema de razonamiento que, siendo sensible al contenido, no esté estrictamente vinculado a él (Johnson-Laird, en García-Madruga, 1988, p. 317).

En 1983, Johnson-Laird introdujo la teoría de los modelos mentales en su monográfico, discutiendo temas lingüísticos, lógicos y psicológicos, incluyendo la comprensión del lenguaje, el formato de las representaciones mentales y la conciencia. Crucialmente, cuestionó la lógica mental, proponiendo una alternativa basada en modelos mentales semánticos para la deducción. Esta teoría se inspiró en la Teoría de Modelos de Tarski (1954) y se apoyó en antecedentes como la teoría de las representaciones de Wittgenstein (1922), los esquemas de Bartlett (1932), y la idea de Kenneth Craik (1943) de que la mente construye modelos reducidos de la realidad para anticipar eventos y razonar. En la teoría de los modelos mentales, estos son representaciones del mundo real o imaginario, contruidos a partir de la percepción, la comprensión del discurso o la imaginación, cuya estructura se *corresponde* con la estructura de lo que representan, conectando así el contenido con la noción semántica de verdad. Los modelos mentales son representaciones verificables; cualquier modelo es verdadero o falso acerca del mundo que pretende representar¹⁶.

Así, se ha pasado de un sistema sintáctico basado en reglas a uno semántico basado en la noción de verdad. Según Johnson-Laird (2008), el antecedente más relevante de su teoría de los modelos mentales se encuentra en el uso de la lógica de modelos, dado que proveen de un método sistemático para la interpretación de las aserciones del cálculo lógico.

¹⁶ En este sentido, yo diría que la teoría de los Modelos Mentales combina la teoría lógica de modelos con la teoría correspondentista de la verdad, teorizada como la congruencia o el isomorfismo estructural de Russell.

12. Nociones generales de la teoría de los Modelos Mentales

La propuesta de los modelos mentales es una teoría computacional desarrollada dentro del marco del procesamiento de la información, que pretende explicar qué es lo que computa la mente, cómo realiza esos cálculos, y cómo se implementa en un sustrato físico. Desde esta perspectiva, el razonamiento opera en tres estadios generales: la comprensión de las premisas, el establecimiento de las conclusiones y la validación de la inferencia.

“La teoría de los modelos mentales se ha pensado para explicar los procesos superiores de la cognición y, en particular, la comprensión y la inferencia. Sugiere un inventario simple de tres partes para el contenido de la mente: hay procedimientos recursivos, representaciones proposicionales, y modelos” (Johnson-Laird, 1983, p. 446).

12.1. Comprensión de las premisas en el razonamiento

La teoría de los modelos mentales sugiere que el razonamiento humano se orienta más hacia la semántica que hacia las reglas de inferencia estrictas, priorizando el contenido sobre la estructura formal. Según esta teoría, las personas forman en su memoria de trabajo modelos mentales que simulan como sería el mundo si las premisas fueran ciertas. Estos modelos, más que proposiciones simples, son representaciones análogas e isomórficas que capturan la esencia estructural de las situaciones sin detalles excedentarios como tamaño o color. De este modo, un modelo general es compatible con una gran cantidad de interpretaciones posibles. Estos modelos, por un lado, facilitan una comprensión mutua entre individuos y, por el otro, permiten diferencias de interpretación. Además, son dinámicos, permitiendo actualizaciones con nueva información. La comprensión de premisas deriva en un modelo mental integrado de éstas, facilitado por el solapamiento de términos con significado idéntico en distintas premisas, similar a la superposición argumental descrita por Kintsch y Van Dijk (1978). En todo caso, si la disposición de los términos en las premisas no permite una integración fácil de la información, se llevarán a cabo operaciones adicionales para intentarlo.

12.2. Establecimiento de las conclusiones en el razonamiento

En la teoría de los modelos mentales, el razonamiento es visto como un proceso de exploración de posibilidades, utilizando representaciones mentales que simulan la realidad. Estas representaciones permiten al sistema cognitivo explorar diferentes escenarios posibles y derivar implicaciones lógicas a partir de las premisas. Cada modelo mental simula un escenario posible compatible con la verdad de las premisas. Las personas interpretan las premisas y, relacionando los términos no explícitamente vinculados en los modelos, derivan conclusiones que reflejen el contenido integrado de las premisas sin información superflua. Así se logra un modelo parsimonioso, novedoso e informativo. El isomorfismo e iconicidad de los modelos mejora la identificación de relaciones y estados no explicitados en las premisas. Por ejemplo, integrando modelos de las premisas 'la taza está a la derecha de la salsa' y 'la cuchara está a la izquierda de la salsa', se facilita deducir la disposición espacial de estos objetos, algo más complejo con la lógica proposicional formal.

Además, los modelos mentales pueden representar relaciones dinámicas, tridimensionales y abstractas, añadiendo versatilidad, subyacen a las imágenes mentales, pese a que algunos componentes no son visualizables.

12.3. Proceso de validación y búsqueda de contraejemplos

En la teoría de los modelos mentales, el proceso de validación constituye el paso final en el razonamiento. Durante esta etapa, se evalúa la validez de las conclusiones extraídas al examinar diversas posibilidades alternativas. Este proceso implica buscar contraejemplos donde las premisas permanezcan verdaderas, pero la conclusión sea falsa. Si todos los modelos generados sostienen la conclusión bajo las condiciones establecidas por las premisas, entonces se considera válida. Por el contrario, si existe al menos un modelo donde las premisas son verdaderas y la conclusión falsa, la inferencia se considera inválida.

Este método de validación refleja una perspectiva más dinámica y flexible del razonamiento que abarca el deductivo, probabilístico y modal. Johnson-Laird (2002) destaca que, dependiendo de la consistencia de la conclusión a través de los diferentes modelos, las personas pueden considerarla necesaria, verosímil o simplemente posible. Así, este enfoque permite a los individuos manejar inferencias en contextos cotidianos, donde a menudo aceptan conclusiones que no cumplen con los criterios de validez lógica.

Inicialmente, se asumía que el proceso de validación en el razonamiento era automático, pero investigaciones posteriores sugieren que las personas no buscan modelos alternativos de forma espontánea. Esta idea se apoya en estudios de Evans, Handley, Harper y Johnson-Laird (1999), que indican que los individuos pueden realizar validaciones, pero no lo hacen automáticamente. Las investigaciones de Schroyens, Schaeken y Handley (2003) revelaron que factores como la presión temporal influyen significativamente en si se inicia o no la validación de inferencias *con materiales abstractos*. Descubrieron que, bajo restricciones de tiempo, las personas tienden a omitir la validación, lo que resulta en la aceptación de más inferencias que no son lógicamente válidas. Además, instruir a los participantes para que se centren en realizar únicamente inferencias lógicas necesarias aumenta la probabilidad de que realicen validaciones. Estos hallazgos sugieren que la validación es un proceso que puede ser omitido bajo ciertas condiciones, contrariamente a la creencia de que siempre se lleva a cabo de manera implícita en el razonamiento humano.

12.4. La participación de la memoria en la teoría de los modelos mentales

12.4.1. El rol de la memoria de trabajo

La teoría de los modelos mentales enfatiza el rol crítico de la memoria de trabajo en el razonamiento, destacando cómo su capacidad limitada afecta la comprensión de premisas, la construcción de modelos y la validación de información. Durante la construcción de modelos, estructuras silogísticas complejas demandan más operaciones cognitivas, lo que puede sobrecargar la memoria e impedir la integración adecuada del modelo o la correcta ejecución de inferencias. Además, la capacidad limitada de la memoria de trabajo restringe la habilidad para manejar

múltiples modelos alternativos, lo que es necesario tanto para validar inferencias como para buscar contraejemplos, incrementando así la probabilidad de aceptar conclusiones incorrectas. La influencia de la memoria de trabajo también se manifiesta en la gestión de diferentes tipos de proposiciones lógicas. Por ejemplo, una conjunción requiere menos capacidad cognitiva que una disyunción, y una disyunción excluyente menos que una incluyente, debido a la cantidad de modelos que cada una necesita mantener activos.

Para adaptarse a estas limitaciones, los modelos mentales aplican el 'principio de verdad', representando explícitamente solo aquellas cláusulas que son verdaderas según las premisas, mientras se dejan marcadores para la expansión de modelos implícitos cuando sea necesario. Por ejemplo, en un condicional como 'si hay un triángulo, entonces hay un círculo', se representa explícitamente el modelo más probable e implícitamente otros compatibles para cubrir todas las posibilidades, omitiendo configuraciones que resultarían en contradicciones con las premisas.

Los modelos mentales obedecen a las restricciones cognitivas de la memoria de trabajo y de conocimientos. Reconocen que las representaciones mentales son siempre parciales e incompletas y que las desviaciones del razonamiento lógico estándar a menudo provienen de limitaciones algorítmicas, no de errores en el estándar normativo de la lógica misma.

12.4.2. El rol de la memoria a largo plazo

La teoría de los modelos mentales destaca cómo el conocimiento almacenado en la memoria a largo plazo influye significativamente en el proceso de razonamiento. Diferenciándose de las teorías que enfocan el razonamiento en estructuras sintácticas y reglas formales, esta teoría subraya la importancia del contenido semántico y pragmático en la comprensión y construcción de modelos mentales. El conocimiento previo sobre objetos cotidianos y situaciones específicas permite a las personas inferir y razonar de manera más efectiva y precisa. Por ejemplo, ante el par de premisas 'si María entra en el ascensor, entonces Isabel saldrá un piso más arriba' y 'María entra en el segundo piso', uno puede inferir que 'Isabel saldrá en el tercer piso' (de Johnson-Laird, 2008). Pero esto no se puede explicar por la mera forma del argumento. Entender este escenario requiere, no solo seguir la lógica del argumento, sino activar un *esquema* de ASCENSOR, comprendiendo los elementos y secuencias involucradas en el discurso. Este esquema ayuda a prever y ordenar temporal y espacialmente los eventos descritos, como el movimiento de Isabel en relación a las acciones de María (Johnson-Laird, 2008). Comprender una premisa y deducir conclusiones lógicas es más intuitivo y directo cuando se basa en conocimientos concretos y contextualizados, en contraste con razonamientos que utilizan elementos abstractos como variables alfabéticas, donde el contenido juega un papel menos crítico. Esta teoría resalta cómo el contenido semántico dirige los procesos de razonamiento más allá de las limitaciones impuestas por la estructura formal de los argumentos.

Los conocimientos previos influyen en la validación de inferencias al facilitar la búsqueda de contraejemplos en la memoria a largo plazo. Schroyens et al. (2003) ilustran esto mediante el siguiente ejemplo: al girar la llave se espera que el coche arranque y se gira la llave, concluyendo

que el coche arranca. Pero aquí podría producirse una búsqueda de contraejemplos recurriendo a conocimientos previos como que no tenga gasolina. No está garantizado que el sujeto inicie la búsqueda de contraejemplos, ni que los encuentre rápido, ni que considere válido el contraejemplo encontrado. Si se halla un contraejemplo válido, puede llevar a rechazar la inferencia inicial o a mantenerla *con menor confianza* (Schroyens et al., 2003), siendo más frecuente el rechazo cuanto más comunes sean los contraejemplos.

Nótese que aquí, el proceso de validación no es del mismo tipo que el que se mencionó en la sección correspondiente a la validación. Aquella, validación tenía que ver con una búsqueda combinatoria en algo análogo a las tablas de verdad de la teoría lógica de modelos, aplicable a materiales abstractos o a historias ficticias. No consistía en un contraste de posibilidades en relación con la información almacenada en la memoria a largo plazo: el conocimiento previo no ayuda a encontrar casos en los que, dándose un triángulo, no se dé un círculo; ni donde, habiendo entrado María en el ascensor, no haya salido Isabel en el piso siguiente. Sin embargo, en el ejemplo del coche, el contraejemplo consiste en buscar una *razón* que *justifique* pue podría darse p y $\neg q$. Que no tenga gasolina es una razón que lo justifica y es, a su vez, un elemento externo a cualquier tabla de verdad. En todo caso, tanto Byrne, Espino y Santamaría (1999) como Quinn y Markovits (1998) tratarán de dar a estos casos una explicación en términos de la ‘tabla de verdad mental’ implicada en los modelos.

12.5. Valoración de la teoría de los modelos mentales desde la presente perspectiva

La teoría de los modelos mentales, desarrollada inicialmente por Johnson-Laird, ha transformado significativamente nuestra comprensión de los procesos cognitivos avanzados como la comprensión y la inferencia. Al alejarse de las orientaciones más tradicionales que se centraban en las reglas sintácticas y formales, esta teoría ha enfatizado la importancia del contenido semántico y pragmático. La teoría de los modelos mentales no solo facilita un razonamiento más adaptado a situaciones reales, sino que también permite simulaciones mentales más intuitivas de posibles realidades. Una característica distintiva de la teoría es su adaptabilidad y la capacidad de actualizar los modelos mentales con nueva información, lo que la convierte en una herramienta invaluable en contextos dinámicos donde las premisas pueden cambiar o expandirse. Esta flexibilidad es crucial para manejar la complejidad y variabilidad del razonamiento humano en situaciones cotidianas. La teoría propone métodos de validación que incluyen la búsqueda de contraejemplos, donde las premisas puedan ser verdaderas pero la conclusión falsa. Esto ha sido fundamental para avanzar hacia un modelo de razonamiento que reconoce y gestiona la incertidumbre inherente a las situaciones que los humanos experimentan.

La integración de la memoria semántica y su impacto en la validación de inferencias a través de la búsqueda activa de contraejemplos destaca cómo el conocimiento del mundo puede influir en la evaluación de las inferencias, permitiendo análisis más detallados y ajustados a la realidad. La idea de que, al encontrar un contraejemplo, la inferencia puede mantenerse, pero *con menos confianza*, introduce la noción de que hay grados de confianza en las conclusiones. Esta idea enlaza

naturalmente con propuestas de carácter más probabilístico y con el estudio de las creencias, sugiriendo una relación más matizada entre conocimiento, creencias y razonamiento. En este sentido, Schroyens et al. (2003, p. 148) creen que los mecanismos analíticos y de interconexión que postulan las teorías sobre el razonamiento humano deben elaborarse sin perder de vista los mecanismos de almacenamiento y recuperación de la información de la memoria a largo plazo (i.e. “los aspectos probabilísticos del conocimiento general asociado con el contenido proposicional acerca del cual uno está razonando”).

La memoria a largo plazo juega un papel crucial en moderar la confianza en las conclusiones, un aspecto que también caracteriza a las perspectivas probabilísticas del razonamiento. En línea con esto, Oaksford y Chater (2007) sugieren que, en su modelo probabilístico, las probabilidades subjetivas se basan en el conocimiento almacenado. Este principio se aplica igualmente en el test de Ramsey, que utiliza conocimientos previos para evaluar afirmaciones. En este sentido, recientemente existe cierta tendencia por adaptar algunos aspectos de la teoría de los modelos mentales con estas propuestas más probabilísticas (Johnson-Laird, Khemlani, y Goodwin, 2015; Khemlani, Byrne, y Johnson-Laird, 2018). Esto será retomado más adelante, en el desarrollo del modelo teórico que se propone en esta tesis.

En conjunto, la teoría de los modelos mentales ha contribuido de manera importante a entender cómo el razonamiento humano es sensible a los contenidos de las representaciones. Esta condición ha sido, a su vez, esencial para entender cómo los procesos de razonamiento pueden adaptarse a las realidades cambiantes propias del mundo cotidiano. De este modo, ha abierto nuevas vías para la investigación y la aplicación práctica en la psicología cognitiva, proporcionando una base que ha posibilitado el estudio del razonamiento cotidiano. Y, consecuentemente, también de las creencias.

13. Estudios psicológicos del razonamiento cotidiano y su naturaleza rebatible

Uno de los objetivos de la argumentación retórica es cambiar la credibilidad de una aserción, lo cual se puede lograr mediante el razonamiento sobre asuntos cotidianos. Las inferencias cotidianas son rebatibles, derrotables y no-monótonas, y no son demostrativas debido a los conocimientos limitados de los individuos sobre el contenido de los argumentos.

El cuerpo de estudios que se expondrá a continuación constituye un acercamiento de la psicología al razonamiento rebatible cotidiano. El razonamiento rebatible se explora en profundidad a través de estudios que utilizan aserciones condicionales, especialmente condicionales causales, para examinar cómo diferentes tipos de información impactan la aceptación de conclusiones derivadas de inferencias, sean estas válidas o no. La búsqueda de contraejemplos en el proceso de validación es una cuestión clave en estos estudios, reflejando el esfuerzo por entender cómo el contexto y el contenido específico afectan el razonamiento. Antes que nada, veamos en qué consiste el razonamiento rebatible.

13.1. El razonamiento rebatible

El razonamiento rebatible ('defeasible') ya fue identificado y analizado por Aristóteles en sus obras de *Tópicos*, *Retórica* y *Analíticos Primeros* bajo las etiquetas de razonamiento dialéctico, retórico y probable. Crucialmente, los entimemas son razonamientos rebatibles. Por tanto, ya desde su origen, el razonamiento rebatible tiene una relación directa con estas áreas de conocimiento.

El razonamiento rebatible discurre sobre generalizaciones del tipo 'los As son Bs' o 'si A, entonces B'. En circunstancias normales, estas generalizaciones se cumplen la mayoría de las veces. Es un razonamiento cuyas inferencias conllevan "generalizaciones que permiten excepciones, donde se infiere lo que ha sucedido o lo que sucederá sobre la base de lo que *normalmente* sucede" (Koons, 2017). Una aserción es rebatible si, implícita o explícitamente, toma la forma de "S es normalmente P" (Vega, 2004). Y un razonamiento es rebatible si se apoya en una aserción rebatible.

El modelo de Toulmin (1958) describe los razonamientos cotidianos como procesos basados en generalizaciones falibles que pueden ser suspendidas por excepciones. En este modelo, los argumentos se estructuran alrededor de tres componentes esenciales: datos, garantes y conclusiones (ver sec. 4.2., cap. 1, pt. I). Los garantes funcionan como reglas condicionales que vinculan los datos con la conclusión, pero pueden ser contrarrestados por *refutadores* que obligan a reconsiderar la validez de la conclusión. Estos refutadores representan "condiciones excepcionales que pueden hacer descartar o rechazar la conclusión justificada" (Toulmin, 1958, p. 137), modificando la inferencia condicional básica de ' $P \rightarrow Q$ ' a ' $(P \wedge R) \rightarrow \neg Q$ ' al añadirse el refutador R.

13.2. Tipos de rebatidores en la propuesta de Pollock

John Pollock (1987) ha proporcionado una visión profunda del razonamiento rebatible, destacando que las inferencias pueden ser inicialmente válidas a partir de un conjunto de premisas, pero pueden dejar de serlo con la adición de nueva información que no contradice necesariamente las premisas originales. Este fenómeno es central en su concepto de *razones prima facie*, definidas como situaciones donde "P es una razón prima facie para que S crea Q si y solo si P es una razón para Q, y existe una R tal que sea lógicamente consistente con P pero (P&R) no es una razón para que S crea Q" (Pollock, 1987, p. 484). Estas razones ilustran la naturaleza contingente de las inferencias, susceptibles de ser invalidadas por información adicional conocida como rebatidores.

Pollock distingue dos tipos principales de rebatidores: *refutativos* (*rebutting defeaters*) y *socavadores* (*undercutting defeaters*). Los rebatidores refutativos son aquellos que directamente contradicen la conclusión esperada, como cuando "R es un rebatidor refutativo para P como razón prima facie de Q si y solo si R es un rebatidor y R es una razón para creer que $\neg Q$ " (Pollock, 1987, p. 485). Un ejemplo clásico sería un coche que no se enciende a pesar de girar la llave, debido a una batería desconectada, invalidando la conclusión de que el coche debería arrancar. R muestra que girar la llave no es suficiente para que arranque. Por otro lado, los rebatidores socavadores cuestionan la conexión entre la premisa y la conclusión sin refutar directamente la conclusión. Pollock los define como relevantes cuando "R es un rebatidor socavador para P como razón prima facie para que S crea que Q si y solo si R es un rebatidor y R es una razón para negar que P no sería

verdad a menos que Q fuera verdad" (Pollock, 1987, p. 485). Un ejemplo sería la percepción de un objeto como rojo: "no me lo parecería a menos que lo fuera"; sin embargo, hay causas alternativas para que me lo parezca, como que esté iluminado por una luz roja. Esto es una razón para negar que "no me lo parecería a menos que lo fuera".

Estos rebatidores ilustran cómo el razonamiento rebatible maneja la incertidumbre y las condiciones contextuales en la vida cotidiana, permitiendo ajustes en el razonamiento basados en nueva información. La capacidad de integrar y responder a contraejemplos específicos es crucial para la adaptabilidad y precisión del pensamiento humano en situaciones reales, destacando la importancia de considerar tanto los rebatidores refutativos, que afectan directamente la validez de las conclusiones, como los socavadores, que influyen en la conexión entre premisas y conclusiones. Esta dualidad subraya la complejidad del razonamiento en contextos prácticos y cotidianos, donde las inferencias están sujetas a revisión.

14. Psicología del efecto de la búsqueda de rebatidores en la supresión de la inferencia

El fenómeno conocido como supresión de la inferencia describe cómo la aceptación de una conclusión inferida desde ciertas premisas puede verse alterada al introducir nueva información rebatidora. Este efecto ha sido un foco de estudio en psicología, especialmente en relación con cómo la adición de información o el acceso a conocimientos preexistentes en la memoria a largo plazo pueden cambiar la percepción de la validez de las conclusiones derivadas de argumentos condicionales. Se han examinado especialmente cuatro formas estructurales de argumento en este contexto, analizando cómo distintos tipos de información afectan la aceptación de las conclusiones, tanto en inferencias válidas como en no válidas. En la Tabla 1, se presenta un ejemplo de cada una de las cuatro formas junto con información acerca de si es una inferencia válida o no lo es.

Desde la perspectiva logicista se consideraba que el razonamiento era un proceso puramente sintáctico. En este sentido, los adolescentes y los adultos, en tanto que habrían alcanzado el estadio de las operaciones formales (Inhelder y Piaget, 1958) debían razonar correctamente de forma relativamente homogénea aplicando reglas mentales de inferencia a los argumentos sin atender a su contenido (Braine, 1978; Henle, 1962; Rips, 1983). Razonar correctamente significa extraer o aceptar sólo las conclusiones válidas de los argumentos. En el caso de condicionales, esto implica entender las relaciones de necesidad y suficiencia entre antecedente y consecuente.

Específicamente, *en un condicional material*, el antecedente P es suficiente pero no necesario para Q, significando que Q puede ocurrir sin P, pero si P ocurre, Q también lo hará sin excepciones. Contrariamente, Q es necesario para P, indicando que P no puede existir sin Q, aunque la presencia de Q no garantiza la de P. Esto clarifica por qué en lógica formal los argumentos de Modus Ponens (MP) y Modus Tollens (MT) son considerados válidos, mientras que la Negación del Antecedente (NA) y la Afirmación del Consecuente (AC) son falacias formales, a menudo malinterpretadas como bicondicionales donde $P \leftrightarrow Q$, es decir, donde la necesidad y suficiencia son bidireccionales.

Tabla 1

Ejemplos de las Cuatro Reglas de Inferencia Condicional Válidas e Inválidas

Modus Ponens (MP): válido	Modus Tollens (MT): válido
$\frac{P \rightarrow Q}{P}$ $\frac{P}{Q}$	$\frac{P \rightarrow Q}{\neg Q}$ $\frac{\neg Q}{\neg P}$
Si se presiona el freno, el coche va más lento. Se presiona el freno. El coche va más lento.	Si se presiona el freno, el coche va más lento. El coche no va más lento. No se ha presionado el freno.
Negación del Antecedente (NA): inválido	Afirmación del Consecuente (AC): inválido
$\frac{P \rightarrow Q}{\neg P}$ $\frac{\neg P}{\neg Q}$	$\frac{P \rightarrow Q}{Q}$ $\frac{Q}{P}$
Si se presiona el freno, el coche va más lento. No se presiona el freno. El coche no va más lento.	Si se presiona el freno, el coche va más lento. El coche va más lento. Se ha presionado el freno.

14.1. Desafíos que afronta el razonamiento cotidiano: estudios sobre los efectos de supresión de la inferencia

Las condiciones ideales de razonamiento lógico, adecuadas para verdades de razón o analíticas, a menudo no se aplican directamente a situaciones empíricas, que están plagadas de excepciones y conocimiento imperfecto. Por ejemplo, la proposición 'si todos lo hicieron, entonces algunos también lo hicieron' es una verdad analítica que sigue fielmente la lógica del condicional material. Sin embargo, realidades empíricas como 'si se presiona el freno, el coche va más lento' pueden no seguir este patrón debido a variables imprevistas, como hielo en la carretera o problemas mecánicos. Este tipo de contingencias demuestra que las generalizaciones en el razonamiento empírico son susceptibles a excepciones, mostrando así la influencia del contexto y el contenido específico en el razonamiento cotidiano.

La investigación en psicología ha explorado extensamente estos fenómenos a través de estudios sobre la supresión de la inferencia y la búsqueda de contraejemplos, donde se examina cómo la información adicional puede inhibir inferencias que serían válidas en un marco lógico estricto. Trabajos significativos en este campo incluyen los de Byrne (1989) y Cummins, Lubart, Alksnis y Rist (1991), que han sido fundamentales para entender cómo los individuos manejan información contradictoria o adicional en el proceso de razonamiento. Estos estudios han revelado que tanto el número como la fuerza de asociación de rebatidores en la memoria semántica juegan roles cruciales en cómo las personas recuperan y procesan dicha información. La metodología desarrollada por Cummins et al. (1991) ilustra cómo la memoria de trabajo y la accesibilidad, ya sea automática o estratégica, a información rebatidora afecta la aceptación de conclusiones.

14.2. El paradigma de presentación explícita de rebatidores

El hecho de que las personas cometieran falacias formales en el laboratorio es difícil de explicar si uno se basa en la lógica mental derivada de la teoría de la demostración. Para hacerlo, recurrieron a explicaciones pragmáticas y de interpretación del discurso. Ante frases como “si recoges la habitación, entonces podrás salir a jugar”, Geis y Zwicky (1971) propusieron que se producía una inferencia invitada o implicatura de que “si no recoges la habitación, no podrás salir a jugar”. De lo contrario, el condicional no sería relevante. Las personas tienden a interpretar las afirmaciones condicionales como bicondicionales, influenciadas por la “lógica de la conversación” (Grice, 1975) y el conocimiento del mundo, más que por reglas formales de inferencia. Pero estas interpretaciones bicondicionales que llevan a aceptar falacias formales pueden suprimirse si el oyente puede acceder a *antecedentes alternativos* de la premisa condicional.

Los trabajos pioneros de Markovits (1984) mostraron experimentalmente que al introducir premisas con antecedentes alternativos se podían suprimir las conclusiones erróneas derivadas de NA y AC. Por ejemplo, si a la afirmación de que una persona se quedará hasta tarde en la biblioteca si tiene un ensayo por escribir, se añade que también podría quedarse si tiene libros por leer, la conclusión de que no se quedará porque no tiene ensayos por escribir se suprime. Esto llevó a estos autores a concluir que en la lógica mental no había reglas de inferencias inválidas, puesto que se podían suprimir. Esto no ocurriría con MP y MT. Esto implica que, si se pudieran suprimir, entonces tampoco habría inferencias válidas MP y MT. Con una metodología análoga, Byrne (1989) añadía a los argumentos otros antecedentes que mostraran que el original no era suficiente. Por ejemplo, que ‘si la biblioteca permanece abierta, se quedará hasta tarde en la biblioteca’. De este modo, pretendía demostrar que incluso las inferencias válidas, MP y MT, podían ser suprimidas mediante la introducción de premisas que sugieren requisitos adicionales no contemplados inicialmente.

En su experimento, se pidió a los participantes que asumieran la veracidad de las premisas. Los resultados revelaron que las inferencias válidas se veían reducidas cuando los argumentos incluían antecedentes adicionales (*r*) que mostraban que el original (*p*) no era suficiente para el consecuente (*q*). Además, observó que en condiciones donde se añadía otra premisa con la información fáctica de que ‘tiene un ensayo que escribir y la biblioteca permanece abierta’ (*p* y *r*) la supresión de la inferencia desaparecía. Estos hallazgos apoyan la teoría de los modelos mentales, según la cual las personas buscan contraejemplos o modelos alternativos que puedan refutar las conclusiones iniciales. Byrne, Espino y Santamaría (1999) argumentan que el proceso de supresión demuestra la capacidad humana para adaptar el razonamiento a nueva información que desafía las inferencias previas. Proponen que el tipo de modelo mental generado—incluyendo una disyunción para antecedentes alternativos para suprimir NA y AC, o una conjunción para antecedentes adicionales para suprimir MP y MT—influye en la disponibilidad y efectividad de los contraejemplos necesarios para producir dicha supresión. Esto se debe a las posibilidades lógicas en las que los antecedentes son verdaderos, algo análogo a una tabla de verdad mental.

Todas estas inferencias pueden ser suprimidas proporcionando información adicional de posibles rebatidores. Byrne et al. (1999) consideran que estos estudios constituyen los primeros avances sólidos hacia la investigación psicológica del razonamiento rebatible o no-monótono que se da en las inferencias cotidianas de las personas.

14.3. El paradigma de presentación implícita de rebatidores

El estudio del razonamiento rebatible avanzó significativamente con investigaciones que exploraban cómo se pueden suprimir inferencias tanto válidas como inválidas al introducir información adicional de manera explícita. Esta metodología permitió a los participantes considerar contenidos adicionales al razonar, evidenciando que el acceso a esta información influye directamente en el proceso de razonamiento. Investigaciones posteriores dirigidas por Cummins et al. (1991) y Cummins (1995) se centraron en entender si los individuos rutinariamente acceden a rebatidores almacenados en la memoria a largo plazo sin necesidad de una presentación explícita. Utilizando condicionales causales en un paradigma causal implícito, estos estudios redefinieron los rebatidores del paradigma explícito como *causas alternativas* (alternative causes) y *condiciones desactivadoras* (disabler conditions). Los primeros consisten en modos diferentes de producir el consecuente (p.ej., “que tenga muchos libros que leer” en el ejemplo de quedarse hasta tarde en la biblioteca), y los segundos consisten en modos de impedir la ocurrencia del consecuente en presencia de una causa viable (p.ej., “que la biblioteca esté cerrada”, en el mismo ejemplo). En la preparación, pidieron a los participantes generar posibles causas alternativas y condiciones desactivadoras para diferentes escenarios causales, con el objetivo de clasificar los condicionales según la cantidad y tipo de rebatidores que podrían evocar.

En los experimentos, se presentaron a los participantes argumentos estructurados en las cuatro formas inferenciales (Modus Ponens, Modus Tollens, Negación del Antecedente, Afirmación del Consecuente) con variaciones en el número y la disponibilidad de causas alternativas y condiciones desactivadoras. Los participantes evaluaron la aceptabilidad de las conclusiones en una *escala graduada* de seis puntos, sin instrucciones de adherirse a estándares lógicos o asumir la veracidad de las premisas. Los resultados mostraron que las condiciones desactivadoras, cuando eran abundantes, disminuían la certeza en las conclusiones de inferencias válidas, indicando que la causa no era suficiente para garantizar el efecto. En cambio, un alto número de causas alternativas reducía la aceptación de las inferencias inválidas, sugiriendo que la causa no era necesaria para el efecto. Estos hallazgos, interpretados en términos de necesidad y suficiencia, resaltan que tanto las condiciones desactivadoras como las causas alternativas actúan como rebatidores *refutativos* y *socavadores* de Pollock respectivamente (Politzer y Bonnefond, 2006), afectando la percepción de la suficiencia y necesidad de las causas en los condicionales causales.

En conclusión, estos estudios ilustran cómo la supresión de inferencias puede ser influida por rebatidores no presentados explícitamente, accesibles mediante la búsqueda rutinaria en la memoria semántica. Este avance en la comprensión del razonamiento rebatible subraya la importancia de considerar cómo la memoria y el conocimiento previo del individuo moldean la

lógica aplicada en situaciones cotidianas, llevando a una evaluación más dinámica y contextualizada de las inferencias.

14.4. Valoración de estos paradigmas en relación con la presente investigación

Los estudios sobre el paradigma de presentación implícita y explícita de rebatidores han hecho contribuciones fundamentales al entendimiento del razonamiento y la formación de creencias desde una perspectiva psicológica y empírica, enfocándose particularmente en cómo se modifica la credibilidad de las conclusiones basada en el acceso a rebatidores. Estas investigaciones han demostrado empíricamente el impacto significativo que los rebatidores tienen en la aceptación de conclusiones, subrayando cómo la presencia o el acceso a información contradictoria puede alterar la aceptación de una afirmación, lo cual está estrechamente relacionado con la credibilidad. Este efecto es crucial para entender cómo las personas ajustan sus creencias en respuesta a nueva información.

Estos estudios han revelado que los individuos acceden rutinariamente a rebatidores almacenados en la memoria a largo plazo cuando no se presentan explícitamente, influyendo en la aceptación de las conclusiones en condiciones similares a situaciones cotidianas reales. La cantidad de rebatidores asociados a un condicional en la memoria y la forma del argumento influyen en el tipo de rebatidores accedidos—ya sean causas alternativas o condiciones desactivadoras—y, por tanto, en cómo se evalúa la aceptabilidad de una conclusión. Crucialmente, las condiciones experimentales de estos experimentos se aproximan al razonamiento cotidiano: las instrucciones eran mínimamente restrictivas, permitiendo respuestas graduadas sin la obligación de adherirse a un razonamiento lógico estricto y sin asumir que las premisas eran verdaderas.

Este marco experimental permite un abordaje más natural que progresivamente se aleja del logicismo formalista para aproximarse al razonamiento rebatible inherente a la argumentación. Al demostrar que la gente maneja activamente la información contradictoria en sus procesos de razonamiento, estos estudios abrieron nuevas avenidas para explorar la credibilidad y la argumentación en contextos empíricos, ofreciendo insights valiosos sobre cómo las creencias y las conclusiones son evaluadas y ajustadas en la práctica diaria.

15. Estudios Subsiguientes sobre los Procesos del Razonamiento Rebatible

Estos estudios iniciales impulsaron numerosas investigaciones sobre el impacto del contenido en el razonamiento cotidiano. Los trabajos posteriores se centraron en definir el alcance de los rebatidores, explorar las condiciones que afectan su eficacia en la supresión de inferencias y analizar cómo funcionan los procesos de búsqueda de contraejemplos. A continuación, se presenta un breve compendio de los principales trabajos y sus resultados, organizados por temas.

15.1. Efectos Pragmáticos del Razonamiento Condicional

Los estudios de Cummins et al. (1991, 1995) demostraron que en el razonamiento con condicionales causales, el acceso a condiciones desactivadoras y causas alternativas disminuye la aceptación de inferencias, ya sean válidas o inválidas. La causa de esto radica en la percepción de necesidad y suficiencia entre causa y efecto. Pero no tiene por qué restringirse solo a las relaciones causales. Thompson (1994) extendió esta observación al razonamiento deóntico, analizando cómo diferentes contextos—obligaciones, permisos, causales y definicionales—afectan a la interpretación de la necesidad y suficiencia de los condicionales. Además, pidieron a los participantes razonar lógicamente asumiendo la verdad de las premisas. Los resultados indicaron que la percepción de necesidad y suficiencia se ve influida por el contexto deóntico y, en la línea de los trabajos de Cummins, que los conocimientos previos intervienen por defecto en el razonamiento, pese a tratar de inhibir su efecto. Fairley, Manktelow y Over (1999) y Dieussaert, Schaeken y d'Ydewalle (2002) continuaron explorando estas ideas, centrándose en cómo la perspectiva de los interlocutores (i.e., hablante vs. oyente) y el control del hablante sobre la ocurrencia del consecuente de los condicionales influyen en su interpretación. Observaron que estas variables pragmáticas son especialmente influyentes en la interpretación de este tipo de condicionales.

Investigaciones más recientes realizadas por Evans, Neilens, Handley y Over (2008) y Thompson, Evans, Simon y Handley (2005) han profundizado en cómo las relaciones de necesidad y suficiencia, junto con el control del hablante sobre los resultados, modulan la fuerza persuasiva de las emisiones comunicativas en razonamientos informales. Estos estudios destacan que, en el razonamiento condicional cotidiano, los individuos no solo procesan premisas lógicas, sino que computan creencias sobre posibles estados del mundo, considerando la probabilidad condicionada y los valores de utilidad asociados con diferentes acciones y consecuencias. Evans (2005) conceptualiza los condicionales como actos de habla que operan en contextos sociales, afectando las creencias y las acciones de los oyentes a través de modelos mentales hipotéticos y probabilísticos. Estos modelos reflejan la probabilidad condicionada de una consecuencia dado un antecedente, integrando así las metas y valores del hablante. Este enfoque resalta la función comunicativa y social de los condicionales, evidenciando cómo el razonamiento está intrínsecamente vinculado con el impacto pragmático de las emisiones comunicativas en las creencias y opiniones de los interlocutores.

15.2. El Rol de la Memoria Semántica en la Recuperación de Rebatidores

Los estudios que se revisan a continuación se han centrado en diversos aspectos del proceso de recuperación de rebatidores y los efectos que esto tiene sobre la aceptación o rechazo de las conclusiones de condicionales causales. Algunos trabajos han examinado la influencia que tiene el número de rebatidores recuperados de la memoria semántica mientras que otros han explorado los efectos de la fuerza de asociación de éstos con la situación causal sobre la evaluación de la conclusión. Otras investigaciones se han centrado más en tratar de comprender y describir los

procesos on-line de recuperación de la memoria semántica, así como del rol que juega la memoria de trabajo en este proceso y si se trata de un proceso automático o estratégico.

15.3. Evaluación de conclusiones

Efectos del Número de Rebatidores: Los estudios de Cummins et al. (1991) y Cummins (1995) destacaron la relevancia del número de rebatidores almacenados en la memoria semántica en la evaluación de conclusiones de condicionales causales, influyendo en la aceptación de inferencias válidas e inválidas. De Neys, Schaeken y d'Ydewalle (2003a) exploraron cómo el número de rebatidores afecta esta aceptación, utilizando tanto un paradigma explícito como implícito. En ambos casos, descubrieron que un mayor número de rebatidores está directamente relacionado con una menor aceptación de conclusiones. Estos resultados muestran que el proceso de búsqueda no se detiene tras la recuperación de un único rebatidor, y que cada uno que se recupera tiene un peso sobre la aceptación de la conclusión.

Además, la investigación de Geiger y Oberauer (2007) indicó que más allá de la presencia de rebatidores específicos, la frecuencia de las excepciones en el cumplimiento de condicionales también modula la credibilidad y la aceptación de las conclusiones. Sus experimentos sugirieron que, en ausencia de información sobre la frecuencia de las excepciones, el número de distintos tipos de condiciones desactivadoras influyen significativamente en la aceptación de las conclusiones de argumentos válidos MP y MT. Sin embargo, cuando se proporciona información sobre la frecuencia de los rebatidores, la aceptación de las conclusiones se basa principalmente en esta, más que en la presencia de múltiples condiciones desactivadoras específicas (i.e., un solo desactivador asociado al antecedente con una frecuencia relativa muy alta puede suprimir totalmente la inferencia válida). Esto resalta cómo la percepción de la probabilidad subjetiva, más que los rebatidores individuales, guía la credibilidad y la aceptación de inferencias, sugiriendo una visión más general de la incertidumbre en el razonamiento condicional.

Efectos de la Fuerza de Asociación: Quinn y Markovits (1998) ampliaron la teoría de los modelos mentales para integrar el impacto de la fuerza de asociación en la aceptación de inferencias condicionales. Sostienen que el proceso de validación activa automáticamente información relevante en la memoria semántica, lo cual facilita la identificación de contraejemplos basados en cómo las causas alternativas y las condiciones desactivadoras están vinculadas a los efectos. Las causas fuertemente asociadas tienden a ser consideradas más necesarias, lo que puede llevar a una aceptación incorrecta de conclusiones inválidas en inferencias de Negación del Antecedente (NA) y Afirmación del Consecuente (AC). En un experimento, pidieron a los participantes que generasen tantas causas como se les ocurrieran para varios efectos en un tiempo limitado. Entonces, se construyeron condicionales con esos materiales. Las causas mencionadas con mayor frecuencia por los participantes se usaron para crear condicionales de fuerte asociación, mientras que las menos frecuentes se emplearon para los de asociación débil.

Posteriormente, otros participantes razonaban sobre esos condicionales, usando las cuatro reglas de inferencia (MP, MT, NA, AC), y debían evaluar la aceptación de sus conclusiones. Los

resultados mostraron que una fuerte asociación aumentaba la percepción de necesidad de la causa, aumentando la aceptación de conclusiones de las reglas inválidas NA y AC. Sin embargo, esta tendencia se revertía con una asociación débil. Los autores interpretaron que estos efectos se debían a que una mayor fuerza de asociación dificultaba el acceso a causas alternativas, mientras que una menor fuerza de asociación lo facilitaba. Al acceder a causas alternativas y disminuir la percepción de necesidad de la causa para el efecto se rechazaban las conclusiones surgidas de inferencias inválidas.

De Neys, Schaeken y d'Ydewalle (2003b) también investigaron cómo la fuerza de asociación afecta la recuperación de condiciones desactivadoras y su impacto en la evaluación de argumentos Modus Ponens (MP) y Modus Tollens (MT). Emplearon una metodología análoga a la anterior, esta vez, para generar desactivadores fuerte y débilmente asociados a condicionales ya construidos. Entonces, construían argumentos MP y MT añadiendo la negación del desactivador fuerte o débil. Por ejemplo, si “el pedal está roto” es un desactivador fuerte para “si piso el freno, el coche se va más lento”, y “hay hielo en la carretera” es débil. Entonces, se añadía negado a la premisa menor “piso el pedal del freno y el pedal no está roto”. Presentaban esos materiales a otros participantes quienes debían evaluar la aceptación de las conclusiones en una escala de 11 puntos. Encontraron que los argumentos que incorporaban la negación de desactivadores fuertemente asociados eran más aceptados que aquellos con desactivadores débilmente asociados, mostrando que la percepción de la suficiencia de una causa puede ser significativamente alterada por cómo está organizada la información sobre las condiciones desactivadoras en la memoria semántica.

15.4. Curso temporal del proceso de búsqueda

La investigación ha demostrado que el número y la fuerza de asociación de los rebatidores son determinantes clave en la facilidad con la que se recuperan estos elementos durante el razonamiento condicional causal. Cuando se razona, los individuos acceden a estructuras en la memoria semántica que contienen rebatidores y los integran en el curso del razonamiento. De acuerdo con algunos modelos de la memoria a largo plazo, como el de Anderson (1983), la probabilidad de recuperar, al menos, un elemento de la estructura de la memoria semántica dependerá del número de elementos que contenga esa estructura: cuantos más rebatidores haya en las estructuras causales para un condicional, más probable será recuperar al menos uno. Por otro lado, si las estructuras causales contienen muchos rebatidores, es probable que se recuperen varios elementos. Conway y Engle (1994) observaron que la duración del proceso de búsqueda en la memoria se extiende en función del número de elementos recuperados. En función de esto, De Neys, Schaeken y d'Ydewalle (2002) postularon que el tiempo necesario para realizar una inferencia debería alargarse cuanto mayor fuera la cantidad de rebatidores a los que se accedía. Esta hipótesis obtuvo respaldo experimental, al observarse un aumento de las latencias de respuesta en la aceptación de las conclusiones conforme aumentaba el número de rebatidores. Esto se interpretó como reflejando una búsqueda más intensiva y extensa dentro de la estructura de la memoria.

En línea con estos hallazgos, Grosset, Barrouillet y Markovits (2005) investigaron cómo la fuerza de asociación entre causas alternativas y sus consecuentes afecta la rapidez y el tipo de respuesta en inferencias consideradas lógicamente inválidas, como las Negaciones del Antecedente y Afirmaciones del Consecuente. Descubrieron que respuestas de rechazo correctas son más frecuentes y rápidas cuando el antecedente está débilmente asociado con el consecuente, comparado con asociaciones fuertes que provocan respuestas incorrectas más rápidas debido al acceso inmediato a causas fuertemente asociadas. Esta dinámica demuestra que tanto el tiempo como la calidad de las evaluaciones están intrínsecamente vinculados a la estructura de la memoria semántica y sus dinámicas asociativas. Esto sugiere que una mayor aceptación de una conclusión conlleva tiempos de respuesta más breves debido a un proceso de recuperación más corto.

15.5. El papel de la memoria de trabajo

El papel de la memoria de trabajo en la recuperación de rebatidores en el razonamiento condicional ha sido un foco de investigación significativo. La memoria de trabajo, estructurada jerárquicamente con componentes de memoria a corto plazo y un ejecutivo central, gestiona los recursos atencionales necesarios para procesar información (Baddeley, 1996; Cowan, 1995; Engle et al., 1999). Rosen y Engle (1997) destacaron que la recuperación de información consume recursos del ejecutivo central, y que tareas concurrentes pueden reducir la eficiencia de esta recuperación. Aunque el acceso a estructuras de memoria semántica durante el razonamiento se ha considerado un proceso mayormente automático y de bajo consumo de recursos (Evans, 2002; Kintsch, 1998), la inhibición de este acceso sí requiere recursos significativos (Evans y Over, 1996; Stanovich y West, 1998). Rosen y Engle (1997) sugieren que, aunque la activación inicial de la memoria es automática, las búsquedas estratégicas posteriores utilizan recursos del ejecutivo central para optimizar la recuperación.

De Neys, Schaeken y d'Ydewalle (2005ab) investigaron si la recuperación de rebatidores es un proceso que depende de la capacidad de la memoria de trabajo. En sus estudios, los participantes con mayor capacidad de memoria de trabajo generaron más rebatidores bajo condiciones controladas, y una carga adicional en la memoria de trabajo redujo la eficiencia en la recuperación de estos rebatidores. Este efecto fue menos pronunciado para rebatidores fuertemente asociados, lo que sugiere que el acceso a asociaciones más fuertes podría ser más automático. El estudio también exploró cómo la capacidad de memoria de trabajo afecta la evaluación de argumentos condicionales. Los participantes con alta capacidad tendieron a adherirse más a los estándares lógicos, rechazando más conclusiones derivadas de inferencias inválidas y aceptando más las válidas en comparación con aquellos de baja capacidad. Esto indica que una mayor capacidad de memoria de trabajo permite una generación más efectiva de rebatidores y una mejor regulación de su impacto en la evaluación de conclusiones.

Estos hallazgos ilustran cómo la memoria de trabajo no solo facilita la recuperación de información relevante, sino que también modula cómo se integra esa información en el

razonamiento, afectando directamente la credibilidad y aceptación de las conclusiones en el razonamiento condicional.

15.6. Activación y disponibilidad de los rebatidores

En estudios sobre la memoria de trabajo y su influencia en la recuperación de rebatidores, se ha investigado la dinámica entre la activación automática y la inhibición consciente de información relevante para el razonamiento. De Neys (2010) combinó tareas de razonamiento con tareas de decisión léxica para explorar cómo se activan y se inhiben las condiciones desactivadoras y causas alternativas durante el razonamiento. Los resultados mostraron que, tras evaluar argumentos inválidos, las respuestas a palabras relacionadas con causas alternativas eran más rápidas comparadas con palabras no relacionadas, indicando una activación automática de conocimientos pertinentes a los rebatidores. Sin embargo, después de argumentos válidos, las palabras relacionadas con condiciones desactivadoras causaban retrasos, sugiriendo que los participantes inhibían activamente esta información para no interferir con la aceptación de conclusiones válidas.

Este proceso fue analizado más detenidamente en un estudio que utilizó EEG para detectar cambios neurofisiológicos asociados con la activación de rebatidores durante el razonamiento. Pijnacker et al. (2010) y Bonnefond et al. (2014) observaron cómo el cerebro procesa y responde a la información desactivadora introducida antes de las conclusiones de los argumentos. Descubrieron que, incluso antes de que la conclusión sea procesada, el cerebro ya está evaluando activamente la información que podría contradecirla, reflejado en cambios como el N2, que indica una violación de expectativas, y el P3b, relacionado con la satisfacción de expectativas. Estos hallazgos subrayan que la recuperación de rebatidores y su integración en el razonamiento ocurren en etapas muy tempranas, y sugieren que la memoria de trabajo desempeña un papel crucial no solo en la recuperación de estos elementos, sino también en cómo gestionamos activamente su impacto en nuestras conclusiones lógicas.

16. Los Inicios de las Aproximaciones Probabilísticas al Razonamiento Rebatible

A principios de los años 90, Mike Oaksford y Nick Chater iniciaron el desarrollo de una teoría del razonamiento humano rebatible, considerando que su estudio era un campo crucial para una ciencia cognitiva adecuada. Sus primeros trabajos criticaban tanto los modelos logicistas predominantes en el razonamiento como la perspectiva general de la teoría de la demostración en la ciencia cognitiva (Chater y Oaksford, 1990; Oaksford y Chater, 1991). Proponían un modelo basado en la teoría probabilística bayesiana para entender los procesos cognitivos humanos. Durante la década de 1990, se desarrollaron numerosas teorías probabilísticas para explicar diversos fenómenos de razonamiento, incluidos los efectos de supresión de la inferencia. Chan y Chua (1994) exploraron cómo las inferencias con incertidumbre que se dan en el discurso cotidiano se ven afectadas por la percepción de los grados de relevancia percibida entre antecedentes y consecuentes en condicionales. Descubrieron que la percepción de relevancia, influenciada por el

conocimiento previo, afecta la magnitud del efecto de supresión. Por ejemplo, un antecedente como 'si nadie ha robado el motor durante la noche...' se consideraba más relevante para el arranque del coche que 'si no hubo niebla durante la noche...'. Estos resultados llevaron a los investigadores a concluir que los condicionales se interpretan como portadores de "una regularidad débil", como expresiones de probabilidades condicionadas.

Siguiendo esta línea, Stevenson y Over (1995) también estudiaron los efectos de supresión desde una perspectiva probabilística, realizando experimentos que mostraron cómo las personas tratan los condicionales como probabilidades condicionadas, variando sistemáticamente la probabilidad cualitativa del antecedente en sus estudios. Por ejemplo, se presentaba a los participantes ítems como el siguiente:

- Si se gira la llave, entonces el coche arranca.
- Si la batería está cargada, entonces el coche arranca.
- La batería está siempre (casi siempre, a veces, raramente, muy raramente) cargada.
- Se gira la llave.

Según Oaksford y Chater (2007), esa manipulación en la tercera frase es, directamente, una manipulación de $P(q | p)$. En los argumentos MP y MT, la aceptación de las conclusiones disminuyó correlativamente a como lo hacían los grados de probabilidad. En estudios adicionales sobre los efectos de supresión, la credibilidad de un condicional reemplaza a los rebatidores directamente. Weidenfeld, Oberauer y Hörnig (2005) señalan que la disponibilidad de excepciones a un condicional disminuye la probabilidad subjetiva de que el consecuente ocurra, dado el antecedente. Esta probabilidad subjetiva influye directamente en el grado de creencia en el condicional y, consecuentemente, en la disposición a aceptar inferencias consideradas lógicamente válidas.

En un estudio realizado por George (1995), los participantes evaluaron la credibilidad de diversas premisas condicionales en una escala de siete puntos. Las premisas, diseñadas para ser realistas, cotidianas y potencialmente controvertidas, como 'si el invierno es duro, entonces enfermará mucha gente', se usaron para construir argumentos de Modus Ponens (MP). A los participantes se les solicitó que, asumiendo la verdad de las premisas, evaluaran la validez de las conclusiones usando la misma escala. Aunque teóricamente el MP debería resultar en una aceptación completa de la conclusión si el condicional es cierto, solo el 48% de los participantes asignó la máxima puntuación, lo que indica que la aceptación no fue universal. Esto revela que, a pesar de instrucciones explícitas para que asumieran las premisas como verdaderas, los efectos del conocimiento previo y la credibilidad son complicados de suprimir. La correlación alta entre el grado de creencia asignado a las conclusiones y a las premisas condicionales indica que la desconfianza en la premisa lleva a desconfianza en la conclusión. Un experimento adicional mostró que solo el 43% de los participantes logró inicialmente ignorar sus creencias personales, mientras que un 98% se dejó guiar por ellas más tarde, destacando que la evaluación de plausibilidad es un proceso habitual pero no obligatorio. George interpreta que los participantes probablemente consideraron

las premisas condicionales desde una perspectiva probabilística, evaluando el soporte del antecedente hacia el consecuente y, por ende, su probabilidad.

George (1997) exploró cómo la expresión explícita de incertidumbre en las premisas condicionales afecta las conclusiones derivadas. Manipulando frases como 'si P, entonces es muy probable que Q' versus 'si P, entonces no es muy probable que Q', descubrió que la incertidumbre de las premisas influye directamente en la aceptación de la conclusión. En experimentos adicionales, al variar y combinar grados de incertidumbre en las premisas condicional y categórica, identificó que las probabilidades asignadas a las conclusiones eran igualmente bajas o incluso inferiores a la premisa menos creída. Este hallazgo respalda el modelo teórico propuesto en esta tesis, que predice dichos efectos de transmisión de incertidumbre.

George (1995, 1997) sugiere que, de forma similar a cómo los argumentos válidos en la lógica estándar preservan la verdad, los argumentos con premisas inciertas transmiten la duda a sus conclusiones, un concepto central en esta tesis. Además, propone que los argumentos cotidianos no se juzgan simplemente por su validez, sino por cómo confirman o apoyan la relación entre premisas y conclusión, lo que Osherson, Smith y Shafir (1986) describen como la *fuerza del enlace inferencial* o *fuerza psicológica del argumento*. Esta fuerza es frecuentemente más influyente que la validez lógica en cómo se aceptan las conclusiones, abarcando procesos de deducción, inducción y abducción.

17. Comentarios finales: del viejo al nuevo paradigma en psicología del razonamiento y de la argumentación

Históricamente, las tradiciones dialéctica y retórica se han centrado en cómo los mensajes lingüísticos influyen en la formación y modificación de creencias para persuadir. Desde Aristóteles, se considera que el aspecto central consiste en plantear el mensaje como un *razonamiento* sobre asuntos *probables para los oyentes*. Es decir, como un dispositivo que dispare ciertos mecanismos formales de inferencia que operen sobre grados de verosimilitud obtenidos a partir de la interpretación de su contenido en función del conocimiento previo y del contexto de una audiencia concreta.

La tradición de vincular la retórica y la dialéctica con la probabilidad, que incluso persistió tras la formalización matemática de la probabilidad por autores como Leibniz y Bernoulli, comenzó a diluirse en el siglo XX. Durante este período, la probabilidad empezó a desvincularse de las creencias, orientándose hacia interpretaciones basadas en frecuencias objetivas propuestas por figuras como von Mises y Fisher. No obstante, gracias a matemáticos como Frank P. Ramsey y Bruno De Finetti, la conexión entre probabilidad y creencias se mantuvo, considerando la probabilidad subjetiva como un reflejo de los grados de creencia. Paralelamente, la retórica y la dialéctica cayeron en desuso hasta su renacimiento en los años sesenta con Chaim Perelman y Stephen Toulmin y más tarde en los setenta con el movimiento de la Lógica Informal y la emergente Teoría de la

Argumentación. Dentro de esta renovada tradición, filósofos como Rescher, Walton o Pollock, vinculados también a la inteligencia artificial y las ciencias cognitivas, desarrollaron sistemas de razonamiento basados en la plausibilidad, rechazando orientaciones probabilísticas cuantitativas.

La psicología del razonamiento inicialmente adoptó la lógica formal como su marco normativo, ocupándose de determinar si los humanos se adherían a las reglas de la validez deductiva. Las respuestas que no seguían estas reglas eran vistas como errores sistemáticos, sugiriendo que la lógica formal podría no ser el marco normativo adecuado. Sin embargo, hacia finales del siglo XX y principios del XXI, investigadores como Oaksford, Chater, Evans, Over y Pfeifer comenzaron a priorizar la probabilidad subjetiva sobre la lógica formal. Esto marcó el inicio de lo que ahora se conoce como "el Nuevo Paradigma en psicología del razonamiento", donde la validez y la verdad lógica han sido reemplazadas por los grados de creencia y la validez probabilística.

Desde 2004, la investigación de Ulrike Hahn y sus colaboradores ha sido fundamental para integrar el razonamiento probabilístico con la teoría de la argumentación contemporánea. Sus estudios han cubierto una amplia gama de temas, incluyendo el análisis de falacias, la evaluación de la fuerza y calidad de los argumentos, y la relación entre la fuente y la calidad del mensaje. Además, han trabajado en establecer un marco normativo bayesiano para la argumentación, vinculando conocimientos de diversas disciplinas. Esta propuesta ha revitalizado la perspectiva clásica de la argumentación, que conjuga discurso, razonamiento y probabilidad personal en el ámbito de la persuasión.

A lo largo de la historia, el impacto del contenido en el razonamiento cotidiano ha sido un elemento central tanto en la retórica como en la dialéctica clásicas. Estas disciplinas han buscado influir en la audiencia y modificar sus juicios de verosimilitud mediante el uso estratégico del contenido de las proposiciones. Tales juicios, que varían en intensidad, se reflejan eficazmente a través de la probabilidad subjetiva. Por tanto, es posible que la psicología del razonamiento humano esté más estrechamente relacionada con las metodologías de la retórica y la dialéctica clásicas que de lo que hasta ahora se había pensado.

La psicología del razonamiento probabilístico ofrece un marco robusto para entender cómo las personas interpretan, generan y evalúan argumentos en condiciones inciertas. A diferencia de los enfoques basados en la lógica formal, este paradigma enfatiza la importancia de la probabilidad subjetiva y el conocimiento previo, ofreciendo así una descripción más precisa y psicológicamente realista del razonamiento humano. No es que el Nuevo Paradigma haya abandonado el formalismo, pues el cálculo de probabilidades es formal. Pero es sensible al contenido, en tanto que de ahí se derivan las magnitudes que posteriormente serán computadas formalmente. Estas investigaciones y desarrollos en la psicología del razonamiento probabilístico están redefiniendo cómo entendemos la argumentación, moviéndose hacia un planteamiento que valora la incertidumbre y la complejidad del pensamiento humano en su evaluación de la verdad y la validez.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

PARTE II: HACIA UN MODELO TEÓRICO INTEGRAL DEL SISTEMA DE CREENCIAS Y LA DINÁMICA DE LA CREDIBILIDAD

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

CAPÍTULO 1

Dinámicas de adaptación en sistemas de creencias: un modelo teórico de interacción, validación y cambio

El estudio de las creencias ha sido predominantemente abordado desde disciplinas como la filosofía epistemológica y ontológica, y la inteligencia artificial, con una atención limitada dentro de la psicología hasta fechas recientes. Este panorama está cambiando conforme se reconoce la importancia de las bases cognitivas de las creencias y su influencia en la conducta humana, comenzando a recibir atención empírica significativa en la psicología cognitiva (Connors y Halligan, 2015; Seitz y Angel, 2020). Van Dijk (2014) señala que los estudios sobre creencias se han centrado en aspectos normativos y en las relaciones entre creencias y el mundo, dejando de lado las estructuras y funciones cognitivas implicadas. La perspectiva racionalista, formalista y logicista de la psicología cognitiva, basada en los métodos de la teoría de la demostración, no puede explicar la no monotonicidad de la inferencia no demostrativa, esencial para la dinámica y actualización de las credibilidades. Además, las reglas sintácticas ciegas a los contenidos de las representaciones impiden la adopción de reglas probabilísticas que, pese a ser otro formalismo, dependen de los contenidos. Hernion (2003) propone la probabilidad subjetiva como la mejor herramienta para describir la dinámica de la incertidumbre y las credibilidades.

Contrariamente a la visión tradicional, la investigación contemporánea destaca la necesidad de una aproximación que aborde la no monotonicidad de la inferencia y la relevancia del contenido de la memoria a largo plazo. Esta visión actual sugiere que un modelo efectivo del sistema de creencias debe tratar estas como entidades complejas y jerárquicas, profundamente interrelacionadas y en constante interacción con el entorno físico y social. Este sistema no solo está destinado a contribuir a la supervivencia del individuo simplificando y esquematizando la realidad para reducir incertidumbre y facilitar decisiones, sino que también debe ser capaz de adaptarse y ajustarse ante los cambios, permitiendo la actualización y revisión continua de sus componentes.

Una propuesta sobre la estructura del sistema de creencias implica determinar y describir los componentes y las relaciones jerárquicas y funcionales que los unen. Es esencial describir cómo estos elementos interactúan dentro del sistema y los principios que rigen estas interacciones, lo cual nos permite entender cómo los individuos integran nueva información y ajustan sus percepciones y acciones de acuerdo con un marco de referencia que es tanto personal como culturalmente determinado. De este análisis emergen hipótesis que pueden ser exploradas y validadas a través de investigaciones empíricas, como se detallará en los experimentos subsiguientes de esta tesis.

1. Prerrequisitos en la formación de sistemas de creencias

Antes de adentrarnos en la exploración detallada de los componentes y la estructura de los sistemas de creencias, es fundamental comprender el contexto en el que estos sistemas operan, ya que el entorno influye directamente en su formación, evolución y funcionamiento. Este modelo teórico del sistema de creencias aborda de manera integral cómo las creencias individuales y culturales no solo emergen, sino que también se validan, se desarrollan y se actualizan dentro de contextos psicológicos y sociales complejos.

Al considerar las creencias como productos de la experiencia, la comunicación y la inferencia, este análisis pone de manifiesto cómo interactúan con el contexto para formar estructuras de conocimiento que son esenciales tanto para la supervivencia individual como para la cohesión social. Además, se enfatiza el papel de la comunicación discursiva en la formación y validación de estas creencias, resaltando su naturaleza intersubjetiva y el impacto de la coherencia y la credibilidad en la aceptación de nuevas informaciones.

Así, a lo largo de las páginas siguientes se profundizará en cómo estas dinámicas influyen en la construcción de un sistema de creencias coherente y significativo, esencial para la comprensión efectiva de la realidad y la auto-comprensión del individuo en su medio.

1.1. Fundamentos teóricos de las fuentes de formación de creencias

Las creencias se forman a partir de la interacción entre el individuo y su entorno, aunque no todos los seres vivos generan creencias. En general, se acepta que las representaciones mentales son una parte fundamental de las creencias y que, por tanto, sólo los seres vivos con cerebro tendrán la capacidad de albergarlas. Sin embargo, no basta con tener un cerebro capaz de generar representaciones; por ejemplo, se debate si seres simples como la ascidia, con un sistema nervioso básico, pueden formar representaciones primitivas del entorno. Algunos autores sostienen que sí (e.g., Glenberg, Jaworski, Rischal y Levin, 2007). Sin embargo, sería arriesgado asumir que albergan creencias sin caer en una posible antropomorfización. Este debate se extiende a lo largo de la escala evolutiva, desde los peces hasta los mamíferos (para una revisión, véase Andrews y Monsó, 2021 y

Andrews, 2020)¹⁷, y plantea interrogantes sobre qué condiciones deben cumplirse para que se generen creencias, lo cual a su vez depende de cómo definamos este concepto, aún sin una definición consensuada.

Se ha argumentado en capítulos anteriores que las creencias constan al menos de una representación mental y de su credibilidad, entendida como la confianza asignada a dicha representación según principios de probabilidad subjetiva. Esta credibilidad es una magnitud que se da en grados y que varía según el sujeto y su contexto temporal y espacial. En humanos, las creencias se agrupan en sistemas más o menos coherentes, y su dinámica respeta cierta consistencia. Las creencias son el producto de la interacción entre los mecanismos cognitivos del individuo y su entorno, originándose principalmente de tres fuentes: la experiencia directa, la comunicación y la inferencia.

1.1.1. La experiencia directa

La experiencia directa juega un papel crucial en la cognición, ayudando a acumular información en la memoria para disminuir la incertidumbre y promover la supervivencia mediante una interacción exitosa con el entorno. A través de los sentidos y procesos cognitivos, interactuamos con nuestro entorno y generamos representaciones del mundo, las cuales, aunque sujetas a error, fundamentan nuestras creencias.

La acumulación y abstracción de experiencias es fundamental para el aprendizaje, incluyendo tanto la categorización (Corter y Gluck, 1992; Nosofsky, 1988; Rosch, 1978) como el aprendizaje asociativo (e.g. Rescorla, 1967; Sedlmeier, 2002). Estos procesos facilitan la generalización y la formulación de predicciones sobre el futuro o hacer conjeturas sobre el pasado, fundamentando así la formación de hipótesis o creencias a partir de las evidencias observadas. Dichas creencias, que presentan variaciones en su credibilidad, se refuerzan mediante la observación de patrones consistentes; por ejemplo, si un individuo situado en un entorno observa que la mayoría de las veces que sucede A, también sucede B, podrá predecir B dado A. La frecuencia relativa de estos patrones estabiliza la credibilidad de nuestras representaciones entorno a un valor. Sin embargo, esta credibilidad no es estática; se ajusta con la incorporación de nueva información, reflejando la

¹⁷ En cualquier caso, la línea de argumentación aquí iniciada es relevante en varios aspectos. Uno de ellos, es el de dar un posible abordaje evolucionista al problema de la caracterización de las creencias. Si los humanos las tienen, será porque la capacidad de tenerlas se ha sometido a un proceso de selección natural con el resultado de que albergar creencias y contemplar posibilidades inciertas ha sido adaptativo para poder responder al medio de la forma más adecuada. Por otro lado, podría tener implicaciones acerca del formato de las creencias. Desde una perspectiva evolucionista, si los humanos las tienen, debería observarse algo parecido en otros animales inmediatamente anteriores. Si los animales no humanos no tienen lenguaje, pero sí tienen creencias, podría discutirse la posibilidad de que el formato de las creencias tuviera lugar en forma de proposiciones sentenciales análogas a “frases en la caja de creencias” que operan según un sistema de deducción natural (Bermúdez, 2003; Camp, 2009; Glock, 2000; 2010; Rescorla, 2009b). Se podría seguir sosteniendo que las creencias operan según un sistema combinatorio de pensamiento con unidades representacionales, pero redefiniendo qué se entiende por proposiciones, y qué se entiende por deducción.

naturaleza dinámica de nuestra interacción con el entorno y la adaptabilidad de nuestra comprensión del mundo ante información emergente.

El aprendizaje a partir de la experiencia incorpora, por tanto, la gradualidad y la incertidumbre propias de la probabilidad subjetiva (Rescorla, 2009) en el camino hacia la formación de creencias. Este proceso, similar a la calibración en epistemología (van Fraassen, 1983) o a los hábitos de formación de opinión ligados a la frecuencia de eventos (Ramsey, 1928), pone de manifiesto que las creencias varían entre individuos o incluso en el mismo individuo en momentos diferentes debido a la diversidad de experiencias.

Las creencias se registran en la memoria semántica y episódica como esquemas y modelos, los cuales llevan asociadas ciertas probabilidades subjetivas. Aunque aún no se comprende completamente cómo se organiza esta información en la memoria, es probable que existan diferencias significativas entre especies. Hay cierto consenso en que, en los humanos, se organizan en dominios dentro de sistemas de creencias complejos.

1.1.2. La comunicación

La comunicación, una experiencia de interacción social esencial, es otra fuente crucial de información que ayuda a reducir la incertidumbre y formar creencias. Los organismos comparten información sobre su entorno, incluyendo fuentes de alimento, peligros, estatus social y estados emocionales, mediante la comunicación. Esta función se amplía en los humanos, abarcando la instrucción, persuasión, cooperación, entretenimiento y, a veces, la manipulación. Generalmente, el objetivo es influir en las creencias del receptor.

Un modelo fundamental en este ámbito es la teoría matemática de la comunicación (Shannon, 1948), que introduce un enfoque cuantitativo para medir la información ('bits') transmitida entre emisores y receptores. Este modelo destaca la importancia del código, el canal y el potencial de distorsión ('ruido'), sentando las bases para entender cómo la información se comparte y se percibe, desde la comunicación animal hasta la humana y las telecomunicaciones.

1.1.2.1. La comunicación como proceso intencional e intersubjetivo

La comunicación se define aquí como el intercambio intencional y mutuo de información entre dos o más seres. Aunque las plantas realizan cierta comunicación química (e.g. Estabrook y Yoder, 1998; Karban, 2015), su falta de cerebro y, por ende, de mente, intenciones y creencias, las excluye de esta definición. Asimismo, algunos animales comunican de forma automática o instintiva sin necesidad de una audiencia, un tipo de comunicación que también queda fuera de esta definición por carecer de intencionalidad.

En cambio, la comunicación *intersubjetiva* involucra la transmisión de percepciones, motivaciones, pensamientos y emociones entre mentes de individuos (Trevvarthen, 2002), mediante emisiones que hacen uso de diversas modalidades perceptivas: olfativas, táctiles, visuales o auditivas. En humanos, se incluye el uso de posturas, gestos, proxemia, expresiones faciales, contacto físico, sonidos y, específicamente, el lenguaje, tono de voz y elementos paralingüísticos.

Estas emisiones alteran el entorno físico y actúan como estímulos ostensivos para captar la atención del receptor con la *intención* de compartir una información específica (Sperber y Wilson, 1986).

Además, este proceso comunicativo ocurre dentro de contextos ecológicos específicos que afectan tanto la selección del tipo de emisión como su interpretación. El significado pretendido de una misma señal puede variar en función del contexto. Por tanto, en la comunicación humana, no solo importan la codificación y decodificación de las señales, sino también los procesos inferenciales que permiten al receptor entender el mensaje intencionado del emisor, considerando el contexto en el que se produce el intercambio.

1.1.2.2. El discurso como paradigma de la comunicación intersubjetiva humana

El discurso se erige como el principal vehículo de la comunicación intersubjetiva entre seres humanos, empleando el lenguaje como su código esencial. Este proceso se desenvuelve en un contexto multifacético que abarca desde el *cotexto*, es decir, el conjunto de elementos lingüísticos que rodean una expresión, hasta la subjetividad de los interlocutores, incluyendo su conocimiento individual (*contexto epistémico individual*) y compartido (*contexto del terreno común*), su cultura, roles sociales, pertenencia a grupos, intereses, y el conocimiento de lo que es lo adecuado en cada situación comunicativa (*contexto social*). Además, el marco espacial y temporal en el que se desarrolla la comunicación (*contexto físico*) también juega un papel determinante.

Estos factores contextuales son cruciales tanto para la creación como para la interpretación del mensaje (Van Dijk, 1997), ya que los participantes construyen modelos mentales subjetivos del contexto en su memoria episódica (Van Dijk, 2006), fusionando así los aspectos lingüísticos con los contextuales. Esto dota al discurso de una riqueza informativa que trasciende su contenido lingüístico puro. Aunque individualmente, los enunciados puedan resultar ambiguos, el contexto global del discurso ayuda a resolver estas ambigüedades, orientando al receptor hacia la interpretación que el emisor pretende.

Sin embargo, la representación mental del mensaje elaborada por el emisor rara vez coincide de manera exacta con la interpretación del receptor. Esto se debe a las diferencias inherentes en la información y experiencias almacenadas en la memoria de cada uno, tanto en su cantidad como en sus contenidos y sus interrelaciones. Además, influyen sus perspectivas únicas y la comprensión del contexto comunicativo. Estas discrepancias, inherentes a la subjetividad individual, significan que la interpretación del mensaje siempre contendrá un grado de variabilidad. Y la credibilidad asignada a un mensaje depende en gran medida de esta interpretación subjetiva.

1.1.2.3. Comunicación intersubjetiva mediante el discurso y su relación con la credibilidad

Se ha reconocido que los modelos basados únicamente en el código lingüístico resultan insuficientes para explicar en profundidad el fenómeno comunicativo. Para una comprensión más integral de cómo se producen y se interpretan los mensajes, resulta esencial incorporar procesos inferenciales que consideren los contextos lingüístico, epistémico y social. A lo largo de los años, este modelo ha sido ampliado y ajustado gracias a las contribuciones de varios teóricos. Por ejemplo, Schramm en 1954 subrayó la necesidad de que emisor y receptor compartan un sistema de signos y reglas,

mientras que Jakobson en 1960 enfatizó el papel del contexto. Berlo, en 1960, añadió la importancia de las habilidades comunicativas, el entorno social y cultural, y las actitudes. En el ámbito de las ciencias cognitivas, se destacan los modelos sistémicos de la escuela de Palo Alto y los modelos pragmáticos de Grice (1989) y posteriormente de Sperber y Wilson (1986, 2004), quienes propusieron el modelo ostensivo-inferencial en su teoría de la relevancia, subrayando el proceso cognitivo en la comunicación.

A continuación, resumiremos algunos de los aspectos clave de la comunicación intersubjetiva. Nos centraremos en los procesos cognitivos involucrados en la comunicación discursiva y su relación con la formación de creencias.

1.1.2.3.1. *La producción del mensaje*

Consideramos la comunicación como un proceso inmerso en un contexto particular donde el emisor, movido por una intención clara de dar a entender algo, planifica y estructura su mensaje. Este proceso implica recuperar y combinar información de la memoria semántica y episódica, ajustándola a un contexto específico. Así, se forma un modelo mental coherente de la macroestructura del mensaje. Este modelo se mantiene activo en la memoria de trabajo durante la codificación lingüística (Van Dijk, 1998). A través de este mecanismo, el mensaje se organiza en proposiciones jerarquizadas según su conexión con el tema central, transformándose en una representación semántica destinada a la comunicación. La elección de palabras y la estructura del mensaje se guían por este modelo mental, decidiendo qué se explicita y qué se omite. Este complejo proceso, aunque aquí simplificado, subraya la riqueza y la complejidad de la producción discursiva en el contexto comunicativo.

1.1.2.3.2. *La comprensión del mensaje*

La señal altera el entorno del receptor. Al ser percibida, éste genera expectativas sobre su relevancia. La señal es decodificada en la memoria de trabajo, transformándose desde su estadio de señal perceptiva inicial hasta convertirse en una representación semántica proposicional. En este punto del proceso, el mensaje se pone en relación con la subjetividad del receptor, activando contenidos relacionados en su memoria a largo plazo. Parte de esta activación ocurre automáticamente, siguiendo una dirección ascendente ('bottom up'), y se propaga a otros elementos relacionados en la memoria. La activación y su propagación dependen de la red de conexiones existente en la memoria de cada persona, influyendo en cómo se interpreta la información recibida y, consecuentemente, en su credibilidad. Simultáneamente, otra parte de la información se activa de forma estratégica, mediante una dirección descendente ('top down'), basado en los intereses y expectativas del receptor en relación con su modelo del contexto. Estos intereses y expectativas determinan a qué partes de la información se presta atención. La activación de información en la memoria a largo plazo, junto con aspectos clave de la estructura semántica, iniciará dos tipos de procesos inferenciales: uno enfocado en conseguir una representación unificada y coherente de la información y otro centrado en la validación de dicha información.

Los procesos inferenciales de *coherencia* conectan la nueva información con la ya almacenada en la memoria, en particular, con el discurso previo y con el conocimiento del mundo. Estos procesos inferenciales enriquecen así la entrada lingüística con detalles no expresados directamente, pero esenciales para la cohesión y el significado del mensaje (Graesser, Singer y Trabasso, 1994). La efectividad de estos procesos, según diversos estudios (e.g. Nieuwland y Kuperberg, 2008; Schroeder, Richter y Hoever, 2008; Singer, 2013; Singer, Halldorson, Lear y Andrusiak, 1992), depende de la *validación*, que comprueba la consistencia de la nueva información con el discurso previo y con el conocimiento del mundo, identificando posibles contradicciones y desajustes (Richter, 2011; Singer, 2013). La entrada lingüística desencadena automáticamente la activación de información en la memoria estrechamente asociada con esa entrada, en un proceso pasivo conocido como resonancia (e.g. Albrecht y Myers, 1995; Albrecht y O'Brien, 1993; Myers y O'Brien, 1998; O'Brien y Albrecht, 1992). Entonces, de manera rutinaria, se comprueba la consistencia entre la nueva información y la activada en la memoria, evaluando así la plausibilidad y verosimilitud de la información recibida (Richter, Schroeder y Wöhrmann, 2009). Dado que los grados de credibilidad otorgados a un discurso derivan de la validación, la cual depende de la memoria a largo plazo —única en cada individuo—, el resultado es una credibilidad que se manifiesta en *grados de probabilidad subjetiva*. Por tanto, frente al mismo mensaje o argumento, diferentes personas pueden asignarle niveles de credibilidad variados.

1.1.2.3.3. El rol de la credibilidad del comunicador en el proceso de validación

La exposición hasta ahora ha explorado cómo la información nueva se integra con el conocimiento previo del receptor para formar una comprensión coherente, permitiendo la validación epistémica y determinando la credibilidad de la información. En situaciones donde el receptor carece de conocimiento previo relevante, puede descartar la información como irrelevante (Sperber y Wilson, 2004) o basar su aceptación en la credibilidad percibida del emisor.

La credibilidad del comunicador, determinada por su percepción como *experto* y *honesto* (e.g. Harris, Hahn, Madsen y Hsu, 2016; Walster, Aronson y Abrahams, 1966), puede influir significativamente en la aceptación de la información, incluso en ausencia de conocimiento previo sobre el tema. Si el oyente considera que es creíble, adoptará su opinión sobre el tema (Gaifman, 1988), mientras que, si considera que no lo es, tampoco lo será su mensaje (Hovland y Weiss, 1951). Cuando el comunicador se considera creíble, la información desconocida se acepta y se interioriza, pudiendo incluso olvidar la fuente comunicativa con el paso del tiempo (Kelman y Hovland, 1953). Las características de experticia y honestidad pueden venir dadas por cuestiones ajenas al discurso (p.ej. fama), o pueden crearse mediante técnicas propias del discurso (Collins, Hahn, von Gerber y Olsson, 2018). Este proceso se complica si el oyente ya posee conocimientos sobre el asunto, donde la credibilidad del mensaje y del emisor interactúan de manera bidireccional (e.g. Collins et al., 2018; Hahn, Harris y Corner, 2009; Hahn, Oaksford y Harris, 2013; Madsen, Hahn y Pilditch, 2020; Petty y Cacioppo, 1984; Wertgen y Richter, 2020).

1.1.2.3.4. *Estado final del proceso de interpretación: el modelo referencial del mensaje*

Richter (2015) argumenta que los procesos de validación son esenciales en la comprensión del mensaje, ya que actualizan y estabilizan el modelo situacional referencial, entrelazando la comprensión con la atribución de credibilidad. Para que la información nueva se integre coherentemente con el discurso previo y con los conocimientos almacenados en la memoria, estos últimos elementos deberán validar al primero evaluando su grado de consistencia. De este modo, a cada fragmento de información le corresponderá un grado de credibilidad resultante de su validación.

Si la información es coherente y resulta creíble, se incorporará como un componente de la red integrada que conforma el modelo mental referencial. Este, a su vez, se unirá a las creencias y conocimientos previos de su dominio en el sistema, dotado de una probabilidad subjetiva inherente. Un supuesto fundamental del presente modelo es que, durante este proceso de integración, la información nueva y la previa ajustan sus grados de credibilidad para mantener la consistencia del sistema dentro de los límites manejables a los que se puede atender. Una vez integrada, la nueva información participa en la selección de futuras entradas informativas sin necesidad de volver a requerir su validación (Schroeder, et al., 2008).

Si la información entrante resulta inconsistente con las creencias previas, existen varias opciones: puede aceptarse, eliminando creencias previas conflictivas (con todas sus consecuencias); puede buscarse una explicación que armonice ambas creencias; puede almacenarse sin aceptarse como verdadera, similar a mitos y leyendas, como la del Rey Arturo; o simplemente puede rechazarse, como la idea de que beber lejía cura enfermedades.

1.1.3. *La Inferencia y el razonamiento cotidiano*

Hasta aquí, hemos explorado dos tipos de interacciones entre el individuo y su entorno capaces de generar creencias: la experiencia directa, y la experiencia comunicativa. En ambos casos, hemos destacado el rol de la subjetividad, de la incertidumbre y de la intensidad variable de la credibilidad, ya fuera por la frecuencia de exposición a eventos o por la conexión de la nueva información con otras creencias. Estas interacciones generan creencias que se almacenan en una memoria a largo plazo dentro de un sistema de creencias relativamente estable, aunque no estático, sino dinámico, abierto a actualizaciones derivadas de la recepción de nueva información, caracterizándose por su naturaleza no-monótona, esto es, sensible a los cambios ambientales y contextuales. Los *procesos inferenciales del razonamiento cotidiano* combinan y reorganizan la información almacenada en la memoria, extrayendo conclusiones y enriqueciendo el conocimiento. Esto, a su vez, permite anticipar posibilidades, hacer predicciones, construir nuevas hipótesis y generar nuevas creencias. Estos procesos dotan al individuo de herramientas para resolver problemas de manera efectiva en contextos variables, facilitando una adaptación ágil al entorno¹⁸. Los debates sobre la teoría

¹⁸ Nótese que aquí sólo se está haciendo referencia a materiales aprendidos por medio de la experiencia. No se está hablando de los contenidos llamados analíticos, a priori, de relaciones de ideas o verdades de razón (según los diversos autores). Éstos se consideran conocimientos a los que se llega por la inferencia misma, como, por ejemplo, el principio de no contradicción, o que el todo es mayor que la parte, etc. o por la reflexión

normativa que rigen estos procesos inferenciales son amplios y continúan evolucionando. Independientemente de cómo se definan, estos procesos se complementan con la búsqueda y recuperación de información de la memoria a largo plazo y las dinámicas de la memoria de trabajo.

Respecto a la teoría normativa que guía los procesos inferenciales, existen principalmente dos enfoques: la lógica y la probabilidad. En psicología, hay estudios que tradicionalmente han priorizado la lógica, como las teorías de la Lógica Mental (e.g. Braine, Reiser y Rumin, 1984; Rips, 1994), mientras que otros han destacado la probabilidad, englobados hoy bajo el apelativo del 'Nuevo Paradigma' (e.g. Douven y Verbrugge, 2010, 2013; Geiger y Oberauer, 2010; Oaksford y Chater, 2007; Over, 2009). Además, la teoría de los Modelos Mentales, originalmente alineada con la lógica de modelos (Johnson Laird, 1983, 2001; Johnson-Laird y Byrne, 1991), ha evolucionado hacia enfoques más probabilísticos (Johnson-Laird, Khelmani y Goodwin, 2015; Khelmani, Byrne y Johnson-Laird, 2018; Khelmani, Lotstein y Johnson-Laird, 2014).

Desde el enfoque logicista, se considera que la memoria almacena los conocimientos como proposiciones, sobre las cuales operan las inferencias lógicas. Estas inferencias, centradas en la estructura sintáctica sin considerar el contenido específico de las proposiciones, siguen principios semánticos de valores de verdad. Por ejemplo, dadas las proposiciones $p \rightarrow q$ y p , se deducirá q de manera válida, independientemente de su contenido. Las proposiciones se componen de símbolos, configurando representaciones. El valor de verdad de estas representaciones se determina mediante su semántica referencial, la cual está influida por la intencionalidad y una teoría de correspondencia asociada. El valor de verdad resultante se preservará a través de todas las operaciones. Este enfoque, basado en la deducción natural, no abarca la inferencia no demostrativa ni contempla la no-monotonidad o la variabilidad de la incertidumbre de las inferencias en la vida cotidiana.

Desde el enfoque probabilístico, la información en la memoria puede no ajustarse necesariamente a formatos proposicionales. Por otro lado, el término "proposición" puede emplearse en un sentido amplio que equivale a "contenido". Estas representaciones son también intencionales al referirse a objetos del mundo. Pero en lugar de disponer de una teoría de la correspondencia que determine los valores de verdad de manera bivalente, este modelo incorpora la ambigüedad en forma de incertidumbre. Así, asigna a los contenidos grados de probabilidad subjetiva, permitiendo valores epistémicos polivalentes y cuantitativos. El nivel de credibilidad otorgado a una representación depende del conocimiento previo, cómo se adquirió la información y su fuente u origen.

Las inferencias probabilísticas computan y combinan estas probabilidades asignadas. Dadas dos premisas como "si se pisa el pedal del freno, el coche va más lento" y "se pisó el pedal del freno",

abstracta, como el "cogito ergo sum", o por el propio significado de los términos, como que ningún soltero está casado, o que todas las madres son hembras. De hecho, pese a que uno pueda creer totalmente que son verdad, es discutible que puedan ser considerados como creencias. Esto nos llevaría al debate acerca de las diferencias entre la creencia y el conocimiento. Pero no es este el lugar para ello.

se puede concluir que “el coche fue más lento”. Y en función de la probabilidad asignada a las premisas, se puede considerar que esa conclusión es cierta, muy probable, incierta, poco probable, o nada probable. Como se puede observar, lo que se dice en la conclusión y su grado de probabilidad son dos resultados distintos, aunque interrelacionados. Estas conclusiones pueden derivarse lógicamente, pero el foco de interés pasa a centrarse en las leyes que gobiernan los grados de creencia adscritos a los contenidos. La validez de la lógica deductiva tradicional no es el principio rector del razonamiento humano. Esto es algo reconocido ya desde hace décadas.

Las probabilidades subjetivas son sensibles a los contenidos. La base de las inferencias probabilísticas son los grados de credibilidad, que se derivan de validar la información entrante. En este sentido, la probabilidad subjetiva otorgada a un contenido nunca será absoluta, sino condicionada a la base de conocimientos o creencias del sujeto. Una vez establecidos, estos grados son combinados, ajustados y manipulados a través de los mecanismos inferenciales en el razonamiento. Así, frente a la misma evidencia A, alguien puede considerar que implica B mientras que otro puede considerar que no, o que casi nunca lo hace. De este modo, asignarán distintos grados de credibilidad a la ocurrencia de B a partir de la observación de A. Además, la credibilidad asignada a un contenido también puede variar con el tiempo incluso para el mismo individuo, ya sea por reconsiderar información previa o incorporar nuevos datos relevantes. La credibilidad, por tanto, es dinámica y se ajusta, reflejando la capacidad de las inferencias probabilísticas para abordar la incertidumbre y la evolución de las inferencias cotidianas. Esta flexibilidad es clave para la adaptación al entorno en un mundo en constante cambio.

1.2. Generación, estructura y función del sistema de creencias en la cognición

El punto de partida para la exposición de este modelo es la consideración del ser humano como un organismo que busca integrar todos los elementos de su vida mental en una unidad coherente y significativa. Se entiende que esta cohesión es esencial tanto para comprender la realidad como para la comprensión de uno mismo en el mundo. Esta tendencia a la integración mejora la fiabilidad y precisión de sus representaciones mentales y las conexiones entre ellas, permitiendo al individuo desarrollar un sistema de creencias coherente y fiable, aunque susceptible de errores. La fiabilidad de este sistema es crucial para interactuar eficazmente con el entorno, logrando objetivos y cumpliendo metas de manera segura. Por tanto, la adaptabilidad del sistema de creencias es fundamental para satisfacer las necesidades biológicas y psicológicas, promoviendo la supervivencia y el bienestar del individuo.

A continuación, esta exposición prosigue detallando dos tipos de representaciones fundamentales para este modelo de credibilidad: por un lado, los modelos mentales, que reflejan las experiencias personales concretas, y por otro, los esquemas o marcos, que capturan la estructura esencial de categorías de experiencias mediante la abstracción y generalización de experiencias individuales. Ambos tipos de representaciones organizan sus elementos a través de relaciones

cualitativas y semánticas de coherencia, y se agrupan por dominios formando parte de sistemas de representaciones más extensos interconectados por diversas relaciones.

En este modelo se propone que los sistemas de creencias se estructuran en función de las relaciones de coherencia, que son esenciales para una percepción unificada de la realidad (Connors y Halligan, 2015; Quine y Ullian, 1970; Thagard, 2000; Tullett, Prentice, Teper, Nash, Inzlicht, McGregor, 2013). En particular, en función de un subconjunto de tales relaciones, que recibe el nombre de *relaciones tópicas*. Consisten en criterios de relaciones binarias que permiten conectar ideas de manera intuitiva, permitiendo la transición de la justificación de unas a otras. Consecuentemente, facilitan la construcción y comprensión de argumentos. Estos criterios abarcan aspectos epistémicos, pragmáticos, contextuales y cuasi-formales, como, por ejemplo, la causalidad, la contrariedad, la semejanza, la definición, lo general-particular, y muchas otras. El concepto de relaciones tópicas está explicado con profundidad en el ap. C.

Este apartado 1.2. concluye destacando que las relaciones de coherencia y justificación son filibles y presentan distintos niveles de solidez. Esta variabilidad se basa en cuán frecuentemente se relacionan las unidades de información en la mente del individuo. A mayor coocurrencia, más sólidas serán estas relaciones, lo que llevará a una percepción más unificada de la realidad. Esto aumenta la sensación de la comprensión y el control del entorno, facilitando la toma de decisiones. En contraste, una menor coherencia en la representación y de justificación para su adhesión generan una sensación desagradable de disonancia cognitiva, desencadenando mecanismos para mitigarla.

1.2.1. In-divi-duo: el ser humano como organismo en busca de la unidad

El concepto de individuo, entendido como un ser humano que busca unificar sus experiencias, aprendizajes y percepciones en una entidad coherente y significativa, es una idea ampliamente aceptada en psicología (e.g. Connors y Halligan, 2015; Gawronsky y Strack, 2012; Festinger, 1957; Ryan y Deci, 2002; Tullet et al., 2013; van Kampen, 2019). Esta tendencia hacia la integración y coherencia implica asimilar nueva información y conectarla con estructuras cognitivas y subjetivas existentes (Kuhl y Fuhrmann, 1998; Piaget, 1971; Sugiura, Seitz y Angel, 2015). Se ve como una inclinación natural hacia la creación de conexiones dentro de la propia psique y con el entorno social (Ryan y Deci, 2002). Sin embargo, esta tendencia puede verse limitada por factores internos, como las capacidades de la memoria, o externos, como una falta de relación real con los eventos en el mundo. Lejos de ser un objetivo en sí mismo, este impulso hacia la coherencia se considera esencial para la adaptación al entorno. En efecto, posibilita la construcción de un "yo" integrado, crucial para la adaptación al medio social (e.g., Kirkpatrick y Ellis, 2006; Sugiura et al., 2015), y la evaluación crítica de la información, crucial para la formación del conocimiento (Kruglanski y Shteynberg, 2012). Ambas cuestiones son fundamentales para una toma de decisiones efectiva que promueva la supervivencia y el bienestar.

El ser humano necesita cubrir necesidades fisiológicas básicas, como alimentación, hidratación y seguridad, para sobrevivir. Además, según diversas investigaciones, para un bienestar óptimo,

también debe satisfacer necesidades psicológicas fundamentales, incluyendo la curiosidad, la experiencia, la autodeterminación, la competencia y la integración social (Ryan y Deci, 2002). La motivación para interactuar con su entorno surge de la necesidad de satisfacer estos requisitos, orientando su comportamiento hacia estímulos beneficiosos y alejándose de los perjudiciales. Para ello, es esencial desarrollar representaciones mentales precisas del mundo circundante. El sistema cognitivo humano facilita la percepción del entorno, la abstracción de patrones, la organización y el almacenamiento de información para su uso futuro, especialmente en la resolución de problemas. Comprender y razonar sobre el entorno, para resolver problemas y diseñar estrategias, implica conectar diversas piezas de información obtenidas a través de experiencias y comunicaciones. Estas piezas deben formar una red de representaciones coherente, fiable y estable (Schroeder et al., 2008). Así, un sistema integrado de conocimientos y creencias mejora la toma de decisiones y la adaptación al entorno.

Algunos psicólogos argumentan que los procesos cognitivos avanzados, como la comprensión y el razonamiento, están orientados a facilitar la acción diaria en un entorno variable y lleno de incertidumbres (e.g. Chater y Oaksford, 2004; Evans y Over, 1996; Gigerenzer y Goldstein, 1996; Glenberg, 1997; Glenberg et al., 2007; Oaksford y Chater, 2007; O'Brien et al., 2003; Schroeder et al., 2008; Zwaan, 1999). En el desarrollo de este modelo, sostengo que esta inclinación hacia la cohesión cognitiva fundamenta la comprensión, la validación de la información y la atribución de credibilidad.

1.2.2. Unidad en las representaciones: modelos mentales y esquemas

Desde que un individuo comienza a explorar su entorno para satisfacer sus necesidades, busca entenderlo, aspirando a construir una representación coherente, tanto a nivel particular como general, que sea consistente, estable y fiable. Las experiencias específicas se almacenan en la memoria episódica, interconectándose y formando modelos mentales de distintas situaciones (Johnson-Laird, 1983; Magliano, Radvansky y Copeland, 2007; Van Dijk y Kintsch, 1983). Estos modelos, definidos por su contexto espaciotemporal, son isomórficos y multimodales, estructurando los elementos de cada situación, sus características y las relaciones entre ellos. Estas relaciones estructuradas proporcionan coherencia a la representación en su conjunto.

Cuando una experiencia se repite a menudo, el sistema cognitivo agrupa sus representaciones, aplicándoles procesos de generalización, abstracción y descontextualización (Sweegers y Talamini, 2014; Van Dijk, 2012). Esto da lugar a los denominados marcos o esquemas (Collins y Loftus, 1975; Graesser y Nakamura, 1982; Schank y Abelson, 1977), que se almacenan en la memoria semántica. A diferencia de los modelos de situación específicos, los esquemas no son multimodales¹⁹, sino representaciones genéricas formadas al abstraer las regularidades de experiencias particulares. Organizados jerárquicamente, codifican los componentes, atributos y relaciones típicas de

¹⁹ No hay un acuerdo general sobre esta última afirmación. Muchos teóricos de la perspectiva del *embodied cognition* (e.g., Barsalou, 2008; Matheson y Barsalou, 2018) defienden que incluso los conocimientos genéricos son multimodales.

situaciones estereotipadas, permitiendo que sus parámetros (o variables) se especifiquen con información concreta (o constante) durante la comprensión, volviendo a formar modelos de situaciones específicas para su aplicación práctica.

La identificación de regularidades y relaciones suele ser un proceso automático e incidental, no necesariamente deliberado (Sweegers y Talamini, 2014). Así, las personas aprenden automáticamente asociaciones como que los truenos preceden a la lluvia, que la ausencia de nubes implica falta de lluvia, que un proyectil lanzado resulta en la destrucción de algo, que ingerir una seta amarilla causa enfermedad, o que los objetos de madera o metal son duros, etc. En esencia, vinculan distintos eventos en unidades de información mediante *criterios de asociación*, creando así una visión integrada de la realidad.

Uno de los primeros problemas que enfrentamos es detallar estos criterios de asociación. Una condición esencial para identificarlos es que la mente los maneje con facilidad al conectar ideas, lo cual debe reflejarse en su uso frecuente. En este contexto, los discursos ofrecen pistas valiosas sobre cómo las personas relacionan diferentes unidades de información, reflejando así cómo los individuos tejen su percepción de la realidad (Magliano et al., 2007).

1.2.3. La coherencia: relaciones semánticas y pragmáticas

En la psicología del discurso, estos criterios de conexión reciben el nombre de relaciones de coherencia. Autores como Sanders, Spooren y Noordman (1993) y van den Broek, Bohn-Gettler, Kendeou, Carlson y White (2011) consideran que la coherencia está constituida por relaciones semánticas y pragmáticas entre las unidades de información. El establecimiento de estas relaciones depende de que los usuarios del lenguaje accedan a los contenidos del discurso y de los conocimientos que tengan acerca de estos contenidos. Así, una misma narrativa puede resultar coherente para alguien familiarizado con sus contenidos, mientras que para otra persona puede no serlo, reflejando cómo estas conexiones se construyen individualmente en la mente (Van Dijk, 1977, 2012).

Dentro de las relaciones de coherencia, la causalidad es especialmente relevante para la comprensión de los eventos del mundo en general y de los textos en particular (e.g., Magliano, Trabasso y Graesser, 1999; Zwaan, Magliano y Graesser, 1995). Además, se han reconocido otras relaciones como semejanza, oposición, contraste, temporalidad, atribución (Wellner, Pustejovsky, Havasi, Rumshisky y Saurí, 2006), así como espaciales, referenciales y lógicas (Kintsch, 1998; van den Broek, 1994; Zwaan et al., 1995). Estas categorías incluyen subtipos tales como consecuencia, explicación, violación de expectativas, reciprocidad, generalización, definición, elaboración, ejemplificación (Wellner et al., 2006), e intenciones de los personajes (Graesser et al., 1994). Sin embargo, aún falta consenso sobre su clasificación y tipología (Hovy y Maier, 1995; Sanders et al., 1993; Scholman, 2019). Aquí asumimos que todas estas relaciones estudiadas en el discurso les son heredadas del sistema de creencias, que es la base de la interpretación del discurso.

1.2.4. La justificación: relaciones tópicas

Los sistemas de creencias van más allá de meros sistemas conceptuales; sus contenidos se consideran como probablemente ciertos en función de las relaciones de apoyo y refutación que se establecen entre ellos. Estas son las llamadas relaciones tópicas. Estas relaciones transfieren la credibilidad o justificación de unas ideas a otras mediante criterios semánticos, como los de causalidad, consecuencia, temporalidad, semejanza, contraste, reciprocidad, categorización, definición y ejemplificación, entre otras, junto con las intenciones y objetivos de los actores involucrados (Dyck, 2002; Schellens y De Jong, 2004; Warnick, 2000). En este sentido, muchas relaciones de coherencia son, a su vez, relaciones tópicas²⁰. Sin embargo, existe una falta de consenso sobre el número y la clasificación de estas relaciones, que varía según el autor y puede resultar en una organización compleja (Olmos, 2012). Según se discutió en el cap. 1, pt. I (y también en ap. C), los tópicos son relaciones binarias que sustituyen la implicación lógica (Dyck, 2002; Vega, 2004), conectando cabalmente dos piezas de información y contribuyendo al sentido y coherencia del argumento. Todas las relaciones tópicas son, a su vez, relaciones de coherencia.

Así pues, las relaciones tópicas son, esencialmente, relaciones de *justificación* entre unidades de información y tienen la función de servir de base para fundar hipótesis (tal como quedó argumentado en la sec. 5, cap. 1, pt. I). Actúan enlazando una unidad ya aceptada con otra hipotética, *garantizando* la transmisión de justificación. Como parte de las relaciones de coherencia, su establecimiento es subjetivo y varía entre individuos; lo que una persona puede ver como una justificación sólida, otra puede que no. En la percepción de cada persona, la fuerza de conexión entre estas unidades influye directamente en la credibilidad otorgada a la hipótesis. Si la razón está completamente aceptada por el sujeto y la considera completamente suficiente para la hipótesis bajo cuestión, el cumplimiento de esta hipótesis se considerará necesario. Este modelo sostiene que las diferencias en la aceptación de la razón o en su percepción de suficiencia resultan en variaciones en la credibilidad de la hipótesis.

1.2.5. Unidad en la representación para la seguridad en la acción

Los individuos conectan eventos de su experiencia siguiendo ciertos criterios, lo que les permite comprender el mundo de manera coherente y tomar decisiones que mejoren el éxito de sus acciones. Estas conexiones entre eventos se almacenan en la memoria a largo plazo, en forma de modelos y esquemas, creando un sistema interrelacionado. Si los eventos A y B se asocian mediante estos criterios, se consideran relevantes entre sí, de manera que la presencia de A (o B) indica la posible existencia de B (o A). Así, los criterios tópicos no solo facilitan una comprensión ordenada del mundo, sino que también apoyan la justificación de hipótesis sobre posibles acontecimientos.

Las relaciones de coherencia, incluyendo las relaciones tópicas, pueden ser falibles o derrotables (Lascarides y Asher, 1993; Walton y Reed, 2002). Esto implica que el razonamiento

²⁰ Alguien podría argüir que el esquema de argumentación basado en la 'opinión del experto' no es una relación de coherencia. Y no lo es. Pero tampoco es una relación tópica, sino una *fuentes de credibilidad*. Así, es posible que no todos los esquemas de argumentación contengan relaciones tópicas.

basado en esquemas y modelos que las aplica es no-monótono²¹: sin necesidad de negar unidades informativas ya aceptadas en el sistema, las conclusiones que se pueden extraer pueden variar con la adición de nueva información. Retomando los ejemplos anteriores sobre relaciones asociativas, puede haber truenos sin lluvia, como en las tormentas secas, o proyectiles que no rompan nada, o setas amarillas comestibles, etc. Pueden darse casos donde no suceda lo que suele ocurrir la mayoría de las veces, mostrando la falibilidad de las reglas. La confianza en estas relaciones binarias depende de la cantidad de veces que se observa que un evento A precede o no a un evento B (Meyniel, Schlunegger y Dehane, 2015), y esta confianza puede ajustarse con nuevas observaciones. Según algunos estudios, una vez establecidas las hipótesis, se ven reforzadas o debilitadas según su éxito o fracaso en guiar acciones que logran objetivos valiosos para el individuo, pudiendo incluso desvanecerse tras errores costosos en la predicción (e.g. Connors y Halligan, 2015; Seitz y Angel, 2020).

Por lo tanto, el sistema genera representaciones coherentes de la realidad cuando sus componentes se ajustan bien entre sí. La estabilidad de estas representaciones proviene de la solidez de sus conexiones, que a su vez se basa en justificaciones provenientes de la experiencia o de otras fuentes de formación de creencias mencionadas previamente. Así, algunas representaciones resultan más integradas y coherentes que otras. Una representación más unificada intensifica la percepción de comprensión del entorno, lo que facilita la toma de decisiones (Tuliet et al., 2013). En consecuencia, desarrollar un sistema de creencias estructurado y fundamentado potencia el sentido de control y confianza en las acciones, optimizando la supervivencia y reduciendo la ansiedad ante lo desconocido.

Al percibirse una incoherencia en la representación o una falta de justificación para mantenerla, se activan procesos cognitivos destinados a restaurar la coherencia. Si falta un criterio de relación, el sistema buscará uno adecuado para conectar los segmentos de información (Connors y Halligan, 2015; Tuliet et al., 2013; van den Broek et al., 2011). Ante la ausencia de justificación, una inconsistencia o contradicción, se experimenta una sensación desagradable de disonancia que impulsa esfuerzos para resolver dicha discrepancia en la representación de la situación (Festinger, 1957; Tuliet et al., 2013; van Kempen, 2019). Como se ha mencionado repetidamente en esta tesis, un sistema de creencias contradictorio resulta ineficaz.

²¹ Este es un término procedente de la filosofía, pero cada vez más común en psicología del razonamiento. Brevemente, un sistema no monótono, también conocido como no monotónico, es un tipo de sistema lógico o de razonamiento que permite la modificación de conclusiones a medida que se adquiere nueva información. En un sistema no monótono, a diferencia de los sistemas monótonos tradicionales, añadir nueva información al conjunto de premisas puede invalidar conclusiones previamente válidas. Esto es crucial en contextos donde el conocimiento es incompleto o incierto, permitiendo que el sistema se adapte y revise sus inferencias a medida que emerge nueva evidencia.

1.3. El colectivo y la diversidad: sistemas culturales de creencias y su rol en la comunicación

En esta sección se desarrollará el concepto de los sistemas culturales de creencias y su rol en la comunicación. Partimos de dos observaciones importantes: primero, es un hecho que existen diferencias en los sistemas de creencias de una misma persona en diferentes momentos, o entre distintas personas, independientemente del momento. Segundo, estas diferencias no se deben a una disparidad en los procesos cognitivos que construyen las creencias, sino a la acumulación de experiencias diversas, que a su vez generan representaciones y relaciones diferentes.

Se sugiere que las diferencias culturales entre grupos humanos derivan en gran medida de sus experiencias, que a su vez están influenciadas por sus ubicaciones geográficas. Los grupos humanos, que también son grupos culturales, comparten representaciones basadas en experiencias similares originadas por vivir en entornos ecológicos comunes. Sin embargo, cabe destacar que esta explicación representa una simplificación intencionada de la realidad, aspecto que se desarrollará más adelante en el texto. Las representaciones compartidas por la mayoría de los miembros de un grupo cultural forman su sistema de conocimientos y creencias, organizadas en modelos y esquemas que reflejan lo que se considera altamente probable dentro de su contexto físico y cultural. Estas representaciones también delimitan las comunidades epistémicas y establecen un terreno común necesario para la comunicación eficaz y la comprensión mutua entre hablantes y oyentes. Mediante el discurso, los hablantes como padres, profesores y periodistas transmiten la forma en que una sociedad entiende el mundo, incluyendo sus tradiciones, prejuicios, juicios y riquezas conceptuales. A partir de representaciones concretas y socialmente compartidas, el discurso permite comunicar contenidos desconocidos, abstractos y desvinculados de la experiencia directa. En este proceso comunicativo, los oyentes interiorizan gradualmente estas creencias culturales. A medida que estas creencias se integran, la información creída y asimilada se conecta de manera coherente y justificada con la información previa, influyendo así en cómo los individuos interpretarán nueva información, adoptando perspectivas subjetivas, particulares, singulares, pero influidas por la cultura.

Este apartado 1.3. concluye destacando dos puntos clave: primero, aunque los entornos geográficos influyen en los contenidos mentales de sus habitantes, las comunidades epistémicas siempre han interactuado y se han influenciado mutuamente, demostrando la naturaleza interconectada de las culturas. Segundo, esta interacción resalta el carácter dinámico de los sistemas culturales de creencias, que están contruidos sobre conocimientos limitados y relaciones falibles. Por lo tanto, los razonamientos que utilizan estos contenidos como premisas resultan ser no monótonos, reflejando la flexibilidad y adaptabilidad de estas estructuras de creencias.

1.3.1. Muchos entornos y una mente, muchas mentes en un entorno

Como hemos discutido anteriormente, la abstracción y organización de experiencias en una estructura coherente y justificada son cruciales para la formación de redes de creencias. A nivel individual, estas redes evolucionan con las experiencias adquiridas. Además, es muy poco probable

que dos personas posean sistemas de creencias idénticos debido a la diversidad de experiencias y la información específica que cada una recibe, lo que resulta en variabilidad interindividual. Por lo tanto, las diferencias en las creencias, tanto dentro de un individuo a lo largo del tiempo como entre individuos, se deben principalmente a las variaciones en sus interacciones con el entorno, y no a diferencias inherentes al proceso de formación de creencias.

1.3.1.1. Diferentes representaciones de la realidad entre diferentes grupos geográficos

Si aceptamos que existen diferencias en el sistema de creencias de una persona en distintos momentos, o entre dos personas diferentes, no debería sorprendernos encontrar variaciones similares entre grupos o poblaciones. Estas diferencias operan bajo la misma lógica, aunque a una escala mayor. Por tanto, es crucial definir qué entendemos por 'sistema de creencias' de un grupo o población. Numerosos estudios (Brown, 1991; Chirkov y Ryan, 2001; Demetriou, Kui, Spanoudis, Christou, Kyriakides, y Platsidou, 2005; Eibl-Eibesfeldt, 1989; Ryan y Deci, 2002) concuerdan en que la arquitectura mental y los mecanismos del sistema cognitivo humano, así como los motivos fisiológicos y cognitivos que impulsan la acción, son universales para toda la especie, conocidos como 'unidad psíquica' y 'universales humanos' respectivamente. Sin embargo, las variaciones individuales en los inputs informativos provienen de las interacciones con sus respectivos entornos físicos y sociales.

Cada grupo humano ocupa un espacio físico único, asentado en territorios con características distintas de orografía, clima, fauna y flora. Aunque las leyes físicas que rigen la naturaleza son universales, los elementos específicos de cada entorno varían en cualidad y cantidad. Por consiguiente, lo que ocurre habitualmente en un lugar puede diferir significativamente de lo que sucede en otro. Sin embargo, los mecanismos cognitivos de categorización, inferencia y comprensión, que contribuyen a la interpretación de la realidad, son constantes en toda la especie. Estos procesos cognitivos uniformes aplicados a distintos entornos variables generan experiencias diversas. Esto lleva a que las regularidades abstractas derivadas de esas experiencias difieran y, por tanto, las representaciones mentales formadas por los miembros de diferentes grupos geográficos también varíen, especialmente aquellas dependientes del entorno físico.

Si bien las categorías genéricas abstractas pueden ser comunes en diferentes entornos, los ejemplares típicos de estas categorías a menudo varían: el ave típica podría ser una paloma gris en un lugar y una urraca en otro; la silla típica podría ser de madera aquí y de mimbre allá. Igualmente, modelos y esquemas más complejos para la misma categoría también varían según el entorno donde se desarrollan. Además, la frecuencia de coocurrencia condicional de eventos puede diferir en diversas latitudes, afectando su observación y abstracción. Por lo tanto, las características del medio ambiente donde residen distintos grupos humanos son parcialmente responsables de las diferencias en algunas características de sus representaciones, así como en los contenidos y la estructura de sus sistemas de creencias. Esto lleva a variaciones significativas entre grupos geográficos (Shore, 1996).

1.3.1.2. Similares representaciones de la realidad dentro de un mismo grupo geográfico

Si un mismo sistema cognitivo produce diferentes sistemas de creencias en entornos ecológicos distintos, es razonable suponer que sistemas de creencias similares surgirán entre miembros de un grupo que comparte el mismo entorno ecológico, utilizando el mismo sistema cognitivo. Aunque estas creencias serán *similares*, no serán idénticas, como se ha discutido anteriormente.

Los individuos de un grupo arraigado en un mismo lugar comparten un conocimiento genérico similar de ese entorno, a pesar de las variaciones subjetivas. Estos individuos están continuamente expuestos a objetos, eventos y relaciones específicas de su entorno, familiarizándose con las frecuencias relativas con las que ocurren instancias de estas clases genéricas. Esas frecuencias se distribuyen de manera relativamente uniforme en el entorno habitual de los observadores / miembros de un grupo. Consecuentemente, la mayoría habrá almacenado en su memoria ejemplares parecidos como los típicos de ciertas categorías. Las representaciones de objetos, sus características y las relaciones entre eventos tienden a ser similares, llevando a un *conocimiento compartido sobre lo que es probable* que ocurra en su entorno habitual (Boyd y Richerson, 1988). De este modo, los miembros del grupo geográfico emplearán las representaciones de los elementos comunes de su entorno para interpretar y explicar la realidad. Estas interpretaciones y explicaciones lo transformarán ahora en un grupo cultural.

Dada la diversidad de entornos y la universalidad de los procesos y motivos cognitivos, distintos grupos ofrecerán respuestas variadas a preguntas similares sobre la realidad. Estas diferencias se reflejan en visiones del mundo diferentes sobre lo que es verdad o no, lo beneficioso y lo perjudicial, promoviendo pautas conductuales que constituyen diversas tradiciones y, en última instancia, generan la diversidad cultural (Henrich y McElreath, 2003).

1.3.2. El rol del sistema cultural de creencias en la comunicación: el terreno común

A medida que las personas se integran en grupos sociales y se comunican entre sí, sus creencias individuales se entrelazan en redes complejas, formando visiones generales del mundo. Desde la perspectiva de la antropología cognitiva, la cultura se entiende como un sistema cognitivo idealizado—un conjunto de conocimientos, creencias y principios—que reside en las mentes de los miembros de una sociedad (Casson, 2002). Este sistema incluye estructuras cognitivas jerárquicas como modelos y esquemas culturales (Holland y Quinn, 1987), que son paquetes de conocimientos y creencias culturalmente definidos. Estos esquemas organizan el conocimiento sobre objetos, situaciones, eventos y acciones que son socialmente compartidos, delineando lo que se considera probable en un entorno particular.

Uno pertenece a una cultura de referencia cuando comparte contenidos cognitivos con el grupo cultural (Van Dijk, 2013). Estos conocimientos y creencias, aceptados por la mayoría de los miembros del grupo, son esenciales para conformar y delimitar una comunidad epistémica (Van

Dijk, 2012)²². Además, este sistema de representaciones comunes facilita un terreno común (Clark, 1996) que es fundamental para intercambios comunicativos eficientes²³. En particular, los principios y creencias compartidos por una comunidad cultural proveen la base para la semántica referencial, necesaria para la producción y comprensión del discurso, sustentando la intersubjetividad en la comunicación lingüística.

Es importante recordar que la aceptación de estos conocimientos, creencias y principios es general, pero no universal. Como se ha mencionado, incluso en condiciones cognitivas y ambientales idénticas, pueden surgir variaciones interindividuales tanto en los contenidos del sistema de creencias como en su aceptación. El estudio de estas diferencias dentro de un grupo cultural es el foco de la teoría del consenso cultural (Romney y Batchelder, 2002; Weller, 1987) y de la investigación sobre variación cultural (Boster, 2002; Wallace, 1961).

1.3.3. Comunicación de contenidos y creencias culturales

Las mentes no se desarrollan de manera aislada, sino dentro de un entorno social y cultural. A través de la interacción con otros miembros de su cultura, las personas adquieren conocimientos y creencias que pueden originarse de dos maneras principales: de forma implícita e incidental mediante el aprendizaje social observacional, y de manera explícita y propositiva a través de la comunicación discursiva.

El discurso es el único medio para adquirir conceptos abstractos y de alto nivel, típicamente resultantes de la interacción social, como 'democracia', 'paradigma', 'educación', o términos más metafóricos como 'inflación económica', 'economía sumergida', y 'oleadas de inmigrantes' (Van Dijk, 1999). Estos conceptos, al ser abstractos, no pueden experimentarse directamente. Asimismo, la comunicación discursiva es una fuente esencial para formar creencias sobre aspectos culturales no experimentados directamente, sino transmitidos por la tradición de la comunidad. A través del discurso, ya sea oral o escrito, se transmiten los contenidos culturales: no solo sobre tipos de objetos y eventos que suelen ocurrir, sino también sobre las interrelaciones establecidas a lo largo del tiempo y sus transformaciones, incluyendo leyendas, religiones, ideas abstractas y el conocimiento acumulado que posibilita el progreso científico y tecnológico de generación en generación. Los contenidos transmitidos por el discurso pueden ser considerados como medianamente o muy ciertos por los hablantes, y así formar parte de un sistema de creencias que se desliga de la experiencia directa.

²² Una comunidad epistémica queda definida por el conocimiento compartido, no por una localización geográfica. Por ejemplo, el grupo de los oncólogos formarían una comunidad epistémica, y otra los entomólogos, y otra los forofos de un equipo deportivo particular, y así en más. Según Van Dijk (2003), las comunidades epistémicas no son meramente grupos sociales o instituciones, sino comunidades de práctica, pensamiento y discurso.

²³ El terreno común (o *Common Ground*) puede hacer referencia al conocimiento compartido por una comunidad epistémica, tal como se apunta aquí, pero también puede consistir en un conocimiento específico compartido sólo por dos personas en una conversación dada.

El discurso permite a los individuos de una cultura reproducir, conservar o transformar conocimientos y creencias, como los esquemas y modelos culturales. Principalmente, este proceso ocurre a través de figuras autoritarias como los padres, los medios de comunicación o la educación formal (Van Dijk, 2014). De este modo, los modelos y esquemas culturales se aprenden y se distribuyen en toda la sociedad.

1.3.3.1. El prisma epistémico: interiorización del sistema cultural de creencias

Al interiorizar el sistema cultural de creencias, este influye en cómo el individuo interpreta y juzga nuevas informaciones. Con este marco cultural, un mismo hecho puede ser interpretado de manera diferente a como se haría antes, estableciendo criterios para evaluar verdades, beneficios, y valores morales.

Así, el sistema cultural, interpuesto entre el individuo y el mundo, se convierte metafóricamente en un *prisma epistémico*. Junto con las experiencias personales, forma parte de la subjetividad del individuo y afecta la credibilidad asignada a nuevas informaciones. Al constituir parte de la base de conocimientos del sujeto, funcionará para conectar, contrastar y validar la información entrante. De este modo, la cultura influye sobre la interpretación de los eventos de la realidad.

1.3.3.2. Distribución de los sistemas culturales de creencias

En cuanto a la difusión del sistema cultural de creencias, este es sólo un punto de partida. A lo largo de la historia, el contacto entre comunidades geográficas y epistémicas —a través de comercio, guerras, peregrinaciones, conquistas y migraciones— ha mezclado y actualizado mutuamente los sistemas culturales. Estas interacciones han enriquecido y modificado las creencias y conocimientos de las culturas involucradas. Debido a las limitaciones espaciales y temporales de los grupos humanos, el conocimiento dentro de sus sistemas culturales también es limitado. Si un individuo solo conoce los elementos y relaciones de su entorno, podría no estar al tanto de las excepciones o contraejemplos a sus generalizaciones, creyendo que lo que ocurre comúnmente en su entorno ocurre de igual manera en todo el mundo. El contacto intercultural puede ampliar este conocimiento, modificando las conclusiones previas de una cultura basadas solo en su experiencia local. Esta actualización se realiza incorporando nueva información que no niega la experiencia previa, sino que la complementa.

Podríamos considerar que la información aceptada por una cultura son unidades de información equiparables a premisas. Supongamos que estas constituyen el conjunto de premisas de un sistema cultural. Si avanzamos hacia una terminología lógica o filosófica, podríamos considerar lo siguiente: un conjunto original de premisas, denotado como Γ , que inicialmente derivaba en una conclusión α , podría, al incorporar un nuevo conjunto de premisas, Δ , llevar a una conclusión opuesta, $\neg\alpha$. Esto ilustra la falibilidad y la no monotonicidad inherentes a los sistemas de creencias, demostrando cómo la adición de nuevo conocimiento puede cambiar nuestras conclusiones.

La formación del sistema individual de creencias y la interiorización del sistema cultural son procesos graduales que se desarrollan a lo largo de la vida del individuo. El prisma epistémico es dinámico y se ajusta progresivamente al integrar nueva información. Además, el sistema cultural de creencias también puede experimentar cambios a lo largo del tiempo debido al contacto con otras culturas o a cambios de paradigma dentro de la misma cultura, influenciados por factores políticos, sociales, científicos, artísticos o filosóficos. Aunque la presentación de estos procesos se ha simplificado aquí para facilitar la comprensión, es importante reconocer que la realidad es mucho más dinámica y compleja, con numerosos elementos interconectados en la construcción de los sistemas de creencias.

1.4. Conexiones entre la argumentación y los sistemas de creencias: lo probable y lo plausible

La comprensión de cómo los sistemas de creencias varían entre y dentro de los grupos geográficos nos lleva a considerar su impacto en las prácticas comunicativas y argumentativas. Estos sistemas no solo influyen en la percepción individual de la realidad, sino que también definen las bases de la argumentación dentro de las comunidades. Las creencias compartidas, forjadas a través de interacciones discursivas y experiencias comunes, se convierten en fundamentales para la argumentación retórica y dialéctica. Estas creencias, entendidas como *édoxa*, o conocimientos generalmente aceptados, juegan un papel crucial al proporcionar el terreno común necesario para la comunicación efectiva y la persuasión argumentativa. Al avanzar hacia una discusión sobre el rol específico de los sistemas culturales de creencias en la comunicación, se destacará cómo estos sistemas, estructurados y difundidos a través del discurso y la interacción social, facilitan o dificultan la transmisión de ideas dentro de y entre los grupos. Así, por todo lo explicado hasta aquí, ya pueden vislumbrarse las conexiones obvias que existen entre los sistemas de creencias individuales o culturales con la argumentación dialéctica y retórica.

A continuación, se aborda la relación entre las creencias derivadas de experiencias individuales y su aplicación en la argumentación mediante entimemas. Se explicará cómo los conceptos de modelos y esquemas coherentes facilitan el razonamiento rebatible, ilustrando brevemente su uso en la modificación de credibilidades. Se discute cómo los sistemas culturales de creencias influyen en la argumentación retórica y dialéctica, enfocándose en su función como *édoxa* — opiniones generalmente aceptadas usadas para justificar argumentos. Esta sección destaca la importancia de considerar estas creencias compartidas al formular argumentos para un grupo específico, sugiriendo el término 'sistema endoxástico' para describir esta interacción entre cultura y argumentación. Se resalta la necesidad empírica de adaptar los entimemas a las opiniones comúnmente aceptadas del grupo al que se dirigen.

1.4.1. Individual – Experiencial

La experiencia directa es una fuente clave en la formación de creencias. Las personas adquieren representaciones sobre lo habitual al *observar* la frecuencia de ciertos sucesos en su entorno. Así, las creencias son representaciones acerca de *relaciones* observadas entre distintos elementos. Pero la observación sólo permite un acceso incompleto a la naturaleza, por lo que esas representaciones estarán sujetas a la falibilidad. Por ejemplo, al notar repetidamente que la presencia de nubarrones suele preceder a la lluvia, se forma el esquema genérico de que 'cuando hay nubarrones, llueve'. Pero la experiencia de casos de nubarrones sin lluvia nos indica que este esquema es falible. Es, por tanto, un esquema que se sostiene en cierto grado. Supongamos que un día se observan nubarrones; basándose en su esquema, una persona puede prever que probablemente lloverá esa tarde. La persona se adhiere con una fuerza relativa a este modelo mental de una situación hipotética en la que llueve. Después de todo, que finalmente llueva o no depende de parámetros desconocidos. Así, a partir de un hecho perceptivo aceptado ('hay nubarrones') y un esquema derivado de experiencias acumuladas, se justifica otorgar credibilidad, en cierto grado, a la hipótesis de que 'esta tarde lloverá'. La credibilidad de la observación perceptiva y del esquema influye directamente en la credibilidad de la hipótesis.

Como se ha mencionado anteriormente, se argumenta a partir de enunciados que el oyente considera subjetivamente probables, verosímiles o creíbles. Estos enunciados, que rara vez son cuestionados por el oyente, se utilizan como premisas para sustentar otros enunciados en debate. La credibilidad de estos contenidos es la base fundamental de los entimemas, que combinan y modifican credibilidades a través de la estructura y conectividad del argumento, un proceso que se explicará en detalle más adelante. De este proceso se deriva un nuevo grado de credibilidad para la conclusión del argumento (la hipótesis en cuestión). Los entimemas utilizan creencias existentes para modificar la credibilidad de otras mediante razonamientos que son inherentemente rebatibles, derrotables o falibles, debido a la naturaleza falible de sus premisas.

1.4.2. Cultural – Comunicativo

Se ha reiterado que los entimemas son fundamentales en la argumentación retórica y dialéctica, técnicas de comunicación que buscan influir en las decisiones y acciones cambiando credibilidades. Como hemos visto, esta comunicación requiere un terreno común de conocimientos, creencias y principios compartidos, esencial para el intercambio intersubjetivo efectivo (Clark, 1985). La efectividad de la dialéctica y la retórica, como disciplinas prácticas, depende de la selección adecuada de premisas. Un entimema es funcional cuando utiliza premisas consideradas probables por un individuo. No obstante, para impactar efectivamente en los oyentes, estas probabilidades deben ser *argumentativamente plausibles*.

Lo argumentativamente plausible se define como los contenidos y reglas de relación ampliamente aceptados por la mayoría o los expertos más prestigiosos de una comunidad epistémica, lo que Vega (1993, 1998) identifica como éndoxa. Esta base común de creencias

proporciona una escala decreciente de plausibilidad²⁴, y es de ahí de donde deben extraerse las premisas para maximizar la efectividad argumentativa, una directriz del modelo de contexto del orador. Según Amossy (2002), el *éndoxa* actúa como un reflejo de la cultura específica, definido como 'lo que parece razonable para todos según la noción particular de racionalidad que el orador desarrolla dentro de un contexto social y cultural' (p. 476). Como señala Wodak (2016), esta idea vincula estrechamente el *éndoxa* (contenidos / opiniones aceptadas por un grupo) con el concepto de comunidad epistémica (miembros de un grupo definidos según un criterio de conocimiento) de Van Dijk (2007, 2008).

Boukala (2016) describe el *éndoxa* como la opinión tradicionalmente aceptada que representa el 'conocimiento' común de un grupo. Es importante destacar que estas opiniones, aunque ampliamente aceptadas, no son universalmente válidas ni se consideran verdades absolutas (Braet, 2005; Kienpointer, 1997). Son contenidos y reglas de relación que disfrutan de un alto grado de aceptación social (Van Eemeren, 2010), pero sólo de forma provisional, siendo susceptibles de ser desafiados después (Walton y Macagno, 2005). Como se puede observar, los sistemas culturales de creencias funcionan como *éndoxa* en la argumentación, denominándose sistemas *endoxásticos* en este contexto. Estos consisten en un conjunto de conocimientos y reglas generalmente aceptados, aunque falibles, que facilitan la intersubjetividad en la comunicación y sustentan inferencias coherentes en el discurso. Específicamente, en la argumentación, sirven como fuente de contenidos con credibilidad o autoridad para justificar hipótesis. Para un individuo que los ha interiorizado, estos sistemas ofrecen un criterio de verificación que se integra con su experiencia personal, formando una base subjetiva para interpretar y evaluar nueva información.

1.5. El guardián epistémico

Esta sección clarifica la relación de todos los conceptos que hemos ido viendo a lo largo de estas páginas precedentes con los procesos psicológicos de validación de la información. A lo largo de la exposición, veremos que la consistencia y la coherencia juegan un rol muy importante. Detallaremos sus especificaciones más adelante. La sección concluye introduciendo algunos conceptos teóricos procedentes de la propuesta de validación epistémica de Richter (2015; Schroeder et al., 2008), que fundamentan ciertos aspectos del modelo de credibilidad desarrollado en esta tesis.

Los sistemas de creencias conformados a partir de la experiencia y la comunicación se organizan en subsistemas temáticos y multidimensionales relacionados con los ámbitos físico y social, y con la autopercepción del individuo en esos ámbitos (Olsson, 2017; Pechey y Halligan, 2012). A medida que estos sistemas se van integrando unos con otros, unificándose en la memoria a largo plazo, irán adoptando progresivamente el rol de *guardián epistémico* (concepto tomado de Richter, 2011; y Schroeder et al., 2008). Este concepto se refiere a los mecanismos cognitivos que utilizan los individuos para evaluar y validar la información, asegurando la precisión y fiabilidad de la información que vaya a ser introducida en el sistema de creencias. Estos mecanismos filtran y

²⁴ Desde una opinión aceptada por toda la audiencia, pasando por la mayoría, o los expertos, y de estos, la mayoría, hasta llegar sólo a los más conocidos y reputados.

evalúan selectivamente la nueva información, contrastándola con el conocimiento existente, permitiendo a los individuos mantener un sistema de creencias coherente y justificado. Así, el sistema se protege de la información errónea, que tendría consecuencias potencialmente adversas para la toma de decisiones y la acción. Durante los procesos de validación de nueva información, esta información entrante activa los elementos relevantes del sistema de creencias almacenados en la memoria a largo plazo (Singer, 2013). La activación puede ser automática, mediante un proceso pasivo de resonancia, o estratégica y voluntaria. Esta información previa activada se constituye en el estándar de contraste para velar por la calidad y la veracidad de la información nueva, permitiendo o denegando su entrada. Según los estudiosos de los procesos de validación (e.g. Cook y O'Brien, 2014; Richter, 2003, 2011; Schroeder et al., 2008; Singer, 2006, 2013), el derecho de admisión está en función de la *consistencia* entre la nueva información y la información previa.

En el modelo que se presenta en esta tesis, se considera que la consistencia tiene un rol fundamental en general y en el proceso de validación en particular, aunque este rol quedará matizado en algunos aspectos fundamentales. En primer lugar, la evaluación de la consistencia se limita a la información relevante activada en la memoria de trabajo (Schroeder et al., 2008). Evaluar dicha consistencia en relación con todo el sistema de creencias convertiría la validación en un proceso computacionalmente intratable (e.g. Cherniak, 1984; Cook, 1971; Garey y Johnson, 1979; Johnson-Laird, 2012; Oaksford y Chater, 2007). Además, exigiría que las personas tuvieran capacidades de omnisciencia lógica y vincularía todos los elementos del sistema de creencias bajo relaciones de consecuencia lógica en un sistema cerrado (propiedad del cierre epistémico). Esto, a su vez, impediría la derrotabilidad y no-monotonidad propias del sistema de creencias, necesarias para su actualización y el aprendizaje de nueva información. En segundo lugar, este modelo considera que la consistencia es probabilística y graduada, no lógica y bivalente. Esto significa que uno puede creer en *α hasta cierto punto*, con una credibilidad finita que varía entre los extremos de imposibilidad y necesidad, representados por los números 0 y 1. Así, se admite la posibilidad de $-\alpha$ en la *proporción complementaria*. En tanto que complementarios, no entran en contradicción. Las creencias ya aceptadas hasta cierto punto se ponen en relación con la información candidata a entrar en el sistema en el contraste implicado en los procesos de validación. En ese momento se computará la consistencia de la nueva información con las creencias previas. El sistema de creencias tiende hacia la consistencia y la unidad, pero tiene limitaciones de capacidad y solo descarta contradicciones internas cuando logra detectarlas.

En todo caso, la consistencia es solo la condición mínima necesaria para la validación y la admisión de nueva información en el sistema de creencias. La detección de inconsistencias es suficiente para iniciar procesos explicativos de elaboración epistémica, con el objetivo de integrar información inconsistente (e.g. Johnson-Laird, 2012; Richter, 2011; Thagard, 2000) o revisar creencias previas (Alchourrón, Gärdenfors y Makinson, 1985; Elio y Pelletier, 1997), o para rechazar información inválida. El sistema intenta evitar la reducción al absurdo en su representación del entorno, pues tendría consecuencias poco adaptativas. La consistencia (ausencia de contradicción) es una condición que consideramos necesaria pero no suficiente para la *coherencia*, la cual a su vez

es indispensable para la *justificación*. Esta última es crucial para la *aceptación* de nueva información. Para que se acepte, la información nueva debe estar relacionada de manera *relevante* con los elementos previos del sistema de creencias, siguiendo *criterios tópicos* que los tomen como razones válidas para su aceptación. Así, la información no solo será consistente y coherente, sino también justificada. En la medida en que los elementos activados del sistema de creencias adopten el rol de ‘guardianes epistémicos’, proporcionarán una interpretación particular a la información entrante. En función de la consistencia, de la coherencia y de la justificación que se le otorgue a esta interpretación, se aceptará o rechazará la nueva información. Así, el sistema de creencias actúa no solo como un prisma interpretativo sino también como un filtro para la información entrante.

En estudios sobre la comprensión de textos, Schroeder et al. (2008) mostraron que, al aceptar información nueva ya interpretada, ésta pasa a conectarse no solo con el modelo de situación referencial del texto en curso, sino también con los conocimientos previos (i.e. con la red de creencias). Para el presente modelo, esto significa que una eventual información entrante, al interpretarse en función de criterios de conexión relevantes, pasará a formar parte integral del sistema subjetivo de creencias. Los autores también demostraron que, una vez integrada, esta información se convierte en un nuevo componente del filtro epistémico, utilizado para validar futura información entrante. Observaron que, al adquirir este rol de creencia, es decir, de información de referencia para el contraste epistémico, la información deja de estar disponible para volver a ser validada, salvo en casos de discrepancias significativas. Argumentan que una susceptibilidad excesiva a pequeñas inconsistencias comprometería la coherencia y estabilidad del sistema, impidiendo la toma de decisiones efectiva para la acción situada. Por tanto, en situaciones de conflicto entre la nueva información y la ya aceptada, se favorece la ya aceptada. El sistema busca maximizar su estabilidad. Si la nueva información es consistente, coherente y justificable con las creencias previas, se validará de manera rápida y sin esfuerzo.

Además, Schroeder et al. (2008) comprobaron que la plausibilidad, o credibilidad, de la nueva información aumenta y se vuelve más resistente a contradicciones futuras cuando se integra en el sistema a través de relaciones de *causalidad*. Esto respalda la idea de que, al vincularse con contenidos previos del sistema mediante una relación tópica, como la causalidad, se produce la transmisión de justificación. De hecho, esta conexión hace que los contenidos previos, por ejemplo, q y r , actúen como razones que justifican los nuevos, por ejemplo, p , aumentando la probabilidad de p dada q y r , (i.e., $P(p) < P(p|q, r)$) tal como se explicó en la sec. 3, cap. 3, pt. I. Según estos autores, el sistema actualiza su contenido con la mínima modificación necesaria. Aunque monitoriza la validez de la información entrante para aproximarse a la verdad, una vez validada, la información pasa a ser protegida para preservar la estabilidad del sistema.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

CAPÍTULO 2

Componentes del sistema de creencias

El sistema de creencias se compone de las creencias y sus relaciones, que se originan por mecanismos como los procesos inferenciales (razonamiento y asociación), perceptivos y aquellos relacionados con la cognición social. Estos mecanismos se derivan de las tres fuentes de formación de creencias previamente mencionadas, que se postulan como las fronteras entre el sistema y su entorno. Toman información del entorno como input para conformar creencias en el sistema como output.

Como se verá más adelante, las creencias que dependen directa y exclusivamente de estos tres dispositivos fronterizos de formación de creencias recibirán el nombre de *creencias básicas* o *fundacionales*. Habrá tres clases de creencias básicas en función de sus tres fuentes: las teoréticas o intuitivas, las empíricas o perceptivas y las discursivas o de testimonio. De ellas dependerá el resto de las creencias del sistema. Las creencias que dependen de otras creencias para su justificación reciben el nombre de *creencias derivadas*. A partir de aquí, luego, se podrá pasar a hablar de las creencias llamadas *implícitas* y *explícitas*, o de las creencias *disposicionales* y *ocurrentes* presentadas la sec. 2, cap. 3, pt. I y en el ap. C. En otro nivel de análisis, se podría hablar también de la distinción entre las *creencias situacionales*, específicas y episódicas y las *creencias abstractas*, genéricas y socialmente compartidas. Evidentemente, esto, a su vez, estará relacionado con sus lugares de residencia: la memoria episódica, la semántica, e incluso, la memoria de trabajo en el momento en que se opera con ellas. Por otro lado, en tanto que las creencias son acerca de algo, el sistema de creencias también albergará representaciones como parte de las creencias. En particular se defenderá que las creencias requieren de representaciones complejas o compuestas. Tal tipo de representaciones serán, principalmente, los modelos y los esquemas. Y, nuevamente, en tanto que estas representaciones complejas están, a su vez, compuestas por diferentes clases de categorías, y por conceptos, relaciones y propiedades, todos estos elementos también estarán presentes en los

sistemas de creencias. Es más, el sistema de creencias hará uso de ciertas propiedades de las representaciones para funcionar.

Finalmente, como se ha dicho al principio del párrafo anterior, el sistema de creencias también está compuesto por relaciones entre creencias. Básicamente, habrá dos tipos generales de relaciones: las de *justificación* y las de *refutación*. Las relaciones de justificación se dan desde una creencia previamente justificada en cierto grado hacia una representación que requiere de justificación para ser admitida en el sistema. Las relaciones de refutación pueden incidir sobre la creencia previamente justificada, sobre la representación candidata entrante, o sobre el propio nexo o relación de justificación, limitando su potencia. La posibilidad de establecer *relaciones tópicas* determina si una creencia previa puede o no funcionar como una buena razón para la justificación de una representación entrante. En el capítulo anterior ya vimos que hay una gran diversidad de relaciones tópicas que permiten la transición de justificación entre representaciones: de causalidad, temporalidad, finalidad, contrariedad, semejanza, etc.

Las relaciones tópicas de justificación son responsables de la *adhesión* que se produzca entre la representación entrante y el sistema de creencias previo. Asimismo, también lo serán para la adhesión que se produzca entre los componentes internos a la propia representación. Por otro lado, igual que las creencias se realizan mediante representaciones complejas, y éstas, a su vez, implican sus componentes, con las relaciones pasa algo similar. Las más relevantes para el sistema son las de justificación y refutación. Pero éstas, a su vez, implican a las relaciones de coherencia puesto que las relaciones tópicas son un subconjunto de relaciones de coherencia (tal como se desarrolla en el ap. C). No en vano, en la bibliografía sobre el procesamiento del discurso, hay cierto acuerdo en que los marcadores de las relaciones de coherencia caen en la categoría de ‘operadores lingüísticos argumentativos’; es decir, son conectores con una función argumentativa en el discurso (e.g. Kamalski, Lentz, Sanders & Zwaan, 2008; Snoek Henkemans, 2001; Verhagen, 2005). Además, en la misma medida en que las representaciones en general albergan distintos tipos de relaciones semánticas, asociativas, de frecuencias de coocurrencia, etc., así también las creencias. Por otra parte, en otro nivel de análisis, se podría decir que las relaciones de justificación ejercen su función de adhesión por relación tópica mediante los mecanismos de inferencia abductiva, inductiva y deductiva (en un sentido probabilístico). Crucialmente, estos mecanismos de inferencia son de *naturaleza explicativa*. Más adelante dedicaré unas páginas a desarrollar esta idea.

En orden a adentrarnos en la descripción de tales componentes, lo primero que se deberá hacer aquí será tratar de dar con una definición operativa del concepto de creencia. Esta tarea no estará exenta de dificultades, pero a medida que se vaya realizando, se irán abordando progresivamente muchos de los componentes listados aquí. Así pues, tratar de dilucidar qué es una creencia conllevará dar una descripción general del sistema de creencias.

1. Análisis anatómico de las creencias

Partiendo de la definición clásica ofrecida por la filosofía del lenguaje y de la mente, se podría considerar que las creencias son un tipo de actitud proposicional, entendida como una cierta forma de posicionamiento mental que guarda un organismo respecto del contenido de una proposición. Desde esta perspectiva, se considera que la *proposición* designa el contenido objeto de creencia, y que la *actitud* es ese posicionamiento mental “de tipo-creencia” (e.g., Fodor, 1975). Y ese posicionamiento de tipo creencia ha sido concebido por muchos autores como una *disposición a la conducta* en función de *considerar* que el contenido de la proposición es realmente el caso (tal como ya vimos en las sec. 1.1. y 1.2., cap. 3, pt. I al hablar del representacionalismo y del disposicionalismo). Así pues, se ha hablado bastante de la proposición, y de la actitud, pero del *organismo*, simplemente, no se dice nada.

Esta definición del concepto de creencia en términos de actitud proposicional es una primera aproximación que permite divisar que se trata de un concepto que está compuesto de partes: de un organismo (O), de la actitud (A) y de una proposición (P), introducida generalmente por una cláusula de relativo, derivando en el esquema de O – A – *que* P.

Sin embargo, pese a su valiosa contribución, esta definición entraña ciertas dificultades en varios sentidos. En primer lugar, no parece necesario asumir que el formato del contenido de la creencia deba ser proposicional. Dependiendo de qué se entienda por ‘proposición’ podría hacer difícil concebir la idea de diferencias de interpretación entre individuos, esto es, de interpretaciones subjetivas. La proposición no parece dar cabida a la ambigüedad existente en gran parte de la información que el sistema procesa. No parece dar cabida a los agregados de información procedentes de la memoria a largo plazo de los individuos. Menos aún a los agregados multimodales que muchas veces se añaden a la información lingüística entrante. En conjunto, esos agregados contribuyen a la obtención de una interpretación subjetiva de la información, y la credibilidad que se le atribuya a la información depende de tal interpretación. Otra dificultad estriba en la posibilidad de que definir la creencia como una actitud “que es de tipo-creencia” sea circular. Parece incluir el *definiendum* en el *definiens*, lo cual es poco deseable para dar una definición, y tampoco lo es para una explicación. Aun especificando que esa propiedad de ser “de tipo-creencia” consista en “una disposición a la conducta en función de considerar que el contenido de la proposición es, de hecho, el caso”, continúa siendo circular, puesto que “considerar” es equivalente a “creer”. Así pues, esa actitud deberá definirse de otra manera. Por último, el silencio acerca del organismo poseedor de la creencia es problemático, puesto que no se especifican las condiciones que debe poseer tal organismo para poder hallarse en ese estado mental de creencia. En conjunto, esta definición podría no acabar de ajustarse a los requisitos de una definición funcional que facilite la investigación empírica acerca de los procesos que operan sobre las creencias de los individuos en su actualización y cambio para la adaptación al medio.

En todo caso, puede ser de utilidad como punto de partida hacia otra definición. La definición hacia la que nos dirigimos es la de entender la creencia como una representación mental compleja

o compuesta, albergada en la memoria de un modo actual o posible, donde la *relación entre sus componentes* esté justificada, ya sea:

- a) Por un principio autorreflexivo, intuitivo o analítico constitutivo de los mecanismos inferenciales del sistema cognitivo, conformando cierto tipo de creencia fundacional o básica.
- b) Por la experiencia empírica de una sensación procedente del mundo que deriva, de forma “*bottom up*”, en creencias perceptivas fundacionales o básicas.
- c) Por la experiencia social con un/a comunicador/a al que se le atribuye honestidad y experticia, haciendo uso de cierta creencia social básica o fundacional.
- d) Por una creencia derivada de la combinación de cualesquiera creencias básicas que funcionan como término medio uniendo coherentemente los componentes de la representación compleja.
- e) O por cualquier creencia que ya se encuentre dentro del sistema de creencias que, también, funcione como término medio en la unión coherente de los componentes de la representación en cuestión.

Como puede observarse, los tres primeros criterios conformarán creencias básicas o fundacionales (introducidas en la sec. 3.1., cap. 3, pt. I), mientras que sólo los dos últimos conformarán creencias derivadas a partir de otras creencias de manera inferencial.

Pero entre esta definición y la primera, referente a la actitud proposicional, hay un paso intermedio. Consiste en transformar el esquema $O - A - \text{que } P$ de la actitud proposicional, a otro aparentemente similar de $S - A - R$, donde S se refiere a la *subjetividad* del organismo, A se refiere a la naturaleza de *adhesión* de las relaciones de credibilidad y R se refiere a una *representación* mental. De este modo, se define la creencia (momentáneamente, en este paso intermedio) como la relación de adhesión que se establece entre la subjetividad de un individuo y una representación. Esta definición presenta la creencia como un compuesto de tres elementos, cuyo núcleo central consiste en una relación. Esta relación es la *credibilidad* y se materializa en forma de adhesión. La fijación de la credibilidad de una nueva representación tiene lugar al vincularla con un elemento justificativo de la subjetividad mediante una relación tópica. La transmisión de justificación produce entonces la adhesión de esa representación al elemento justificativo. La intensidad de esa adhesión varía en grado y se mide como una probabilidad subjetiva. En ese sentido, se comporta como lo hace la probabilidad, respetando los axiomas de Kolmogorov. En ciertas condiciones, la interacción de esa intensidad con las demandas ambientales da lugar a la acción situada.

A partir de aquí, se procederá a tratar de describir estos tres componentes generales de la creencia. Este paso, en todo caso, conlleva ciertas dificultades. Esto es así, por un lado, porque la representación objeto de creencia lo es cuando pasa a ser creída, es decir, cuando pasa a ser parte del sistema de creencias del sujeto; y el sistema de creencias es, precisamente, una parte de la subjetividad. Por otro lado, porque la adhesión, entendida como el proceso de asignación de credibilidades en virtud de la justificación, también es un proceso subjetivo; es decir, se produce

dentro de la subjetividad de tal modo que la magnitud de credibilidad otorgada a una representación particular depende del estado puntual de configuración del sistema de creencias y de las leyes de funcionamiento de la propia subjetividad. Así pues, en las próximas páginas procederé primero exponiendo los aspectos de la subjetividad que son relevantes para el modelo, tratando de mostrar cómo su forma de operar puede dar lugar a unas primerísimas creencias primitivas llamadas básicas o fundacionales. Después, pasaré a exponer un análisis de las condiciones mínimas que debe cumplir la representación para poder ser una creencia. Finalmente, pasaré a describir el proceso de adhesión y la dinámica de la asignación de credibilidades.

1.1. La subjetividad

En cierto modo, se podría decir que la subjetividad es la condición de posibilidad del sistema de creencias y proporciona las herramientas iniciales para su construcción. Estas herramientas consistirán en ciertos aspectos del *self* del sujeto. Por un lado, consistirán en las *capacidades autorreflexivas* del individuo. Por otro lado, y teniendo en cuenta las tres fuentes de formación de creencias, esas herramientas consistirán también en los *mecanismos con los que opera el sistema cognitivo*: de inferencia, perceptivos y pragmáticos. Finalmente, también consistirá en el *espacio de la memoria declarativa* (episódica y semántica) que albergará a todo el sistema. En este sentido, además de proporcionar las bases para la construcción del sistema, la subjetividad también estará constituida por el propio sistema de creencias albergado en la memoria declarativa, por sus *contenidos* y por su *organización*.

Así pues, decir que una creencia es la relación de *adhesión* que se establece entre una *representación* en cuestión y la *subjetividad* del individuo significa que dicha representación se adherirá a:

- 1) Las propias capacidades de autorreflexión (como la representación de que “pienso”).
- 2) La propia forma de funcionar de los mecanismos inferenciales (como la representación de que $\alpha = \alpha$, o de que α no puede ser igual a $\neg \alpha$ a la vez y bajo el mismo aspecto).
- 3) La forma de funcionar de los mecanismos de asociación (como el paso de la percepción reiterada del par α y luego β al supuesto de que $\alpha \rightarrow \beta$).
- 4) La forma de funcionar de los mecanismos perceptivos (como la representación de que este objeto es rojo porque lo experimento sensitivamente como rojo).
- 5) La forma de funcionar de los mecanismos pragmáticos (como, para un niño, la representación de sus padres como comunicadores honestos y expertos).
- 6) Una organización particular en un momento dado de los contenidos de la memoria declarativa (como cualquier nueva representación r que se derive de la conexión entre los contenidos previos $c \rightarrow r$ y c activados en la memoria en un momento concreto).

Así pues, la subjetividad del organismo relevante para esta propuesta no consistirá únicamente en los contenidos previos de su memoria declarativa, su número y su organización, como venía

asumiendo desde el inicio de esta Parte II. Es cierto que estos contenidos y su organización determinan la interpretación subjetiva de la nueva información entrante y que, por tanto, juegan un papel crucial en la subjetividad. Asimismo, en tanto que son contenidos aceptados por el individuo, conforman las creencias del sistema de creencias. Además, en muchas ocasiones, constituirán las creencias previas que dotarán de justificación a las nuevas representaciones permitiendo su adhesión y nueva incorporación en el sistema. Pero como puede observarse, si los aspectos relevantes de la subjetividad se redujeran sólo al sistema de creencias, entonces volveríamos a caer en la circularidad, pues una representación sólo pasará a ser una creencia si se adhiere a un sistema de creencias en curso. La subjetividad puede incluirlo, pero debe admitir algo más, esto es, otros elementos externos al sistema de creencias, pero causalmente conectados con él.

Estos otros elementos incluidos en la subjetividad serán los que permitan la generación del sistema de creencias. Deberá haber representaciones que puedan adherirse a algún aspecto de la subjetividad que no sea, a su vez, una creencia. Esa adhesión debe producirse de manera no-inferencial. Esto significa que debe haber creencias que no obtengan su justificación a partir de otras creencias previas mediante inferencia, sino que tomen su justificación de algo externo al propio sistema de creencias. Esto se hace aún más evidente si tenemos en cuenta que las creencias, finalmente, deben ajustarse a cómo es el mundo reflejándolo de una manera más o menos fiel.

Tales creencias son las denominadas básicas o fundacionales: creencias que no se derivan de otras anteriores, pero que sirven de base sobre la que fundar el resto de las creencias que contenga el sistema. En este sentido, el sistema de creencias que aquí se propone se acerca a los postulados del fundacionalismo. En concreto, del fundacionalismo moderado (ver sec. 3.1., cap. 3, pt. I). Esto implica, por un lado, que no hay un límite preestablecido acerca del número de creencias básicas que puede haber y, por otro lado, que tales creencias básicas, pese a poder tener inicialmente una fuerte presunción de verdad (i.e., una alta credibilidad), son solo *prima facie*, pudiendo ser corregidas, revisadas o actualizadas por la adición de nueva información. Las creencias básicas aquí son derrotables, rebatibles o falibles.

1.1.1. Creencias básicas de un sistema fundacionalista moderado

Es cierto que una creencia puede estar justificada sobre la base de otra anterior, derivándose de esta por inferencia. Y ésta, a su vez, de otra anterior. De hecho, esto es un aspecto central en el presente modelo. Pero como ya se argumentó extensamente en la sec. 3, cap. 3, pt. I, es necesario que esta regresión se detenga en algún momento. Por otra parte, una justificación circular, donde todas las creencias se justifican entre sí de tal modo que, al final, cada creencia se justifica a sí misma, es igualmente inaceptable. Para evitar la regresión infinita de razones y la objeción del aislamiento, debe haber una justificación inicial y ésta no puede surgir del interior del propio sistema de creencias. Por tanto, debe haber una serie de creencias básicas que no se adhieran a sistema alguno de creencias, sino que sirvan de base para su fundación. La cuestión es, entonces, saber a qué se adhieren esas creencias fundacionales. La respuesta bien podría ser la de que se

adhieren a los propios mecanismos de funcionamiento del sistema cognitivo. La credibilidad inicial consistirá en una sensación de seguridad y confianza en dichos mecanismos.

Más arriba se desarrolló la idea de que hay tres fuentes básicas de formación de creencias: la inferencia / razonamiento, la experiencia / percepción, y el discurso / comunicación. Es pertinente preguntarse hasta qué punto estas fuentes de formación pueden generar creencias básicas que no dependan de otras anteriores para su justificación.

1.1.1.1. Creencias teoréticas básicas o intuiciones

Los mecanismos inferenciales no se limitan a combinar creencias previas para derivar creencias nuevas. También pueden generar creencias fundacionales, de naturaleza axiomática e intuitiva (Poston, 2016). Estas incluyen verdades formales sintácticas como las tautologías, el principio de identidad, o el principio de no contradicción o algunos principios lógicos y matemáticos. También incluyen verdades de razón o analíticas, como que la parte es inferior al todo, que toda causa tiene un efecto, o que el alimento es comestible.

Además, se suele considerar que otras podrían ser de carácter autoreflexivo, relativas al *self* (e.g. Fumerton, 2006; Pryor, 2000), como el famoso ‘pienso, luego existo’, donde ‘pienso’ es una creencia básica que no requiere de justificación adicional de otras creencias y ‘existo’ es una creencia derivada de la anterior. Lo mismo se podría decir de ‘siento’ y ‘percibo’. Estas creencias son básicas en tanto que su justificación no procede de la inferencia a partir de otras creencias anteriores. Las creencias básicas procedentes de esta fuente son las que tienen una carga de credibilidad mayor (casi siempre, total). Son creencias invariables que no dependen del mundo. Sin embargo, su credibilidad consiste en una sensación de seguridad y convicción en la forma de operar del sistema y en la consciencia. Y como apuntan Quine y Ullian (1970), desde Kant a Gödel sabemos que ni siquiera las verdades matemáticas son autoevidentes por sí mismas, lo que implica que la intuición no garantiza una adhesión plena a estas creencias.

1.1.1.2. Creencias Empíricas Básicas o Percepciones

Las creencias empíricas básicas o percepciones son fundamentales para vincular directamente la mente con el mundo real, evitando así la crítica de aislamiento cognitivo. Estas creencias se originan a partir de la experiencia sensorial y la percepción, formando una base esencial para la correspondencia entre las representaciones mentales y la realidad externa. Esta conexión es crucial para sustentar la intencionalidad, permitiendo que nuestras percepciones y experiencias directas formen creencias que reflejen fielmente el mundo exterior, proporcionando una base sólida y fiable para el conocimiento empírico.

Se ha propuesto que las creencias empíricas básicas son aquellas que versan sobre las apariencias (e.g. Huemer, 2001, 2013; Pryor, 2000; Tucker, 2010). Las creencias perceptivas sobre las apariencias estarían justificadas de forma no inferencial por las *sensaciones* puras, estimulares, inmediatas, preconceptuales, “subdoxásticas”. Tales sensaciones no requieren de capacidades conceptuales ni presuponen ninguna categoría: ‘creo que esta flor es roja porque me parece roja. Es decir, porque *experimento* algo a lo que llamaré *rojez* al ver eso a lo que llamaré *flor* cuando tenga

conceptos y palabras que categoricen y denominen los elementos de esa experiencia'. Esa experimentación no es una creencia, sino una sensación, una estimulación particular del sistema nervioso. Esa estimulación desencadena los procesos responsables de las representaciones perceptivas. Así, un estado de no-creencia produce y, a su vez, justifica de manera no inferencial una creencia perceptiva²⁵. Las percepciones no requieren de argumentos que justifiquen su resultado²⁶. Se trata, nuevamente, de una adhesión al modo de operar del sistema cognitivo, una confianza en su funcionamiento (siempre y cuando se esté en condiciones óptimas de observación). No se es consciente necesariamente de esa confianza. Esas creencias sobre las apariencias conforman la base de los episodios singulares de nuestra experiencia subjetiva.

De este modo, las creencias empíricas dependen ontológicamente de la estimulación sensorial procedente del mundo real, contribuyendo así a la dirección de ajuste mente-mundo. Estas creencias empíricas se producen de forma involuntaria y obligatoria en un "proceso guiado por los datos" o "*bottom-up*". En principio, en tanto que la arquitectura cognitiva es la misma para todos los individuos de una misma especie (Demetriou et al., 2005), los mecanismos que producen las creencias perceptivas son también los mismos. En igualdad de condiciones ambientales, esto garantizaría que todos los individuos de una misma especie pudieran tener las mismas creencias empíricas básicas. En este sentido, seguirían siendo subjetivas, pues toda experiencia lo es, pero compartibles con los otros miembros de la especie²⁷.

En cualquier caso, esa igualdad de condiciones ambientales no siempre se logra, pudiendo variar las condiciones de observación de un individuo a otro, o de un momento a otro. Además, puede haber diferencias individuales de varios tipos. Por ejemplo, en el ámbito de la percepción del color, aun en igualdad de condiciones ambientales, una persona podría ver la pizarra de un aula de color verde oscuro, y otra persona de color gris, o negro. Así, puede haber diferencias en función de la parcela del entorno del objeto en la que se fijan las personas para juzgar el color, afectado por el

²⁵ En tanto que es una justificación sin razones, Burge (2003) ha preferido llamarla "autorización" o "legitimización". De este modo, la sensación *autoriza* o *legitima* una creencia perceptiva (para una revisión crítica véase Kalpokas, 2014).

²⁶ Por lo menos, no en el sentido tradicional de justificación y argumentación. Se puede entender, al modo de Helmholtz (1866), que la percepción es un proceso inferencial (Fodor y Pylyshyn, 1981). Las propiedades del estímulo detectadas por el transductor (i.e., conversor de energía estimular) son epistémicamente inaccesibles, pero el sistema cognitivo extrae de ellas una información que funcionará a modo de premisas para construir la creencia perceptiva. Así, según estos autores, la *creencia perceptiva* es la conclusión de una especie de "argumento" que se construye de manera inferencial, pero las premisas de las que parte *no* son inferidas. En este sentido, independientemente de que esas premisas sean creencias o no, dado que no son inferidas, sigue siendo compatible con el fundacionalismo (moderado) sobre las creencias empíricas que estamos presentando aquí.

²⁷ No se puede decir que sean propiamente *objetivas* pese a que fueran compartidas por todos los miembros de la especie. Lo objetivo depende del objeto, no de la especie que lo percibe. Por ejemplo, el sistema visual de las abejas es sensible a luz ultravioleta. Esto les permite ver el polen con colores muy vivos, destacando entre el resto. Los humanos, sin embargo, no son sensibles a la luz ultravioleta, por lo que su percepción visual del mundo será muy distinta a la de las abejas pese a observar el mismo objeto. *El sistema perceptivo determina los aspectos de la realidad que podrán ser percibidos y el modo en que serán percibidos*. El sistema perceptivo muestra la parcela de realidad que las distintas especies de organismos necesitan para su supervivencia.

contraste contextual (Jonaskaite, Dael, Parraga, Chèvre, Sánchez, y Mohr, 2020); o podrían estar categorizando la misma sensación de distinta manera; o puede haber cierta variabilidad genética en la fisiología responsable de la percepción del color (Deeb, 2006); o incluso, pese a ser un proceso *bottom-up*, algunos autores (e.g. Clifford, Holmes, Davies, y Franklin, 2010; Lupyan, 2017; Maier y Rahman, 2019) sugieren que podría verse influido por procesos y estados “guiados cognitivamente” o “*top-down*” como las otras creencias y expectativas previas o el conocimiento categorial del sujeto durante el curso de la construcción de la creencia perceptiva (aunque hay mucho debate sobre esto; e.g. Firestone y Scholl, 2014, 2015, 2016).

Así pues, hay toda una serie de posibilidades que podrían generar ligeras diferencias intra o interindividuales en la conformación de esas creencias perceptivas o empíricas básicas. Y es que, como se mencionó en la sec. 3.1.2., cap. 3, pt. I, la relación entre la realidad y su transducción (sensorial y funcional), su codificación y su representación no es obvia, ni todos los mecanismos implicados son infalibles. Por tanto, las creencias perceptivas pueden conllevar cierta carga de *incertidumbre* en la señal informativa. En este sentido, se dice que las creencias empíricas o perceptivas son creencias justificadas “*prima facie*” (Huemer, 2001, 2013): si al sujeto *S* le *parece* que *P*, entonces, en ausencia de refutadores, *S* tiene algún *grado* de justificación para *creer* que *P*. Por tanto, son creencias cuya justificación no depende de otras, y uno tiende a aceptar por defecto.

En este sentido, los estudios de neuroimagen de Harris, Sheth y Cohen (2008) y de Sachs y Hirsch (2008) han mostrado que los seres humanos aceptan la apariencia como realidad hasta que se demuestre lo contrario. En la línea del presente argumento, también mostraron que, pese a que inicialmente dichas creencias puedan tener una fuerte presunción de verdad, pueden ser actualizadas por la confirmación o refutación de nueva evidencia. Esto es, se revisan y corrigen a la luz de información adicional. Puedo ver un objeto a lo lejos y creer que es, por ejemplo, una vaca, porque me parece una vaca, pero tener ciertas reservas. Luego, al acercarme, veo que era una roca. Son creencias básicas falibles, rebatibles o derrotables. Esto implica que la adhesión que se produzca puede llegar a ser variable, pudiendo pasar a darse en distintos *grados*.

1.1.1.3. Creencias comunicativas básicas

En tercer lugar, finalmente queda preguntarse por el potencial de la fuente social, comunicativa o discursiva para la generación de creencias básicas fundacionales. Pese al amplio desarrollo de la epistemología social y del testimonio en las últimas décadas, la cuestión acerca de las creencias básicas en este ámbito es bastante nueva y está relativamente poco desarrollada. En la medida en que se trata de una fuente social, posiblemente intervengan aspectos pragmáticos y emocionales. Aquí podría ser que las supuestas creencias básicas relativas a esta fuente se adhieran al modo en que operan los mecanismos de la cognición social, que a su vez están relacionados con el sistema emocional de recompensa (Hiser, y Koenigs, 2018; Olsson, y Ochsner, 2008). Existen aquí dos posiciones diferenciadas.

1. Una asume que las creencias básicas se corresponden con el propio testimonio, cuya credibilidad encuentra su justificación en el hecho mismo de la existencia de la comunicación: nos creemos algo porque nos lo han dicho, sin necesidad una inferencia ulterior.
2. La otra asume que la creencia básica consiste en que las emisiones de comunicadores honestos y expertos son aceptables, y se justifica por episodios concretos de experiencia social. Así, del mismo modo que las creencias empíricas básicas quedaban justificadas de manera no inferencial por la experiencia perceptiva, esta creencia comunicativa básica queda justificada de manera no inferencial por la experiencia social. Pero, mientras que los contenidos de las creencias empíricas podían ser extremadamente variados, en este caso sólo hay una creencia básica. Ahora bien, esta creencia básica es la puerta de entrada de una infinidad de creencias derivadas de ésta. Todo lo que nos diga un comunicador considerado honesto y experto será creído.

A continuación, se desarrollarán brevemente estos dos puntos. Se podría decir con relativo acierto que el llamado *Principio de Aceptación* de Burge (1993) caracteriza el primero de ellos. Hasta donde sé, el segundo posicionamiento no ha recibido un nombre que lo identifique. En tanto que no me atrevo a enunciar un principio, me aventuraré a llamarle, simplemente, *posicionamiento de la honestidad y la experticia* (aunque procede, como veremos, de la satisfacción de necesidades y el logro de metas por vía social). No estoy seguro de que sean los nombres más afortunados, pero nos servirán para continuar con la exposición.

1.1.1.3.1. Posicionamiento del Principio de Aceptación

Como mencionamos al explorar cómo la comunicación contribuye a la formación de creencias, típicamente involucra a un hablante que busca influir en su interlocutor y a un oyente que desea disminuir su incertidumbre. Sperber (2013) sugiere que la comunicación solo se habría seleccionado naturalmente en la evolución como un mecanismo de intercambio de información honesta. Sin esta base de honestidad, perdería su valor adaptativo esencial para la supervivencia del individuo y la especie. Por lo tanto, la transmisión veraz del conocimiento ha sido, y sigue siendo, fundamental para la existencia misma de la comunicación.

Esto sugiere que las personas tienden a ser honestas en sus declaraciones, ya que mentir requiere un esfuerzo adicional y consume recursos cognitivos, lo que hace que la sinceridad sea la opción más eficiente (Van Bockstaele et al., 2012; Vrij, Fisher, Mann, y Leal, 2006; Vrij, Granhag, Mann, y Leal, 2011). En otras palabras, ser sincero resulta más sencillo y económico que engañar. Esta tendencia natural hacia la honestidad respalda el principio conversacional de presunción de sinceridad de Grice (1989), que asume que, por defecto, las personas dicen la verdad, y solo se desvían de ella cuando el beneficio justifica el esfuerzo. Como Van Dijk (2014) destaca, fuera de entornos jurídicos, en la vida cotidiana se considera que el testimonio ajeno es una fuente de conocimiento válida y fiable. Esto es aplicable tanto en contextos formales de educación, donde se transmiten los conocimientos acumulados por la humanidad, como en situaciones diarias como preguntar la hora, preguntar por una calle o escuchar relatos de otros.

En el campo de la epistemología social, distintos autores como Burge (1993), Foley (1994), Hardwig (1985, 1991) y Lackey (2008) sostienen que el testimonio debe considerarse como una fuente básica y *prima facie* de evidencia o justificación. Esto implica que las personas generalmente aceptan que lo que otros dicen es verdad, a menos que haya motivos convincentes para dudar. Esta postura se encapsula en el Principio de Aceptación de Burge: “[Una] persona está legitimada para aceptar como verdadero algo que se le presenta como verdadero y que le es inteligible, a menos que existan razones más fuertes para no hacerlo” (Burge 1993, p. 457). Conforme a las experiencias diarias mencionadas, generalmente no exigimos pruebas de la honestidad o experticia de quien habla para creer en sus palabras. No pedimos evidencia adicional sobre su fiabilidad. Esto es lo que otorga al testimonio una autoridad para conformar creencias fundacionales, no derivadas de otras creencias (Foley, 1994). Según esta perspectiva, la información recibida se acepta por defecto, quedando la posibilidad de cuestionarla y revisarla para momentos posteriores (Gilbert, 1991; Gilbert, Tatarodi, y Malone 1993).

Pero, por otro lado, el mismo recurso a la evolución como determinante de las capacidades y mecanismos de la comunicación humana también proporciona argumentos contrarios a este principio. Sin negar la importancia de la honestidad en el papel adaptativo de la comunicación, también hay que tener presente que el error de interpretación, la manipulación de la información y la mentira son un hecho. Un hecho que, en determinadas circunstancias puede jugar a favor del hablante en la competición por la supervivencia y en detrimento del oyente en el mismo juego. En este sentido, Sperber (2001) postula que debe haber algún tipo de precaución automática que someta la información transmitida a un procedimiento de verificación básico, con el fin de probar su validez.

1.1.1.3.2. Posicionamiento Básico de la Honestidad y la Experticia

La comunicación honesta y experta es crucial para la supervivencia y adaptación, especialmente en la transmisión de conocimiento de padres a hijos. Es importante entender por qué los niños confían en sus padres como fuentes epistémicas. Contrariamente a la creencia de que los niños son ingenuos, estudios recientes muestran que son selectivos al aceptar información de terceros (Call y Tomasello, 1999; Couillard y Woodward, 1999). Esto apoya la idea de la existencia de una *vigilancia epistémica* (Sperber, Clément, Heintz, Mascaro, Mercier, Origg, y Wilson, 2010) según la cual se produce una precaución automática de verificación de la información. La confianza de los niños en sus padres y cuidadores se basa en experiencias sociales positivas, donde el afecto, el cuidado y la coherencia entre los compromisos (promesas y amenazas) y su cumplimiento satisfacen sus necesidades y promueven la confianza (Hakim et al., 2012; Rotenberg et al., 2013). Esta confianza no solo es un rasgo humano, sino que se observa en otros animales desde el nacimiento, radicando en la satisfacción de necesidades y percibiendo al comunicador como alguien que no busca dañar, indicando su *honestidad* y buenas intenciones.

La mera amabilidad u honestidad no garantiza la confianza epistémica en las afirmaciones de alguien. Goldman (2001) argumenta que los niños también desarrollan confianza en la autoridad epistémica de sus padres o cuidadores a través de interacciones que demuestran la competencia de

estos últimos. Basándose en el concepto de Baron-Cohen (1995) sobre el "detector de la dirección de la mirada" y el "mecanismo de atención compartida", destaca que las primeras expresiones lingüísticas a las que se exponen los niños tratan sobre objetos o eventos conocidos por el hablante. De este modo, prosigue Goldman, *la experiencia temprana del niño es la de que los hablantes (adultos) hablan sobre lo que saben*, sirviendo de base empírica para considerar a los padres o cuidadores como *expertos* y competentes en sus conocimientos.

Los niños suelen confiar en la información que reciben de adultos, especialmente cuando esta ayuda a interactuar con su entorno de manera efectiva. Por este motivo, desde el punto de vista del desarrollo ontogenético, uno cree la información procedente de estas fuentes comunicativas, asumiendo los mensajes como relevantes para su *self* (Seitz, Paloutzian, y Angel, 2018). Aunque esta información inicial puede ser revisada con el tiempo (Sperber, 1996; Tomasello, 2008), los niños tienden a dar crédito a quienes muestran experticia y honestidad. Desde edades tempranas, muestran preferencia por informadores que demuestren competencia (Burdett et al., 2016; Clegg, Kurkul, y Corriveau, 2019; Clément, Koenig, y Harris, 2004) y amabilidad (Landrum, Mills, y Johnston, 2013; Mascaro y Sperber, 2009; Williams y Danovitch, 2019). Valoran más la amabilidad, lo que a veces les lleva a una percepción elevada de la experticia del comunicador (Landrum et al., 2013). La familiaridad con la fuente también afecta esta confianza (Williams y Danovitch, 2019), sugiriendo que los niños naturalmente ven a sus padres o cuidadores como fuentes fiables de información, formando una base para una creencia social básica.

La creencia básica en la comunicación, basada en las experiencias positivas tempranas de interacción, sugiere que los mensajes de un emisor honesto y experto son fiables. Esta creencia se considera básica porque no deriva de otras creencias ni se justifica inferencialmente. La evaluación hedónica positiva de estos comunicadores facilita y guía la aceptación de creencias subsiguientes, derivando en preferencias por determinados cursos de acción y en la adopción de determinadas actitudes. Este fenómeno subyace a algunas estrategias en publicidad, marketing y política, resaltando la relevancia del *ethos* retórico.

Sin embargo, es crucial distinguir entre la creencia general en la fiabilidad de comunicadores honestos y expertos y la aplicación específica de esta creencia a individuos concretos. Aunque un emisor considerado fiable pueda transmitir información errónea involuntaria o voluntariamente (por manipulación o engaño), esto no invalida la creencia básica. Solo indica que se ha aplicado mal para ese caso específico. La creencia básica de que "la información procedente de fuentes honestas y expertas es aceptable" sigue siendo válida pese a que la creencia derivada de que "Juan es honesto y experto" no lo sea. Los niños, atribuyen experticia y honestidad de manera inferencial, por un proceso inductivo, teniendo en cuenta los episodios pasados de acierto y exactitud de la información recibida de esa fuente (Clément, Koenig, y Harris, 2004; Birch, Vauthier, y Bloom, 2008; Sperber et al., 2010). Al basarse en frecuencias de experiencias pasadas, obtienen una comprensión graduada de la credibilidad y la competencia sobre las fuentes, en lugar de una percepción binaria.

1.1.1.3.3. *De la honestidad y la experticia de los cuidadores a la credibilidad de la mayoría*

Una parte sustancial de las creencias que uno acabará albergando vendrán introducidas por comunicadores particulares considerados honestos y expertos.

La confianza que los niños desarrollan hacia sus padres o cuidadores es fundamental para la construcción de confianza en terceros a lo largo de su vida (Ainsworth, Blehar, Waters, y Wall, 2015; Bowlby, 1969; Rotenberg, 1995, 2010). Con el tiempo, estas dinámicas de confianza se abstraen y aplican de manera selectiva a otras personas fuera del círculo de cuidado directo, como resultado de un proceso de aprendizaje (Harris y Corriveau, 2011; Rotter, 1971; Sperber et al., 2010). Esta expansión de la confianza favorece comportamientos prosociales como la ayuda y la cooperación, esenciales para la formación de comunidades y el fortalecimiento del sentido de pertenencia (Dirks, 1999; Gambetta, 1988; Misztal, 2013). Este carácter gregario de los seres humanos satisface necesidades de seguridad y podría explicar por qué las personas suelen confiar en la opinión mayoritaria (éndoxa) en situaciones de incertidumbre. Además, a medida que se desarrollan, las personas internalizan las normas culturales de su comunidad (Harkness y Super, 1996; Seitz et al., 2018; Sperber, 1996), configurando gran parte de sus sistemas de creencias. La gran mayoría de los contenidos de esos sistemas vienen dados por el discurso social más que por el razonamiento independiente o la experiencia directa.

En esta sección hemos tratado de detallar los aspectos de la subjetividad que consideramos relevantes para poder albergar creencias, soslayando la dificultad de la circularidad aparente que habíamos apuntado más arriba. Se mantiene la consideración de que una creencia consiste en la adhesión de la subjetividad de un individuo a una representación, así como que parte de esa subjetividad consiste en la red de creencias previas del sujeto. Pero los aspectos de la subjetividad relevantes para el modelo quedan ahora ampliados.

Partiendo de las tres fuentes de formación de creencias introducidas al inicio de este capítulo, la subjetividad del individuo se puede entender desde dos perspectivas. Primero, como el modo en el que opera su sistema cognitivo (i.e., en el razonamiento, en la percepción y en el establecimiento de interacciones sociales). Segundo, como los contenidos específicos de su memoria declarativa albergados en su sistema de creencias. De este modo, en el proceso de llegar a creer una nueva entrada informativa o representación, esta podrá adherirse tanto al modo de operar del sistema cognitivo (como creencia básica), como a una creencia básica, como a otras creencias derivadas que previamente ya formen parte del sistema de creencias.

Así pues, una creencia será una representación mental albergada en la memoria y justificada, ya sea por una sensación, una inferencia, una fuente experta y desinteresada, una combinación de cualquiera de estas o por otras creencias precedentes.

1.2. La representación

Este es el otro término de la relación de adhesión en el esquema $S - A - R$. En este punto es pertinente detallar qué condiciones debe cumplir la representación para ser susceptible de constituir el objeto de la credibilidad. Unas cuantas páginas atrás se anunció que debía tratarse de representaciones complejas o compuestas por distintas categorías. Otra condición necesaria para que la representación sea susceptible de ser creída es que sus categorías puedan relacionarse unas con otras de manera coherente en la mente del individuo. Y, finalmente, la condición suficiente es que esas relaciones de coherencia sean de tipo justificativo; es decir, que sean las relaciones tópicas que producen la adhesión. Si bien esta última condición es suficiente, no se puede dar sin que se den las dos anteriores.

1.2.1. Complejidad de la representación

Para exponer estas ideas, nuevamente, lo más sencillo es empezar recurriendo a las aportaciones de la filosofía y del estudio del lenguaje.

1.2.1.1. Las creencias y la estructura de las representaciones mentales

Desde estos campos es ampliamente reconocido que los términos lingüísticos aislados no pueden tener valor de verdad. Sustantivos como “ventana”, verbos como “romper” o adjetivos como “rojo” no son, sin más, verdaderos o falsos. Sólo cuando estos términos *componen* enunciados aseverativos o asertivos se hacen susceptibles de ser verdaderos o falsos. Por ejemplo, “Juan rompió la ventana” o “los pájaros surcan los cielos” son enunciados a los que, en determinadas circunstancias, se les puede atribuir un valor de verdad u otro. Y este tipo de enunciados están compuestos por un sujeto y un predicado que se unen en la afirmación y se separan en la negación. Consisten en la relación de dos términos, en *la atribución de un predicado a un sujeto* (ver ap. C).

En el ámbito de las representaciones mentales, los conceptos aislados tampoco admitirán grados de credibilidad. Las creencias versan sobre *relaciones* de elementos. Por tanto, sólo las representaciones compuestas de dos o más unidades pueden ser o no creídas.

Desde una perspectiva en la que se considere que las representaciones mentales consisten en símbolos intencionales estructurados en un formato proposicional, la unidad mínima susceptible de ser creída como verdadera o falsa será la propia proposición ya que está compuesta por predicados y argumentos. Por ejemplo, las oraciones anteriores se corresponderían con unas proposiciones que se anotarían como ROMPER (Juan, ventana) y SURCAR (pájaros, cielo), donde a lo que está fuera del paréntesis se le llama predicado, y a lo que está dentro se le llama argumentos. Estas proposiciones simples ya pueden ser portadoras de valores de verdad; y, a su vez, pueden enlazarse con otras mediante conectores lógicos para formar proposiciones más complejas (lo cual es la base de la productividad del sistema). En esos procesos de combinación se preservaría su valor de verdad.

Los modelos mentales de situación y los esquemas o marcos son formatos alternativos de representación mental. En lo que sigue, la exposición se centrará en los modelos de situación, pero,

como se comentará más adelante, gran parte de esta explicación también se puede aplicar a los esquemas o marcos.

1.2.1.2. *Fundamentos y niveles en la comprensión del discurso*

La propuesta de los modelos mentales de situación se desarrolló en el ámbito de la psicología del discurso (van Dijk y Kintsch, 1983) y del razonamiento (Johnson-Laird, 1983), aunque, actualmente, muchos autores consideran que es un formato general de representación mental. En la comprensión del discurso había tres niveles de representación del texto (van Dijk y Kintsch, 1983). El primero, la *forma superficial*, consistía en la representación ortográfica, fonológica, léxica y sintáctica literal de la forma lingüística del texto. El segundo, el *texto base*, consistía en la representación semántica de las relaciones proposicionales explícitas en el texto. El tercero, el *modelo de situación referencial*, consiste en una reconstrucción mental multimodal y no lingüística de la situación sobre la que trata el texto (Barsalou, 2008; Glenberg, 1999; Varela, Thompson y Rosch, 1991), que va más allá del contenido proposicional explícito en el texto, incorporando información fuertemente asociada con éste, procedente de la memoria episódica y semántica. Estos modelos son representaciones complejas consistentes en una red interconectada de unidades de información procedente del texto y de la memoria (e.g. Fletcher y Bloom, 1988; Kintsch, 1988, 1998; van den Broek, Yong, Tzeng y Linderholm, 1999). En tanto que interviene información de la memoria del individuo, la *interpretación subjetiva* de la información entrante se produce en este nivel. En este sentido, la interpretación de la oración “los pájaros surcan los cielos” será distinta en función de que uno se esté representando el prototipo de pájaro como algo parecido a una paloma o como algo parecido a un avestruz, con sus consecuentes repercusiones sobre la credibilidad.

Como se ha mencionado, muchos investigadores consideran que los modelos mentales de la situación son el tipo de representaciones mentales subjetivas que la gente construye acerca de su experiencia cotidiana concreta (e.g. Van Dijk, 2014; Zacks y Swallow, 2007). En el proceso de comprensión de la experiencia, la mente construye un modelo isomórfico, a pequeña escala, del fenómeno referente. Este modelo imita su funcionamiento, al menos, en el modo en que la mente puede captarlo. En este sentido, un modelo siempre es una interpretación de la realidad.

1.2.1.3. *Categorías generales en modelos mentales*

Estos modelos, constituidos por una red interconectada de información entrante y de la propia memoria, se estructuran en el marco de unas *categorías generales* mediante las cuales la mente organiza la realidad en representaciones coherentes. Hay consenso en que estas categorías generales y no lingüísticas fundamentan la *predicación* en el lenguaje y el discurso.

Las categorías generales y abstractas que *componen* los modelos mentales permiten organizar la estructura de las representaciones de la experiencia (directa o comunicada) en un esquema canónico de las situaciones. Estas categorías son, principalmente, las de escenario o marco (combinaciones de fragmentos de espacio y tiempo), participantes animados o inanimados (animales, personas, cosas), relaciones entre participantes (roles, tamaño relativo, posición, posesión, causación, etc.), estados, acciones o eventos dinámicos (definidos como cambios de

estado), agentes, objetivos y metas, géneros (clases, conjuntos) y números o cantidades (e.g. Radvansky y Zacks, 2011; Shipley y Zacks, 2008; Van Dijk, 1994, 2014).

1.2.2. Dinámica de actualización de los modelos mentales de situación

La forma en la que estas categorías se relacionan entre sí dota de una estructura simple y jerárquica a los modelos de situación (Mannes y Kintsch, 1987; Van Dijk, 1980). De este modo, los modelos pueden organizar el flujo dinámico de la experiencia cotidiana de manera rápida y eficiente (Van Dijk, 2014). Segmentan ese flujo en episodios discretos, significativos y relevantes delimitados por cambios de estado en las relaciones establecidas entre las categorías (Van Dijk, 2014; Zacks y Swallow, 2007). Esto es, las transiciones y cambios de estado se aplican sobre las relaciones de causalidad, temporalidad, acción, agencialidad, metas, localización espacial, etc., es decir, sobre las relaciones que permiten la *coherencia* en la representación (e.g. de Vega, 1995; Zwaan y Radvansky, 1998). Cuando se produce un cambio de estado en estas relaciones, es necesario que el modelo de situación se actualice en orden a mantener la coherencia. Consistentemente con esto, sucesivos experimentos en psicología del discurso han mostrado que los tiempos de lectura se incrementan cuando se producen rupturas en estas relaciones al procesar un texto (p. ej. en una analepsis o *flashback* entre dos hechos o escenas). Son, precisamente, las relaciones entre las variables de las categorías las que permiten o impiden la producción de representaciones coherentes en la mente del individuo (Zwaan y Radvansky, 1998).

Por tanto, estas categorías juegan un papel fundamental, en varios sentidos. En primer lugar, son lo que permite que los modelos puedan ser representaciones *compuestas*, lo cual era el primer requisito para que pudiera llegar a ser creída. En segundo lugar, son los componentes que la mente emplea para establecer *coherencia* en la representación. Y la coherencia era la otra condición necesaria para poder recibir credibilidad. Y, en tercer lugar, al asumir que las relaciones tópicas son un subconjunto de las de coherencia (a saber, el subconjunto que permite la transmisión de *justificación* entre representaciones) se obtiene ya la condición suficiente. En cualquier caso, en tanto que esa transmisión fluye desde las fuentes originales de credibilidad a través del entramado de representaciones, es más una propiedad del sistema que de las representaciones *per se*.

Pero, los méritos de las categorías no acaban aquí. También son las que reciben las transiciones y los cambios de estado permitiendo la actualización y la falibilidad (i.e., su *dinamismo*). Las relaciones de coherencia que se establecen entre los elementos de la representación y su actualización influyen sobre la interpretación (Albrecht y O'Brien, 1993; Van den Broek, y Helder, 2017). Y, a su vez, la interpretación influye sobre la credibilidad. Además de esto, las relaciones de coherencia posibilitan distintas formas de razonamiento sobre las representaciones, como el razonamiento relacional, causal, espacial, temporal, analógico y otros (e.g. Bloom, Fletcher, Van Den Broek, Reitz, y Shapiro, 1990; Fletcher, & Bloom, 1988; Gattis, 2003; Gentner, 2001; Gentner, y Holyoak, 1997; Gentner, Imai, y Boroditsky, 2002; Kubose, Holyoak, y Hummel, 2002; Trabasso, 2005; Trabasso, Van den Broek, y Suh, 1989). Esto permite, entre otras cosas, que unos modelos se combinen con otros conectando sus componentes (i.e., su *productividad*).

Finalmente, estas categorías, pese a no ser lingüísticas, fundamentan la *predicación*. En tanto que los modelos de las experiencias tienen las mismas propiedades generales que los modelos de situación para el discurso, estas categorías organizan también las representaciones semánticas de las cláusulas y de las oraciones que describen esas situaciones (Van Dijk, 2014). Esto es relevante para la credibilidad en tanto que *una representación traducida a un formato lingüístico será tanto más creíble cuanto más predicable sea el objeto o atributo respecto del sujeto oracional*. Es decir, más adherido esté en la mente del individuo aquello que se predica de quien se predica. Será predicable en función de los conocimientos y creencias previas del individuo. Así, según ciertos conocimientos generales compartidos, el predicado “...surcan los cielos” es predicable de “los pájaros” o “los aviones” pero no de “los peces” o de “las mesas”, pese a que se pueda construir una oración gramaticalmente correcta.

1.2.3. Dinámica de actualización de los esquemas cognitivos de marcos y guiones

La memoria episódica, donde se construyen los modelos de situación, almacena información sobre el grado de asociación entre los distintos componentes de las experiencias singulares (Cohen y Eichenbaum, 1993; Shastri, 2002; Sutherland y Rudy, 1989). Esta información constituirá la materia sobre la que se producirán los procesos de abstracción, generalización y descontextualización que permiten la construcción de los esquemas de marcos (Minsky, 1974) o guiones (Schank y Abelson, 1977) en la memoria semántica. Son procesos inductivos basados en la frecuencia relativa de la repetición y coocurrencia de componentes a través de situaciones parecidas (ver sec. 1.2., cap. 1 & sec. 2.2. cap. 4 de esta pt. II). En función de esto, los esquemas de conocimiento genérico heredarán la mayoría de las categorías de los modelos de situación. No así para aquellas que indiquen singularidad como, por ejemplo, las de espacio y tiempo (Winocur, Moscovitch, Sekeres, 2007). De este modo, los modelos tienen una influencia decisiva en la conformación de los esquemas. Por otra parte, los esquemas también influyen sobre los modelos. Cuando se consolidan, contribuyen a la comprensión de las situaciones permitiendo enlaces y relaciones entre las categorías de los modelos de situación a modo de puentes inferenciales.

Para formar un esquema de conocimiento genérico estable y cristalizado, es crucial una alta frecuencia de repetición y coocurrencia de componentes en situaciones similares. Esta alta frecuencia de repeticiones fortalece las conexiones entre las categorías dentro del esquema, creando estructuras estables. Esto permite que la activación de una categoría se propague rápidamente a otras categorías estrechamente vinculadas dentro del esquema (e.g. Cohen y Eichenbaum, 1993). Cuanto más exclusivo le sea el contenido de la categoría activada, más probable es que acabe activando el esquema completo. Por ejemplo, la categoría 'camarero' activará más específicamente el esquema 'restaurante' que la categoría 'comida'. Así, los esquemas se perciben como unidades complejas y relativamente estables, aunque pueden cambiar lentamente con la incorporación de nueva información, según sea su consistencia con el estado actual del esquema.

Siguen siendo estructuras flexibles. Otra consecuencia de esa consolidación por inferencia inductiva a partir de la frecuencia relativa es que los esquemas constituyen creencias *derivadas*, y

su grado de credibilidad se corresponderá con el valor de esa frecuencia relativa. Por ejemplo, mi creencia de que el esquema de restaurante incluye al agente camarero dependerá de la frecuencia relativa restaurantes con camarero almacenada en mi memoria. Puesto que, tal como se ha dicho, esa frecuencia debe ser relativamente alta, la credibilidad de la representación esquemática también lo será. Por eso mismo formarán parte del sistema de creencias.

1.2.4. Acerca de los contenidos de las representaciones

Sólo un breve apunte. Hasta aquí se ha abordado la cuestión de las condiciones necesarias de la estructura de la representación para que pueda ser o no creída. En cuanto a su contenido, en principio, las creencias pueden ser acerca de cualquier asunto. En todo caso, en páginas anteriores ya se expuso que, debido al propósito principal del sistema de creencias, lo adaptativo de éstas es que se ajusten a cómo es el mundo; así pues, el *entorno ecológico* es uno de los determinantes de sus contenidos. A su vez, en tanto que hacer predicciones y conjeturas es importante para la supervivencia, las *inferencias* basadas sobre esos contenidos experienciales determinarán nuevos contenidos en el sistema de creencias. Finalmente, también vimos que el bagaje y el *acervo social* y *cultural* también moldea y determina las creencias de los individuos pertenecientes a una comunidad epistémica.

En cualquier caso, acerca de esta última fuente de formación de creencias, cabe puntualizar un matiz. Las creencias, además de centrarse en lo que puede, pudo o podrá ser, mediante la transmisión cultural podemos sostener creencias insertadas en *mundos ficticios* que no creemos. Puedo creer que fue Arturo y no Merlín quien arrancó Excalibur de la piedra, o, leyendo una novela policíaca o viendo una película, puedo creer que al final el asesino es, digamos, Juan y no otro personaje. Puedo sostener redes de creencias dentro de cúmulos informativos que no creo. Nichols y Stich (2000, 2003) proponen que estas creencias estarían contenidas en un espacio de trabajo mental separado al que han llamado “caja de los mundos posibles”. Consideran que tal espacio es parte de la estructura básica de la arquitectura mental de la mente humana, y que funciona igual que la dinámica de las creencias basadas en lo real (para una ampliación, ver Badura y Berto, 2019).

Lo dicho, a su vez, es compatible con el supuesto que se hace en el presente modelo de que, cuando el referente es la realidad, en el sistema de creencias sólo constan contenidos considerados como tentativamente verdaderos, es decir, probables (e.g. Lewis, 1946; Alcock, 2018) o tentativamente correctos (e.g. Sant’Anna, 2018; Van Dijk, 2014). Esto implica que el sistema de creencias no alberga contenidos considerados absolutamente inciertos ni improbables, pues considerar que el sistema de creencias almacena contenidos que no son creídos llevaría al absurdo. En cualquier caso, en tanto que la credibilidad es una cuestión de grado sobre un total, siempre que me creo algo *hasta cierto punto* estoy dando cabida a la posibilidad (que no creo) de su contrario por la irrupción de algún refutador. Si considero algo como tentativamente cierto, automáticamente, estoy considerando su contradicción como tentativamente falsa. En la presente propuesta, pensar en lo falso o poco probable consiste en una operación que se aplica sobre lo

considerado tentativamente verdadero. Estas cuestiones se volverán a tratar en apartados subsiguientes.

1.3. La adhesión

Llegamos aquí al tercer elemento del esquema S – A – R, consistente en el vínculo que se produce, en primera instancia, entre la representación y la subjetividad (que puede, a su vez, constar de otras representaciones) y, más en detalle, entre las categorías de la representación en cuestión.

1.3.1. Elementos básicos implicados en el proceso de adhesión

Para proceder con su descripción, puede ser adecuado adelantar que, en la presente propuesta, se considera al sistema de creencias como una red compuesta de nodos y arcos en la memoria a largo plazo del individuo.

Hay distintos tipos de nodos en función del nivel de granularidad al que uno atienda. Para los presentes objetivos, interesan principalmente dos niveles. Por un lado, el nivel de las representaciones globales macro-estructuradas (RG), como los modelos y los esquemas genéricos, que conforman las representaciones del sistema de creencias albergado en la memoria del sujeto. Por otro lado, el nivel de las categorías que componen esas representaciones mentales globales, constituyendo un nivel local de representaciones micro-estructuradas, como conceptos (i.e., participantes, acciones, etc.) o propiedades. Los arcos son los elementos encargados de producir la adhesión. Se corresponden con enlaces que expresan tipos de relaciones entre nodos. En el sistema de creencias, estas serán las relaciones tópicas, un subconjunto de relaciones de coherencia.

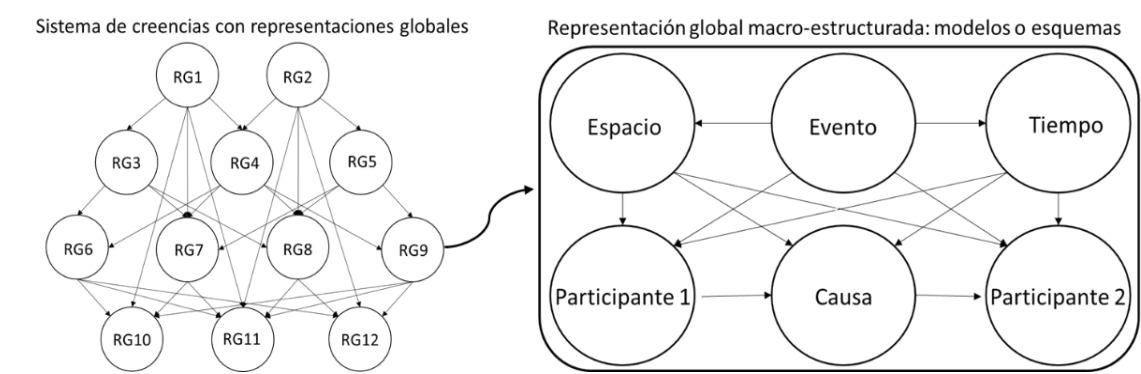
El establecimiento de relaciones tópicas entre dos o más nodos permite la transmisión de justificación de unos a otros. En este sentido, los arcos o enlaces de relación tópica permitirán transmitir justificación de una representación global previamente justificada en el sistema de un individuo (i.e., una creencia) hacia otra representación que está menos justificada, o no lo está. La adhesión que se produce en este nivel es de carácter *inter-representacional*.

Las categorías que componen estas representaciones globales (listadas más arriba) son de tres tipos generales: categorías que incluyen conceptos de objetos y acciones, categorías de propiedades y categorías de relaciones. Estas últimas ejercen de arcos dentro de la propia representación global, enlazando los nodos locales de las categorías conceptuales entre ellos (p. ej. participantes con acciones) y con los de propiedades. De este modo, estos arcos logran estructurar la representación global en una unidad coherente en la mente del individuo.

Con un mero fin ilustrativo, la Figura 2 esboza el sistema de creencias a través de representaciones generales (a la izquierda) y detalla la estructura interna potencial de una de dichas representaciones, mostrando sus diversas categorías (a la derecha).

Figura 2

Ilustración de la Composición del Sistema de Creencias y sus Representaciones



Nota: a la derecha, los círculos representan nodos o contenidos del sistema de creencias. Esos contenidos consisten en representaciones globales (RG): representaciones complejas compuestas por categorías y relaciones. Las flechas representan las relaciones que se establecen entre esas RG. A la izquierda, se presenta la estructura y composición interna de una posible RG, cuyo contenido podría tratar acerca de un evento como “el otro día Juan empujó a Pedro en la cola del cine”.

1.3.2. Proceso de adhesión

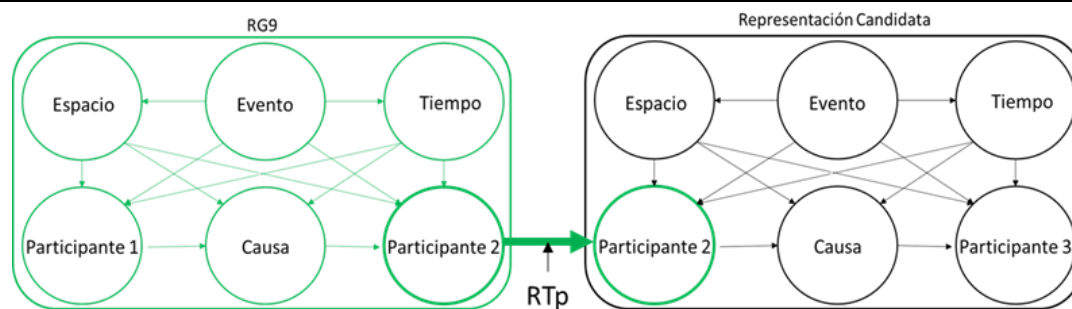
En el curso del procesamiento de la información, muchas veces, la entrada informativa puede resultar vaga, con vacíos, rupturas y nodos que no están conectados explícitamente, o que no están especificados. En el proceso de comprensión de dicha entrada, se le aplica gran cantidad de información procedente de la memoria a largo plazo en orden a obtener un modelo más concreto y específico en el que sus unidades locales queden interconectadas; es decir, en orden a obtener una representación coherente. Y, como ya se ha dicho en repetidas ocasiones, es en ese proceso cuando generamos una interpretación subjetiva de la información. La evaluación de la credibilidad se realiza sobre la base de esa interpretación subjetiva, *ya enriquecida* por contenidos procedentes de la propia memoria del individuo.

La idea general de la presente propuesta es que el sistema de creencias previo, justificado desde sus niveles más básicos y fundacionales, podrá dotar de justificación a una representación entrante vinculándose con ella mediante una relación tópica (i.e., adhesión inter-representacional). Al hacerlo, debido a su relevancia, determinadas categorías de las representaciones globales previamente justificadas del sistema se enlazarán con determinadas categorías de la representación global entrante candidata a ser justificada (RC). En ese momento, en virtud del criterio que rige la relación tópica (RTp), cierta magnitud de la justificación contenida en la representación global previa se transferirá a la nueva representación candidata, tal y como pretende ilustrar la Figura 3 (abajo).

Esa justificación transferida pasará a discurrir ahora también por los arcos de las categorías locales relacionales que componen la representación global entrante o candidata RC. Esos arcos, que ya permitían la construcción de un modelo de la situación coherente y comprensible de RC, ahora, además, permitirán su integración en un modelo justificado y creíble. La adhesión que se produce en este nivel es de carácter *intra-representacional*. La Figura 4 pretende ilustrar este proceso. Sus

Figura 3

Conexión Tópica entre Categorías Locales de RG y RC en la Transmisión de Justificación



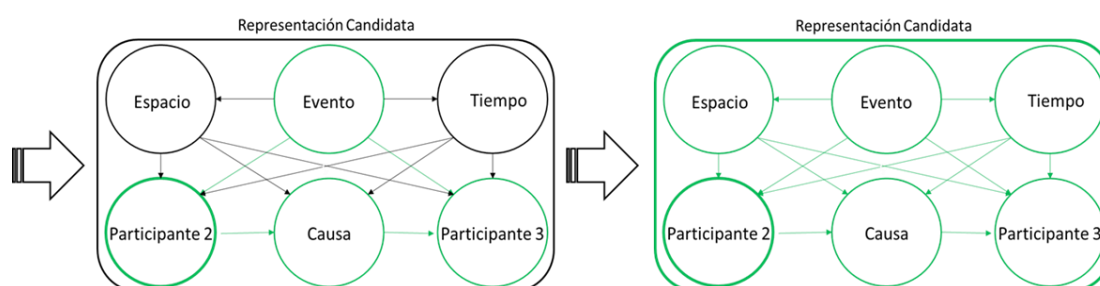
Nota: Tomemos aquí la Representación Global 9 de la red presentada en la figura anterior. El color verde significa que RG9 ya tiene una carga de justificación suficiente como para formar parte de la red de creencias. La representación candidata queda conectada mediante una relación tópica (*RTp*) a RG9. De hecho, la conexión tópica entre ambas representaciones se produce en el nivel de las categorías locales (en esto, similar a la superposición argumental). Esta conexión permitirá el paso de la justificación de RG9 a la candidata.

respectivos componentes pasarán a adherirse y cohesionarse entre ellos de tal modo que la representación global entrante podrá quedar fijada y adherida al sistema de creencias. Así pues, el proceso es de ida y vuelta: el sistema envía justificación, se produce la adhesión intra-representacional en RC y solo ahora se puede dar la adhesión a nivel inter-representacional, por la que RC pase a ser un nuevo miembro *RGn* de la red de creencias del individuo.

Por todo lo dicho, se puede decir que la adhesión se produce a dos niveles: a un nivel interno a la propia representación candidata vinculando justificadamente sus categorías, y a un nivel más general, vinculando justificadamente la representación candidata con el sistema de creencias.

Figura 4

Flujo de Justificación en la Conexión de Categorías Internas a RC.



Nota: Una vez ha recibido cierta magnitud de justificación procedente de RG9, esa justificación fluye a través de los arcos de RC conectando justificadamente las categorías pertinentes de RC. Pongamos por caso que yo no tenía ninguna razón para suponer que “el otro día Juan empujó a Pedro en la cola del cine” pese a entender el significado de esa oración. Pero sí que tenía en mi memoria a largo plazo la creencia de que “Juan es impulsivo”. Cuando esta creencia previa entra en contacto con “Si alguien es impulsivo, entonces puede ser violento”, la superposición argumental del término medio “impulsivo” permite establecer una justificación para creer que “Juan puede ser violento”. Esto es, que la categoría de propiedad “violento” queda vinculada justificadamente con el participante “Juan”. Así, “violento” se puede predicar de “Juan”. Esto hace que ahora le atribuya más probabilidad a la representación que me he formado de Juan empujando a Pedro en una cola.

1.3.3. Características de la adhesión: grados de intensidad y probabilidad

La adhesión es la propiedad por la cual se unen distintas entidades, y varía en *intensidad*. En una creencia derivada, la intensidad de su adhesión viene determinada por la magnitud de la justificación que ha heredado de las creencias previas con las que se ha enlazado. Asimismo, si esa magnitud es lo *suficientemente* elevada, esta nueva creencia podrá pasar a justificar ahora a nuevas representaciones candidatas.

Esa intensidad de la adhesión se medirá como *grados de probabilidad subjetiva*. La dinámica del proceso de transmisión de justificación funcionará dentro de los cauces delimitados por el cálculo de probabilidades, respetando una *consistencia probabilística local*. Por otro lado, conviene no confundir la intensidad de la adhesión con su *estabilidad*. Brevemente, la intensidad de la adhesión es un reflejo de la credibilidad medida como una probabilidad subjetiva, mientras que la estabilidad de la adhesión hace referencia a la persistencia o permanencia de la unión entre los elementos en cuestión. La intensidad depende, como se ha dicho, de la magnitud de la justificación recibida, independientemente del número de creencias previas que le otorgan justificación. Sin embargo, la estabilidad depende crucialmente de ese número (cuanto mayor sea el número de conexiones, mayor será la estabilidad de la creencia). De este modo, puede haber creencias derivadas con cualquier combinación de magnitud de intensidad y de estabilidad. Por ejemplo, mi creencia de que “el otro día Juan empujó a Pedro en la cola del cine”, puede quedar justificada en cierto grado a partir de mis creencias previas de que “Juan es impulsivo” y de que “si alguien es impulsivo, entonces puede comportarse de forma violenta”. Además, si María, quien presenció el evento y no tiene motivo para engañarme, me lo relata, la estabilidad de mi creencia aumenta, aunque esto no incrementa necesariamente su credibilidad. Pues bien, tal y como se dijo en el punto anterior, cuando la representación candidata se pone en contacto con la subjetividad, ésta la interpreta y evalúa hasta qué punto sus elementos están unidos. El resultado de esta evaluación dependerá de encontrar o no otras creencias en el sistema mediante las que pueda suministrar justificación a los arcos que unen las piezas internas de la representación. De ser así, esa representación candidata estará hasta cierto punto justificada para el sujeto, y será creída. De lo contrario, no.

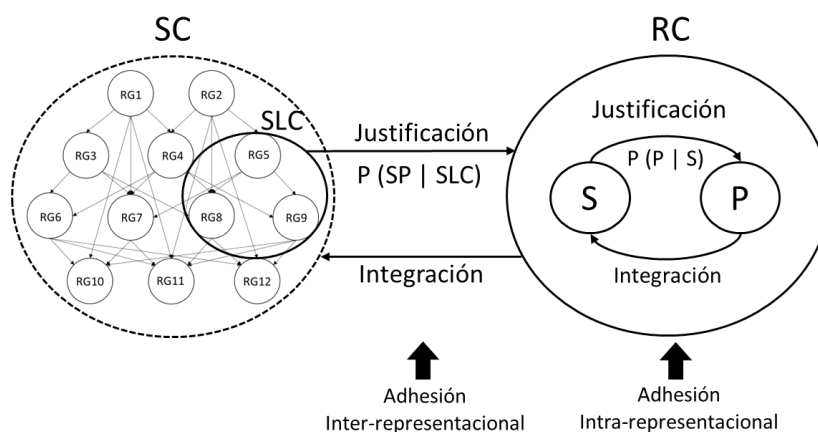
Pasando ahora del nivel de las representaciones mentales al nivel de las categorías de la expresión lingüística, diremos que el contenido de una oración pasará a ser creído cuando una *razón* previa justifique la atribución de un *predicado* a un *sujeto*. Esto significa que, cuando algo no se cree, es porque no hay una subestructura de red que establezca adecuadamente la relación entre la categoría representada por el sujeto y la representada por el predicado en la mente del individuo. Esto es compatible con una idea que ya vimos en la sec. 5, cap. 2, pt. I, a saber, que, en términos lingüísticos o semánticos, una verosimilitud se correspondería con el *grado de participación condicionada de un predicado en un sujeto*. Así, la pregunta de cuán creíble o verosímil resulta el contenido de una oración podría reformularse como la pregunta de cuán probable es un predicado

(P), dado un sujeto (S) para un individuo en cuestión²⁸. Es decir, podría reformularse como la probabilidad condicionada $P(P|S)$. Nótese que estamos en el nivel de justificación interno a la propia representación candidata (adhesión intra-representacional).

Debemos tener en cuenta que hablar de probabilidad en el ámbito de las creencias significa hablar de probabilidades subjetivas. Y, como también vimos en la sec. 11.2., cap. 2, pt. I, las probabilidades subjetivas que se atribuyan a una representación particular dependen de los conocimientos y creencias previas de los individuos. En este sentido, esa probabilidad vendrá determinada por representaciones previamente justificadas que ejerzan de razones. Y que estén previamente justificadas significa que ya residan en el sistema de creencias del individuo (SC). Es decir, que, finalmente, la pregunta es la de cuán subjetivamente probable resulta el contenido de una oración compuesta por un sujeto y un predicado (SP), dado su sistema de creencias (i.e., $P(SP|SC)$). Algo así como “si mi sistema de creencias SC está bien calibrado, entonces consideraré probable el contenido SP hasta cierto punto X”. Como veremos más adelante, esa probabilidad no le vendrá derivada inmediatamente de *todo* el sistema de creencias, sino de las razones contenidas en un *foco local relevante* (L) de dicho sistema, activado por la información entrante (i.e., $P(SP|SLC)$). Nótese que estamos ahora en un nivel de adhesión más general: el que se produce entre la representación global candidata entrante y el sistema de creencias que forma parte de la subjetividad del individuo (adhesión inter-representacional).

Figura 5

Adhesión Inter e Intra-Representacional.



Nota: En la imagen, el círculo de líneas discontinuas está representando el sistema de creencias del individuo (SC), mientras que el círculo interior con línea gruesa continua está representando el foco local de creencias activado por la representación candidata entrante (SLC). Esa representación candidata será contrastada con SLC, y la verosimilitud de su contenido será evaluada en función de SLC, dando como output una probabilidad subjetiva (i.e., $P(SP|SLC)$). La evaluación de la verosimilitud del contenido de la representación candidata es equivalente a plantearse hasta qué punto se puede predicar P dado S en función de un SLC concreto. De este modo, RC se puede integrar en SC.

²⁸ He adoptado aquí estos términos lingüísticos, por un lado, para evidenciar su relación directa con las cuestiones debatidas en la sec. 5, cap. 2, pt. 1, y por otro, por mera comodidad ya que hablar en términos de sujetos, predicados y atribuciones es más ágil. Pero lo central siguen siendo las representaciones y sus constituyentes.

La figura 5 ilustra gráficamente los procesos de adhesión inter e intra representacional, que conforman el núcleo del modelo de la dinámica de creencias que se presenta en esta tesis.

1.3.4. Credibilidad, probabilidad y refutadores en la memoria

Ese grado mencionado de *participación condicionada* del predicado en el sujeto implica que, en según qué condiciones, podría participar y que, en según qué otras, podría no participar. La plena participación en *todas* las condiciones *conocidas por el individuo* le justifican para *asumir* que la probabilidad puede ser total, pudiendo establecer una relación necesaria del predicado respecto del sujeto. El conocimiento de esa plena participación es la razón por la cual podría juzgar la oración SP como verdadera. Por ejemplo, puedo considerar verdadera la oración “el cuervo x es negro” puesto que sólo sé de la existencia de cuervos negros. Por otro lado, la participación en *la mayoría* de las condiciones conocidas por el individuo es la razón que le justifica para asumir una relación de probabilidad dentro de la posibilidad contingente. Por ejemplo, “el pájaro x vuela”. En este caso, habiéndose dado S, *prima facie* se dará P, contemplando la posible intervención de determinadas condiciones que dificultan o impiden su aparición. Estas condiciones son los rebatidores o refutadores. En razonamientos cotidianos deductivos como los entimemas, esos refutadores serían las denominadas condiciones desactivadoras / inhibitoras (i.e. los *rebatidores refutativos* de Pollock, 1987). Y en razonamientos inductivos como la ejemplificación, serían los antecedentes alternativos (i.e. los *rebatidores socavadores* de Pollock, 1987). De estos dos tipos de refutadores se habló extensamente en la sec. 13.2., cap. 4, pt. I, al abordar el razonamiento rebatible. Así pues, se puede decir que, en el sistema de creencias, además de las relaciones positivas de justificación, también existen relaciones negativas de *refutación*. Estas relaciones de refutación limitan las de justificación, produciendo el efecto contrario a la adhesión. En cualquier otro contexto se le denominaría efecto de repulsión. Pero dada la similitud del sistema de creencias con la argumentación (que justificaré de aquí a un momento), le denominaremos efecto de *disensión*. De todos modos, su efecto sobre las creencias tiene una incidencia menor al de las relaciones positivas de justificación; de otro modo, no serían creencias.

En lo sucesivo emplearé preferentemente las expresiones de refutadores o rebatidores para referirme a cualquier tipo de elemento derrotador que produzca este efecto de falibilidad y disensión. Pero no está de más precisar que se pueden distinguir, como mínimo, tres clases generales:

- 1) Refutadores explicativos: pueden adoptar la forma de contraargumentos u objeciones. En el caso del condicional “si piso el pedal del freno, entonces el coche se detiene”, concebir la posibilidad de que “si hay hielo en la carretera, el coche no se detiene” y de que “hay hielo” sería un contraargumento que explicaría la ausencia del consecuente pese al antecedente. Asimismo, simplemente aducir que “hay hielo en la carretera” también lo sería. Implica recurrir a un conocimiento cuya posible presencia *explica el porqué* de la no ocurrencia del predicado o consecuente. Se trata de refutadores procedentes de la memoria semántica.

- 2) Refutadores asociativos basados en casos: representaciones de colecciones de datos que conforman un cuerpo de evidencia contraria respecto a la representación crítica bajo cuestión; pueden adoptar la forma de contraejemplos o de meras excepciones, crucialmente, sin poder explicativo. Por ejemplo, la representación crítica de “los pájaros surcan los cielos” se podría ver refutada o matizada por la activación del conocimiento de contraejemplos, almacenados en la memoria semántica, como el avestruz, la gallina, el pingüino, o cualquier pájaro recién salido del cascarón. De todos ellos, uno simplemente podría saber que, aun siendo pájaros, no vuelan. Sin saber por qué. De modo similar, ante la representación de que “si hay nubarrones, entonces llueve”, a uno se le podría activar el conocimiento episódico de que, simplemente, “a veces ha estado muy nublado y no ha llovido”. Es decir, el conocimiento de meras excepciones, en este caso, almacenadas en la memoria episódica, sin identificar siquiera ningún tipo de categoría de nubes. Uno puede desconocer totalmente las clases de nubes portadoras de lluvia, o los motivos o condiciones por los cuales, pese a estar nublado, no ha llegado a llover. Implica saber simplemente que a veces no pasa, sin poder recurrir a ningún tipo de conocimiento más. Así pues, en ninguno de estos casos se aducen contraargumentos explicativos.
- 3) Refutadores de evidencia negativa: reconocimiento de las limitaciones del estado epistémico personal (autorreflexivo) y general (científico y social). Por ejemplo, con relación a una creencia como “Marte es un planeta inerte y desierto”, aun creyéndolo, uno podría albergar ciertas reservas del tipo: “puede que hayan descubierto vida y que yo no lo sepa” o “el hecho de que no hayan descubierto vida en Marte no implica que no la haya”. Como puede observarse, no es que haya evidencia contraria a la creencia que uno sostiene, sino que la evidencia que uno conoce en un momento dado no es suficiente para estar absolutamente seguro de que el contenido de esa creencia crítica sea verdadero.

Dentro de cada una de estas tres clases se deben poder distinguir varios subtipos, muchos de los cuales podrían ser relevantes para el análisis de la argumentación. Los trabajos de Walton (e.g., Godden y Walton, 2007) sobre las “preguntas críticas” van en esta dirección, clasificando tipos de refutadores en función del esquema de argumentación que rebaten (ver ap. C). Se centra, sobre todo, en los refutadores explicativos. Pero, sin duda, debe haber también subtipos de refutadores asociativos. En este trabajo no atenderemos más esta distinción. Como se verá más abajo, aquí nos centraremos en la medición de la credibilidad como probabilidad subjetiva. Cualquier refutador, independientemente de su clase, tendrá unos efectos que se verán reflejados y recogidos en (el complementario de las) probabilidades subjetivas, tal como se especificará en la sec. 4, cap. 6, pt. II acerca de la escala de medición.

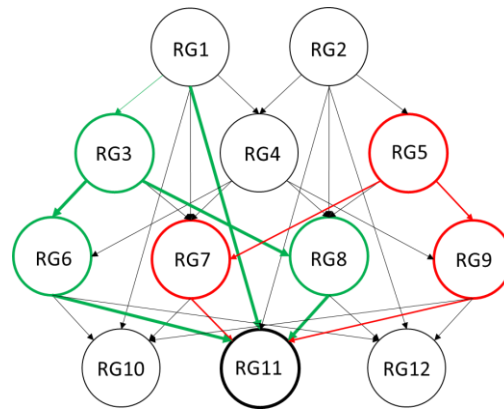
1.3.5. El Sistema de creencias entendido como un mapa argumentativo

El presente modelo considera que el sistema de creencias está compuesto por nodos de representaciones globales y arcos de relación. En páginas anteriores ya se ha hablado de los aspectos que determinan la credibilidad de las creencias fundacionales, y de que éstas son siempre *prima facie*. Por otro lado, la credibilidad de una nueva representación global derivada queda

determinada por la magnitud de la justificación y refutación que recibe de los nodos representacionales con los que se enlaza en el sistema. Esas magnitudes de justificación y refutación fluyen a través de los arcos de relación, tal como pretende mostrar la Figura 6.

Figura 6

Credibilidad como Función de las Relaciones de Justificación y Refutación del Sistema.



Nota: El color verde se corresponde con las *relaciones de apoyo* y el rojo con las *de refutación*, y el grosor de las flechas se correspondería con la intensidad de la relación. Así, aquí se muestra que el nodo o representación global RG11 recibe tres arcos gruesos de apoyo y dos finos de refutación. Esto implica que será relativamente estable y que su credibilidad será alta, pero no total.

Esto está en consonancia con las observaciones (previamente discutidas en las sec. 14 y 15, cap. 4, pt. I) sobre condiciones desactivadoras y causas alternativas: varias investigaciones experimentales muestran que los individuos acceden de manera rutinaria a estructuras relevantes de su memoria a largo plazo, donde almacenan tanto elementos de apoyo como refutadores (e.g. Cummins, 1995; Cummins, Lubart, Alksnis y Rist, 1991; De Neys et al., 2003a; Quinn y Markovits, 1998). Esos estudios, a su vez, indican que los individuos integran estos refutadores en el curso del razonamiento (e.g. Bonnefond, Kaliuzhna, Van der Henst y De Neys, 2014; De Neys et al., 2002; De Neys et al., 2005a; Grosset, Barrouillet y Markovits, 2005; Pijnacker, Geurts, van Lambalgen, Buitelaar, y Hagoort, 2010; Quinn y Markovits, 2002).

Además, existe abundante evidencia demostrando cómo, durante el proceso de comprensión y validación del discurso, las personas enlazan la información recién recibida con contenidos relevantes previamente almacenados en su memoria a largo plazo (e.g. Kintsch, 1988, 1998; Kuperberg, Lakshmanan, Caplan, y Holcomb, 2006; Myers y O'Brien, 1998; O'Brien y Albreht, 1992; O'Brien, y Cook, 2016; O'Brien, Cook, y Guéraud, 2010; Rizzella y O'Brien, 2002; Singer, y Ritchot, 1996; Van Dijk y Kintsh, 1983). En ese proceso, el sistema cognitivo puede detectar de forma automática y rutinaria posibles inconsistencias y contradicciones entre la información entrante y la almacenada en la memoria (e.g. Albreth y O'Brien, 1993; Isberner y Richter, 2013; Myers, O'Brien, Albreht y Mason, 1994; Piest, Isberner, y Richter, 2018; Richter, Schroeder, y Wöhrmann, 2009; Singer, Halldorson, Lear, y Andrusiak, 1992). Así pues, estos hallazgos apoyan también la idea de un acceso rutinario (y on-line) a estructuras con refutadores.

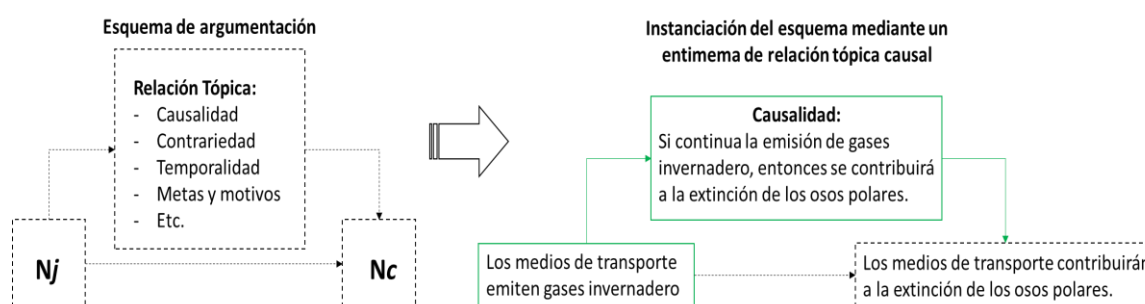
1.3.5.1. Relaciones entre la dinámica de la justificación y los esquemas argumentativos

Dada la complejidad de las relaciones de apoyo y refutación dentro del sistema de creencias, algunos nodos o representaciones globales serán totalmente creídas, mientras que otras lo serán en menor grado. Los arcos que permiten la adhesión conectarán un nodo o representación justificada hasta cierto punto con otro nodo o representación menos justificada. Esta conexión se produce estableciendo una relación tópica entre ambos nodos (por ejemplo, una relación causal de X a Y). Esta relación tópica es el criterio que establece la relevancia justificativa de la conexión entre ellos, haciendo *pertinente* a la representación previa como *razón* para creer la representación candidata (en función de los conocimientos del individuo: en el ejemplo, este debe saber que X causa Y). Tales relaciones tópicas son concebidas como reglas hipotéticas generales que actúan como un puente inferencial *garantizando* el paso de cierta proporción de justificación entre las representaciones o nodos. Este conjunto de ‘representación justificada + representación a justificar + regla de relación justificativa’ instancia un *esquema de argumentación*. Como se desarrolla en el ap. C, los esquemas de argumentación se pueden expresar formalmente como condicionales genéricos, plausibles y rebatibles, del tipo “si $R(X, Y)$ y X , entonces *normalmente* Y ” (Vega, 2012). En este esquema, R denota una *relación* que se establece entre dos representaciones o nodos X e Y . En el contexto del modelo que estamos desarrollando aquí, lo expresaríamos como “si $Tp(Nj, Nc)$ y Nj , entonces *probablemente* Nc ”, donde Tp denota la existencia de una relación tópica entre Nj y Nc . Por su parte, Nj denota el nodo previamente justificado o razón y Nc el nodo candidato a ser justificado por Nj .

La Figura 7 (que sintetiza otra más detallada desarrollada en el ap. C) ilustra un diagrama de esquema argumentativo con un ejemplo concreto. Así se clarifica la posición de los elementos que se discuten en el ap. C —como premisas menor y mayor, y conclusión— dentro de esta conceptualización psicológica.

Figura 7

Esquema de Argumentación en Términos de Nodos y Arcos Relacionales.



Nota: A la izquierda, se muestra la estructura del esquema de argumentación, identificando a Nj como un nodo o representación global ya justificado y a Nc como un nodo candidato a justificación, junto con la relación tópica que facilita esta transmisión de justificación basada en criterios específicos. A la derecha, se presenta una instanciación de esta estructura con contenidos específicos.

El diagrama del esquema de argumentación de la Figura 7 se representa en líneas discontinuas dando a entender su carácter potencial, relativamente abstracto. Deberá ser actualizado e instanciado por materiales relativamente concretos. Como se mencionó en el cap.1, pt. I, el esquema de argumentación se compone de una serie de variables o celdillas vacías que deberán ser rellenadas por constantes fijas concretas (idea que se desarrolla más extensamente en el ap. C). En el entimema causal que lo ejemplifica, el elemento con líneas discontinuas se corresponde con la representación candidata entrante N_c pendiente de ser justificada. Los elementos en verde se corresponden con las creencias y relaciones ya justificadas y aceptadas por un supuesto individuo. De este modo, en el intento de dotar de justificación, cuando el sistema recibe la entrada informativa N_c , buscará un N_j al que conectarla mediante una relación tópica.

Los mecanismos que transmiten justificación de N_j a N_c no son puramente formales, reflejando la naturaleza no formal de las relaciones tópicas. Decir que X *implica* Y , una relación formal, es distinto a decir que X *causa* Y , una relación no formal. También hay que distinguir entre los tipos de relaciones: no es lo mismo decir que X causa Y , que decir que X *es similar a* Y . La naturaleza y la cualidad de estas relaciones varían significativamente. Los tipos de relaciones constriñen el tipo de creencias a que pueden dar lugar. Por ejemplo, una relación de semejanza entre X e Y no hace creer en la existencia de Y a partir de darse X , aunque la existencia de una propiedad z en X sí puede justificar subjetivamente la presencia de una propiedad z en Y , siempre y cuando se conozcan bien estas conexiones entre X e Y . Por tanto, las relaciones tópicas son, en esencia, epistémicas, vinculadas al conocimiento que posee el individuo.

1.3.5.2. Interacción entre contenidos y refutadores, incertidumbre y falibilidad

Más relevante aún es el hecho de que su instanciación en contenidos concretos es lo que hace que tales relaciones no sean dicotómicas en la mente del individuo. La incertidumbre y los refutadores se dan sobre las reglas y los conocimientos que los individuos tienen acerca de los asuntos del mundo. Eso es lo que está en la base de su falibilidad y probabilidad. De este modo, manteniéndonos en las relaciones tópicas de causalidad, la creencia de que “ X causa Y ” pasa a la de que “ X , normalmente (o probablemente), causa (o causará) Y ”.

La relación causal del esquema puede ser directa, o puede requerir de un *respaldo*²⁹, como en el caso del ejemplo anterior del rol de los medios de transporte en la extinción de los osos polares. En efecto, la relación causal allí expuesta requiere de una cadena de relaciones más compleja que no figura explícitamente en el ejemplo (i.e., la *implicación* presente en esa relación causal requiere de una *explicación*):

- a) Que la emisión de gases invernadero contribuye al aumento del calentamiento global.
- b) Que, si aumenta el calentamiento global, entonces se derretirán los polos terrestres.
- c) Que el polo norte es el hábitat del oso polar.

²⁹ Para más información sobre este concepto, ver el modelo de Toulmin (1958), presentado en sec. 4.2. cap. 1, pt. I.

d) Que, si se destruye el hábitat de una especie, normalmente esa especie se extingue.

Todos estos elementos parecen necesarios para poder aceptar la relación causal presentada en el condicional del ejemplo. En ese sentido, es plausible asumir que toda esa información se activa (o pre-activa) en la memoria al establecer la relación causal, aunque no devenga consciente necesariamente. Como se desarrolla en el ap. C, los argumentos causales cotidianos son la punta de un iceberg discursivo complejo (Van Dijk, 2012; Vega, 2004). El modelo mental del que es consciente el individuo estará limitado por la información más accesible que viene a la mente (Taylor y Fiske, 1978; Wyer y Albarracín, 2005). Pero su aceptación también involucra la activación no necesariamente consciente de subcomponentes causales con probabilidades variables (Hahn, Bluhm y Zenker, 2017; Hastings, 1962).

1.3.5.3. *Integración de la red de creencias y argumentación: un enfoque dinámico y relacional*

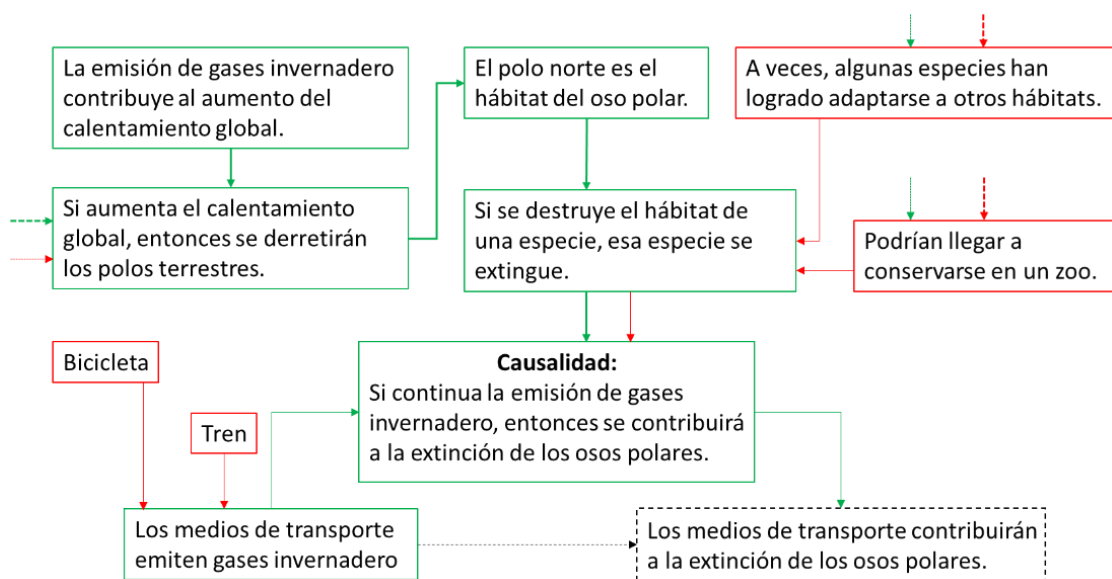
En todo caso, esta activación permite, como consecuencia, producir la adhesión de N_c a partir de N_j . Además, también se activará la información correspondiente a sus respectivos refutadores, lo que produce la cualificación de ese “normalmente” o “probablemente”. En este caso, los únicos refutadores a los que yo puedo acceder (en función de mis conocimientos previos) son los de que “a veces, algunas especies han logrado adaptarse a otros hábitats” (donde tendría en cuenta ese *a veces* y ese *algunas*), y que “podrían llegar a conservarse en un zoo” (teniendo en cuenta la propia reserva expresada por ese *podrían*). Luego, es posible que la creencia justificada “los medios de transporte emiten gases invernadero” también permanezca almacenada con algún refutador en la memoria del individuo que la cualifique, y que esto, además, dependa de la interpretación que haga tal individuo. En mi caso, si la interpreto tal cual está expresada, en este formato lingüístico (en un tiempo verbal presente), sólo podré acceder al tren y a la bicicleta como posibles refutadores. Pero si la interpretase de una manera atemporal, entonces podría recurrir a toda la gama de vehículos eléctricos que está por venir, así como a todos los tipos de transporte empleados con anterioridad al s. XVIII. En función de esto, el grado de aceptabilidad de esa supuesta creencia almacenada variaría. Así, podríamos enriquecer el esquema anterior al modo ilustrado en la Figura 8 (abajo).

La ilustración de la Figura 8 (abajo) pretende mostrar la complejidad del *fragmento* de la red de creencias plausiblemente activada en la mente de un individuo en el proceso de justificación de una representación candidata a ser o no creída. Precisamente, el sistema local de creencias consiste en esa red activada en un momento dado. La Figura 8 permite comprender la similitud de esta propuesta de red de creencias con la argumentación, especialmente con el modelo de Toulmin (1958) presentado en la sec. 4.2., cap. 1, pt. 1. También se asemeja a varias propuestas de diagramación o mapeo de argumentos en inteligencia artificial (e.g., Rapanta y Walton, 2016; Twardy, 2004; van Gelder, 2002, 2007), donde se presentan redes argumentativas que integran respaldos y refutadores, algo propio del razonamiento rebatible. Igual que en las figuras anteriores, el verde representa contenidos almacenados de apoyo y el rojo, de refutación. Las flechas discontinuas verdes y rojas indican que esas representaciones y relaciones pueden verse, a su vez, apoyadas y objetadas por otros elementos del sistema, no necesariamente activados (pese a que puedan tener influencia sobre el grado de justificación de los contenidos activados). Desde luego,

dadas las limitaciones de la memoria de trabajo, no sería plausible asumir que siempre deba darse un rastreo exponencial de razones hasta las fundacionales. Si bien las fundacionales son necesarias para la justificación del sistema, no siempre tienen por qué influir directamente sobre cualquier creencia. La justificación se propaga de unos nodos a otros a través de las distintas relaciones inferenciales que el individuo puede establecer entre ellos, incorporando el efecto de los refutadores en el trayecto. En esta red, unos mismos contenidos pueden ejercer ora como conclusiones, ora como premisas en los razonamientos que participan en la transmisión de justificación entre representaciones.

Figura 8

Sistema Local de Creencias Entendido como un Mapa Argumentativo.



Nota: Los mapas argumentativos (e.g. van Gelder, 2007) constan, típicamente, de unas premisas de apoyo y unas objeciones a una conclusión en concreto. Pero tanto esas premisas de apoyo como esas objeciones lo son en tanto que sean creídas o consideradas correctas por el individuo. Aquí, estamos describiendo el sistema de creencias como una red de nodos y arcos con relaciones de justificación y refutación en base a los contenidos que los instancian. Tales elementos solo tendrán efectos si logran ser (pre)activados en la mente del individuo.

En líneas generales, este modelo es consistente con la mayoría de las teorías actuales sobre argumentación y los estudios sobre razonamiento cotidiano y comprensión del discurso. A su vez, es consistente con la idea de que, tras aceptar una representación en el sistema, esta sirve como criterio para validar nueva información. El ejemplo de la figura muestra una estructura de red concreta donde las ideas se apoyan o refutan de manera lineal. Sin embargo, también existen estructuras no lineales, coordinadas, que combinan múltiples estructuras lineales para otorgar estabilidad a ciertas creencias. Por otro lado, este ejemplo sugiere que ciertos elementos pueden activarse plausiblemente en la mente del individuo durante el proceso de justificación, implicando la memoria de trabajo y de largo plazo. Este fenómeno está ligado a procesos mentales específicos, como la resonancia y validación, previamente mencionados.

1.3.6. Conclusiones

A lo largo de estas páginas se ha tratado de dar una definición operativa del concepto de creencia, lo cual ha llevado a realizar, a su vez, una descripción de los componentes del sistema de creencias que este modelo considera relevantes. Como se ha podido observar a la hora de abordar esa definición, más que la sustitución total del viejo esquema $O - A - P$, se trata de una especificación de dicho esquema, derivando en otro del tipo $S - A - R$. En este esquema, los términos extremos (S y R) se corresponden con la subjetividad del individuo y una representación global respectivamente, mientras que el término medio (A) se corresponde con la relación de adhesión que se produce entre ambos extremos.

En primer lugar, ello ha conllevado adentrarnos en determinados aspectos de la subjetividad relevantes para las creencias según el presente modelo. Tales aspectos atañen a los mecanismos cognitivos para la formación de creencias fundacionales a partir de las tres fuentes clásicas y al lugar donde tales creencias quedan albergadas. Este lugar es la memoria a largo plazo. Esto ha llevado a realizar un pequeño recorrido teórico por los distintos tipos de creencias fundacionales desde la perspectiva de un fundacionalismo moderado. En este tipo de fundacionalismo, todas las creencias fundacionales son totalmente creídas, pero *prima facie*.

Después, se ha pasado a describir el otro de los términos extremos implicado en las creencias, a saber, la representación. En ese apartado, la propuesta clásica de proposición ha quedado sustituida por el concepto más general de representación global. Estas representaciones se entienden aquí como modelos mentales de la situación, por un lado, y como esquemas cognitivos por el otro. En ambos casos se trata de representaciones complejas compuestas de una red de elementos coherentemente ligados entre sí mediante relaciones de justificación. Tal y como se ha visto, estas dos propiedades (complejidad y coherencia) son cruciales para permitir la productividad y el dinamismo requerido en la actualización de credibilidades.

Por último, se ha abordado el término central del esquema: la adhesión. En este punto, se ha dicho que el presente modelo concibe el sistema de creencias como un compuesto de nodos y arcos de unión. Las representaciones son los nodos, y las relaciones tópicas son los arcos. Se ha definido la adhesión como la propiedad por la cual se unen distintas entidades. Cuando esas entidades son representaciones globales macro-estructuradas, hablamos de una adhesión de tipo inter-representacional. Esto es lo que sucede cuando una representación global pasa a quedar integrada en el sistema de creencias. Por otro lado, esas entidades también pueden corresponderse con los componentes de las representaciones globales. Cuando es así, hablamos de una adhesión de tipo intra-representacional. En las creencias derivadas, aquello que permite la adhesión intra-representacional es la justificación procedente de otras creencias previas del sistema que funcionan a modo de razones.

También se comentó que la adhesión varía en intensidad, y que esta intensidad se mide como una probabilidad subjetiva condicionada. Los elementos que implementan esas probabilidades son los componentes representacionales de las creencias. Recordemos que esos componentes o

categorías fundamentan la predicación. En ese sentido, cuando esos componentes representacionales quedan instanciados en categorías lingüísticas, diremos que la intensidad de la adhesión intra-representacional se expresa como la probabilidad subjetiva de un predicado (P) dado un sujeto gramatical (S) o, lo que es lo mismo, como $P(P|S)$. Esa probabilidad interna a la representación viene determinada por la justificación que le haya sido transmitida desde las otras creencias del sistema. Dicho de otro modo, la probabilidad de que P se predique de S para un individuo concreto depende de su sistema de creencias SC. Dado que la representación global candidata está compuesta por S y P, se podría decir entonces que la probabilidad subjetiva de la representación global candidata entrante se expresa como $P(SP|SC)$. En cualquier caso, dadas las limitaciones del sistema cognitivo, SP no se contrastará con todo SC, sino con una región local activada (L) a la que llamo SLC. Por tanto, la adhesión inter-representacional podría expresarse como $P(SP|SLC)$. En cualquier caso, esta distinción es meramente conceptual puesto que ambas probabilidades siempre coinciden. Tal como se ha dicho, la intensidad de la adhesión se mide como una probabilidad subjetiva, y las probabilidades se basan en una estimación (o razón) de casos favorables acerca de algo respecto de la totalidad de los casos conocidos por el individuo. Cuando lo favorable coincide con lo conocido, la probabilidad es total y, por tanto, también la adhesión y la credibilidad. Pero cuando, alternativamente, hay menos casos favorables que la totalidad de los conocidos, entonces ni la probabilidad, ni la adhesión, ni la credibilidad será total. En esta situación, habrá también casos *conocidos* que *no* sean favorables o se contemplarán casos *posibles* desfavorables. Llamo a estos casos refutadores o rebatidores. Si los casos favorables producen el efecto de adhesión, los casos desfavorables producen el efecto de disensión. Las relaciones de justificación del sistema de creencias se ven limitadas por las correspondientes relaciones de refutación. Esto significa que todas las creencias serían totalmente creídas de no ser por el efecto de los refutadores. Por otro lado, esto también significa que todas las creencias que no son totalmente creídas son creencias derivadas, esto es, con una credibilidad obtenida por inferencia. En el texto, se han distinguido tres clases generales de refutadores: los explicativos, los asociativos basados en casos y los de evidencia negativa. De este modo, el sistema de creencias va quedando organizado en la memoria por relaciones de justificación y refutación. Por lo menos, esos son los criterios mínimos cruciales.

Finalmente, se han puesto de manifiesto los paralelismos entre los elementos implicados en la argumentación con los implicados en el sistema de creencias. La conclusión de un argumento se sostiene por una premisa menor concerniente a algún hecho conocido por el individuo, y por una relación tópica en la premisa mayor mediante la cual se le transfiere justificación. La premisa mayor comporta el esquema de argumentación que permite estructurar los otros dos elementos, mostrando la relevancia de la menor para apoyar la conclusión en función del criterio de relación tópica (estas ideas están ampliamente desarrolladas en el ap. C). Por otro lado, consideramos que todo argumento constituye la parte manifiesta de un iceberg discursivo, donde la mayor parte del conocimiento requerido para su interpretación y aceptación depende de muchos elementos que se dan de forma implícita y que tienen asociados probabilidades variables. En la sección

inmediatamente precedente, acabamos de ver cómo todo eso se ve reflejado en el modelo del sistema de creencias que estamos delineando. Aquí hemos visto que una representación candidata entrante (N_c) se vinculará con representaciones justificadas (N_j) del sistema de creencias almacenado en la memoria mediante relaciones tópicas (T_p). Estos enlaces tópicos son los que permiten otorgarle justificación. Los elementos manifiestamente activados (N_c , N_j , y T_p) no son los únicos que intervienen en el proceso. Este modelo asume que se activarán más elementos de dominios de conocimiento relevantes, que constituyen el sistema local de creencias. Tal y como se ha señalado, la investigación psicológica ha mostrado que tanto en los procesos de comprensión del discurso como en los de razonamiento, uno no sólo accede a la información consistente con la información entrante (o que incluso la apoya), sino que también accede a información contradictoria, a refutadores. En este sentido, la trama de enlaces que se activa en los dominios de SLC está compuesta por unas relaciones de justificación y por otras de refutación. Esta red de relaciones operará de manera conjunta derivando en una magnitud final de justificación que le será transferida a la información entrante N_c .

Pues bien, hasta aquí con la descripción de los tres componentes básicos del esquema de creencia $S - A - R$. A continuación, se procederá con la descripción de la naturaleza esencialmente explicativa de los mecanismos inferenciales que permiten la adhesión, profundizando de manera gradual en los principios abductivos, inductivos y probabilísticamente deductivos que los rigen.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

CAPÍTULO 3

Inferencias explicativas como mecanismos de la justificación

Como se mencionó al comienzo de este capítulo, el objetivo principal del sistema de creencias es minimizar la incertidumbre para mejorar la supervivencia del organismo. Así, se generan conjeturas sobre eventos pasados y presentes, o predicciones futuras basadas en la evidencia disponible. Si la información procedente del medio confirma las expectativas, el proceso fluye sin sorpresas, ya que esta información quedaba explicada por la evidencia previa. En esencia, lo que el individuo conoce fundamenta estas suposiciones. No obstante, si la nueva información contradice las expectativas, sugiriendo incoherencia o inconsistencia con la evidencia activa, entonces necesita obtener una explicación. Esta discrepancia alerta al individuo de la necesidad de adquirir conocimientos adicionales para navegar su entorno con confianza.

Los individuos buscan comprender por qué sucedieron, suceden o sucederán ciertos eventos, buscando explicaciones que les permitan entender y también controlar su entorno. Para lograr esa comprensión, buscan cómo conectar coherentemente las diversas piezas informativas de las que disponen. De acuerdo con Graesser et al. (1994), las relaciones de coherencia son, básicamente, relaciones explicativas. En efecto, proporcionan las bases explicativas necesarias para la comprensión. Dentro de estas, las relaciones tópicas se especializan en transmitir justificación: aceptamos algo cuando entendemos su *‘porqué’*. Estas relaciones explicativas y justificativas toman formas: abductiva³⁰, inductiva³¹ y deductiva. Como se mencionó anteriormente, estas relaciones permiten la adhesión a nuevas creencias al ofrecer una base explicativa que satisface la búsqueda de comprensión del individuo.

³⁰ En todo caso, la abducción (y la analogía) pueden ser consideradas como especies de inducción, por lo que los mecanismos inferenciales podrían reducirse a dos: inductivos y deductivos.

³¹ Pero, como veremos, la inferencia inductiva puede cumplir varias funciones. Las funciones de construcción y justificación de hipótesis explicativas no se consideran *relaciones de explicación*. Pero su función de generalización o proyección sobre casos o datos nuevos, no observados o desconocidos sí que se consideran relaciones de explicación. Esto es, tales datos se explican por inferencia inductiva. Así pues, si bien las inferencias deductivas y todas las inductivas permiten transmitir justificación, sólo las deductivas y cierto tipo de inferencias inductivas (además de las abductivas) lo hacen explicativamente. Se desarrollará algo más este tema más tarde.

1. La explicación

Una explicación consiste en dar cuenta y razón de algo para proporcionar su comprensión (e.g. Álvarez, 2012; Díez y Moulines, 1999). La comprensión es un proceso cognitivo de alto nivel por el cual se reconoce la conexión entre varias piezas o fragmentos de conocimiento (Van Camp, 2014). Según varios autores (e.g. Friedman, 1974; Kitcher, 1989; Thagard, 2003), las explicaciones promueven la comprensión al unificar las diversas piezas informativas en un sistema de información más general que funciona a modo de marco integrador común. En este sentido, explicar consiste en subsumir una unidad informativa en un conjunto o dominio general atribuyéndole ciertas propiedades de ese conjunto. Ahora, la unidad pasa a pertenecer al conjunto general. Crucialmente, ese conjunto debe ser más conocido que la unidad en cuestión, y aceptado como válido, cierto o verdadero. Por tanto, las explicaciones acomodan la nueva información en el contexto de las creencias previas (Lombrozo, 2006).

2. La Necesidad de Explicación ante la Información Inesperada

Se requiere una explicación cuando nos encontramos con información inesperada que no encaja con nuestros conocimientos y creencias actuales. La falta de explicación implica que no entendemos algo ni tenemos razones para aceptarlo.

En este modelo, dicha información inesperada se refiere a una representación candidata N_c (o en el contenido de la oración SP, o en la atribución de P en SP). Por otro lado, ese conjunto general más conocido y aceptado mencionado más arriba consiste en el sistema de creencias y conocimientos almacenado en la memoria del individuo, o en un subdominio de tal sistema (i.e., SC o SLC, o lo que sabemos de S de la oración SP). La explicación, entonces, conlleva integrar esta nueva información de manera coherente dentro de nuestro sistema de conocimientos, creencias, o convicciones (Lombrozo, 2006) estableciendo relaciones que la hagan comprensible y justificable. Una vez integrada, la información deja de ser sorprendente y se convierte en parte de nuestro conjunto de creencias aceptadas. Esto refleja la idea de que la información ya validada raramente se cuestiona de nuevo (Schroeder et al., 2008), aunque el carácter adaptable del conocimiento humano permite su reevaluación ante nuevas evidencias o contradicciones.

La explicación surge en respuesta a información inesperada, abordando típicamente interrogantes como '¿por qué?' o '¿para qué?' respecto a su *contenido*³². Además, en el marco de este modelo, la explicación también responde al porqué de la *posesión*, de la *aceptación* o de la *preferencia* de dicha información. Es decir, aquí, la explicación responde a su justificación en un

³² En ocasiones también se considera que, además de las cuestiones mencionadas, las explicaciones responden también a las preguntas de "¿qué?" y "¿cómo?".

sentido amplio. Preguntas como '¿por qué?', '¿para qué?', '¿qué?' y '¿cómo?' son las que comúnmente inician el proceso de búsqueda de coherencia y comprensión en el discurso, según Graesser et al. (1994). Por lo tanto, la nueva información (Nc) impulsa la búsqueda de referencias dentro del sistema de creencias (SC) o un subdominio de este (SLC) que puedan explicarla adecuadamente.

3. Integración argumentativa de nuevas ideas en el sistema de creencias

En la explicación, la entrada informativa pasará a quedar integrada de forma coherente con esa información de referencia en una *estructura argumental*. Así, la información de referencia recabada pasa a funcionar a modo de razón para la justificación de la representación candidata entrante (i.e., Nc), que queda como la conclusión de esa estructura argumental.

Esta integración argumental de Nc en un dominio de referencia trae como consecuencia que, ahora, Nc puede ser inferida. En este sentido, la explicación de cierta información inesperada consiste en mostrar que existe otra información que hace *esperable* su ocurrencia o su verdad (Díez y Moulines, 1999). Si bien no es necesario que la haga *totalmente* esperable, sí que es menester que la haga *más* esperable *que* en su ausencia. Así pues, consiste en aducir premisas que muestran la representación candidata como *probable*, o más probable que en su ausencia (i.e., como vimos en la sec. 3, cap. 3, pt. I: $P(Nc) < P(Nc \mid SLC)$)³³. En ese sentido, esos contenidos previos ordenados del modo adecuado permiten ahora predecir la ocurrencia de la representación candidata. Según Álvarez (1995; 2012), cuando la unidad informativa ha quedado explicada, pasamos a estar en un estado mental diferente; uno en el que se ha incorporado la conclusión del argumento como aporte de nuestro nuevo estado de creencias. Disponer la información de referencia relevante en esta estructura argumental explicativa permite inferir información sobre cosas pasadas, así como hacer predicciones sobre ocurrencias futuras. La estructura permite que el *explanandum* o Nc se infiera del *explanans* o SLC. Esas inferencias explicativas y predictivas son las nuevas creencias que absorbe el sistema en su dinámica.

4. Relaciones tópicas y explicación inferencial

Tradicionalmente, se reconoce que la estructura argumental del *explanans* incluye dos componentes principales como premisas: *una regla o ley general* basada en regularidades bien establecidas y, si lo que se explica (i.e., el *explanandum*) es particular, unos *datos o condiciones antecedentes particulares* (e.g. Hempel, 1965). La regla general facilita la 'esperabilidad', formando así el núcleo de la explicación. Dependiendo de si el *explanandum* es general o particular, ambas premisas consistirán en leyes generales o bien contendrán condiciones antecedentes particulares.

³³ En dicho capítulo lo vimos en la forma $P(A) < P(A \mid B)$.

Estos componentes, las condiciones antecedentes y la regla general, proporcionan la *cobertura legal* que, mediante inferencia, hace esperable el contenido de la representación candidata. El esquema que se presenta en la Figura 9 pretende ilustrar la estructura argumental de la explicación.

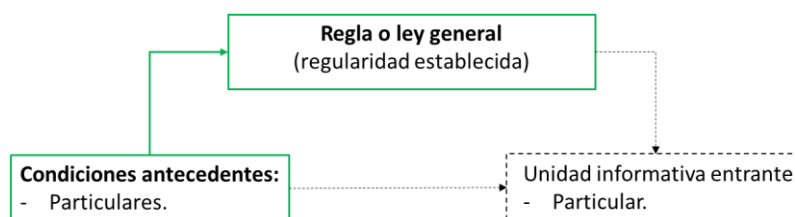
Figura 9

Estructura Argumental de la Relación Explicativa

1) Esquema de argumento explicativo con explanandum general



2) Esquema de argumento explicativo con explanandum particular



Nota: Los elementos presentados en líneas sólidas verdes se corresponden con los elementos ya aceptados que conforman la cobertura legal del *explanans* y hacen esperable al *explanandum*. El elemento presentado en líneas discontinuas se corresponde con la información del *explanandum*. Este último puede ser particular o general. Si es particular, entonces el *explanans* debe contener una condición antecedente particular. Los entimemas que conformarán los materiales experimentales de esta tesis son de este tipo.

Una diferencia del presente modelo respecto del concepto de explicación tradicional empleado en la investigación filosófica es que, en esta última, la explicación genuina es causal (e.g. Lewis, 1986; Salmon, 1984; Thagard, 2009). Sin embargo, aquí la explicación no se remite necesariamente a la historia causal del fenómeno a explicar.

Cómo se señaló más arriba, la explicación subjetiva que tiene lugar en la justificación de las creencias responde, no sólo al porqué del contenido sino también otros “porqué” de orden más pragmático. Por ejemplo, el de porqué un hablante expresa tal o cual contenido o, más relevante aún, los motivos para la *aceptación* de dicho contenido por el oyente. Por ejemplo, en los casos en los que exista una causa C que produzca simultáneamente los efectos E1 y E2, en la concepción filosófica no se podrá decir que E1 explica E2 puesto que no es su causa. Al tener una causa común, E1 correlaciona con E2, pero no lo explica. Sin embargo, en la presente propuesta, observar E1 justifica la creencia en E2; es decir, explica la aceptación de la creencia en E2 dado E1, pese a ignorar la historia causal C de ambos. En este caso, la relación tópica del signo o coocurrencia (desarrollada en el ap. C) explica por qué uno puede creer en E2 habiendo observado E1. Lo mismo se puede decir del tópico de la mayor o menor probabilidad o *a fortiori* (ver ap. C): imaginemos tres equipos de fútbol A, B y C, y que A es muy malo, y B es más o menos igual a C siendo ambos regularmente buenos y, por tanto, mejores que A. Imaginemos que ha sucedido un hecho (como condición

antecedente) en el que A ha ganado a B. Un par de semanas después, B se enfrenta a C. A partir de esto, uno podría sostener la creencia de que C ganará a B. La explicación que subyace es el hecho establecido (i.e., que A ganó a B), añadido al tópico de que “si lo menos probable ha pasado, entonces también pasará lo más probable” (como cobertura legal inferencial). Pero esta explicación, que justifica esta creencia, no es una explicación causal: que A haya ganado a B no es la causa de que C vaya a ganar a B. Y así con el resto de las relaciones tópicas. En este sentido, todas las relaciones tópicas proveen de explicación para las entradas informativas candidatas a ser o no creídas a modo de cobertura legal o esquema de justificación inferencial, siendo la relación causal sólo una de ellas.

En todo caso, es evidente que la relación causal juega un papel destacado, tanto por el lugar que ocupa en los procesos psicológicos del aprendizaje, del razonamiento y de la comprensión del discurso, como por constituir, efectivamente, la explicación genuina a nivel filosófico. De hecho, algunos estudios apuntan hacia la posibilidad de que las personas prefieren basarse en explicaciones causales sobre otros tipos de explicación a la hora de evaluar la probabilidad de que una afirmación sea verdad (Rehder, 2006; Sloman, 1994, 1997). Por eso mismo, en los experimentos que se presentan en esta tesis se ha escogido trabajar con entimemas cuyo esquema de argumentación instancia la relación tópica denominada ‘de la causa al efecto’ (también introducida más extensamente en el ap. C).

5. Características de la explicación causal

La explicación causal difiere de simplemente incrementar la probabilidad subjetiva de un evento, enfocándose en proporcionar una comprensión de su historia causal. Aunque un evento puede tener múltiples causas parciales que juntas conforman su causa total, solo se requiere información sobre algunos factores causales relevantes, determinados por el contexto de la explicación. Estas causas relevantes son seleccionadas por reglas de cobertura o patrones de regularidad que vinculan propiedades mediante leyes generales, con las causas siempre precediendo temporalmente a sus efectos en una relación asimétrica.

En la práctica, para explicar un evento particular (N_c) a través de su causa (N_j), no es necesario activar toda la red causal previa del individuo (SC), sino solo un segmento relevante (SLC) para hacer comprensible y aceptable el evento. Lombrozo (2006) indica que son las reglas de cobertura, patrones o regularidades implicadas en la explicación las que determinan qué información previa es causalmente relevante para activarse en una inferencia específica. Así, una propiedad de N_j solo será considerada como causa de N_c si el individuo dispone de conocimientos previos que sustenten esta conexión causal, evidenciando cómo la relevancia causal está condicionada al acervo de conocimientos y creencias del individuo.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

CAPÍTULO 4

Tipos de relaciones explicativas en el sistema de creencias

Exploraremos tres tipos de relaciones explicativas: abducción, inducción y deducción, cada una con un papel distinto en el presente modelo. Comenzaremos con la abducción, a la cual dedicaremos una atención mínima. La inducción, por otro lado, es fundamental para el modelo y merece una discusión ampliada. La deducción, por su parte, es la que recibirá mayor atención, dado el título de esta tesis, detallando los elementos y sus mecanismos para la credibilidad.

Como veremos, la inducción tendrá dos facetas importantes. La primera faceta no es de carácter explicativo, pero es básica para el sistema de creencias. Implica la creación de reglas hipotéticas generales y la abstracción de regularidades de experiencias para producir esquemas genéricos en la memoria a largo plazo. Aunque este aspecto podría considerarse parte del proceso de la construcción de representaciones, elegimos abordarlo aquí para tratar la inducción de forma unitaria dentro de la exposición sobre las relaciones inferenciales en el sistema de creencias. Su segunda faceta es propiamente explicativa. Veremos cómo permite dotar de explicación tanto a información ya conocida como predecir información nueva a partir de la pasada. En el primer caso, será la propia explicación la que reciba credibilidad, mientras que, en el segundo, lo será la nueva información en virtud de la explicación. Luego, nos centraremos en las relaciones deductivas, fundamentales para los entimemas en nuestros experimentos. El modelo sugiere que el sistema de creencias utiliza deducciones a partir de las representaciones inducidas, operando sobre las probabilidades que estas representaciones conllevan.

1. La abducción

La abducción es el razonamiento que genera explicaciones para lo inesperado o novedoso, infiriendo las causas de las observaciones particulares. Este proceso genera hipótesis sobre las causas que podrían explicar lo observado recurriendo a los conocimientos previos almacenados en la memoria. Por este motivo, estas explicaciones son susceptibles de ser incorrectas o refutables.

Las hipótesis sugeridas por la abducción versan sobre la relación entre eventos que ocurren juntos en circunstancias particulares. Lo sorprendente despierta interés porque lleva a cuestionar y buscar explicaciones, intentando resolver la disonancia cognitiva producida por el estado de ignorancia y alcanzar una comprensión satisfactoria (Gopnik, 2000). Se aplica en muchos casos, desde diagnósticos médicos frente a síntomas diversos, en la formación de teorías científicas que abarcan hechos dispares, hasta en situaciones diarias como, por ejemplo, interpretar huellas en la arena como indicio de presencia humana en una isla.

El patrón clásico de abducción empieza con la observación de un hecho sorprendente C. El conocimiento previo sugiere que, si A fuera cierto, C sería lo esperable, llevando a concluir A como explicación ($C; A \rightarrow C; \vdash A$). Sin embargo, múltiples hipótesis A pueden explicar C, haciendo necesario que el contexto determine la explicación más adecuada. Es una inferencia a la mejor explicación. Aliseda (2011) describe la abducción como un proceso que implica revisar y modificar creencias, siendo un tipo de razonamiento que puede ser plausible, rebatible o incluso considerado por algunos como inductivo. Este proceso genera una explicación estructurada que vincula tres componentes: la observación inesperada, su explicación abductiva como condición previa y una teoría de fondo que hace relevante esta condición previa a la observación (dada por SLC).

En este contexto, un evento dado resulta sorprendente solo en relación con las creencias previas (SLC). En el momento de la observación del evento, dichas creencias estaban configuradas de tal modo que lo hacían no esperable. Es decir, no explicable. A través de la abducción, lo que inicialmente sorprendía se normaliza al ajustar el conjunto de creencias SLC inicial, permitiendo integrar ahora al hecho observado de manera coherente.

2. La inducción

La inducción, estrechamente vinculada con la abducción, es un mecanismo inferencial que conduce a conclusiones que son plausibles, probables o susceptibles de ser cuestionadas. Ambas comparten características clave que llevan a algunos autores a considerar la abducción como un subtipo de la inducción. Una de estas características es su naturaleza no analítica, es decir, no son deductivas. Esto significa que la información obtenida en las conclusiones de estos razonamientos no está totalmente contenida en la de sus premisas, sino que va más allá. En este sentido, la conclusión puede ser más general que las premisas de las que parte. Por ello, se describen como razonamientos ampliativos, que generan nueva información. Sin embargo, esto también significa que la certeza de las premisas no asegura la de las conclusiones, sino que simplemente las hace creíbles hasta cierto punto, incluso cuando podrían ser incorrectas.

2.1. La inducción y su papel central en el sistema de creencias

La inducción frecuentemente genera conclusiones más generales que sus premisas, que suelen basarse en datos o hechos específicos. Por ejemplo, a partir de observaciones particulares, puede

inducirse una regla general, una hipótesis o un esquema causal que explique esas observaciones como efectos suyos. Este proceso permite transformar experiencias particulares en conocimientos genéricos, donde las evidencias respaldan, hasta cierto punto y con una cierta probabilidad, las hipótesis explicativas. Aunque la inducción no es un mecanismo preservador de la verdad, sí permite su transferencia parcial (Díez y Moulines, 1999). Este mecanismo juega un papel crucial en la justificación no demostrativa de muchas explicaciones y creencias, apoyándose en la probabilidad para definir su grado de certeza.

En general, las formas de justificación inductiva serían las siguientes:

- a) Inducción por enumeración: Método clásico donde, a partir de observaciones específicas finitas (por ejemplo, todos los cisnes vistos por un individuo son blancos), se generaliza una regla esquemática (los cisnes son blancos).
- b) Inducción estadístico probabilista: Basada en regularidades estadísticas (la mayoría de M son P), conduce a conclusiones ampliativas sobre que (todo) S es P porque S es M. Su estructura se asemeja a un silogismo deductivo, pero sus conclusiones son ampliativas (pues 'todo' abarca más que 'la mayoría'), tentativas y sujetas a revisión.
- c) Inducción por inferencia a la mejor explicación o abducción: Se justifican las hipótesis que mejor explican los hechos observados. Funciona operando mediante premisas implícitas sobre regularidades. Por ejemplo, al observar huellas humanas en una isla, en tanto que la mayoría de las veces que hay huellas humanas se debe a la presencia humana (premisa implícita), entonces se concluye la presencia humana como la explicación más plausible.
- d) Inducción eliminativa: Es una forma de poner a prueba una hipótesis, explicación o creencia previamente obtenida como conclusión por otro mecanismo de inducción (p.ej. la enumerativa). Consiste en buscar contraejemplos y refutadores para falsar esa conclusión hipotética explicativa. Si los casos encontrados, pese a existir, no favorecen a otra hipótesis alternativa que uno tenga en mente, entonces la hipótesis original gana en justificación y credibilidad.
- e) Inducción por métodos hipotético-deductivos: Tras formular una hipótesis, se derivan predicciones específicas. La observancia de estas predicciones en la realidad refuerza la hipótesis.

En cierto modo, se podría decir que el mecanismo inferencial inductivo juega tres papeles relevantes en las relaciones de explicación que discurren por el sistema de creencias.

1. Formación de hipótesis: a partir de una colección finita de datos particulares tomados como premisas, por inducción, se puede pasar a obtener ciertas *hipótesis generales* como conclusiones. Tales hipótesis generales pueden consistir en *regularidades o reglas esquemáticas*. Esa obtención no es lógica, sino psicológica.
2. Consolidación y actualización de hipótesis: los mecanismos inductivos contribuyen al aumento o la disminución de la justificación que se le transmite a tales hipótesis generales

en función de la observación de nuevos datos que la confirmen o la refuten. En ese sentido, esas hipótesis o esquemas serán relativamente flexibles y mutables.

3. Generalización y proyección: finalmente, las hipótesis construidas y justificadas podrán funcionar como las leyes de cobertura del *explanans*, operando como hipótesis explicativas. Es decir, permitirán dotar de explicación y justificación a nueva información entrante a la manera inductivo-probabilística expuesta en el punto b) anterior.

Desde una perspectiva filosófica, se considera que, de estos tres casos, sólo en el último la inferencia inductiva es propiamente explicativa. En los dos primeros casos, lo que se obtiene en la conclusión es la propia hipótesis explicativa, ya sea porque las premisas particulares la originan o la justifican (pudiendo también refutarla). Sin embargo, en el tercer caso lo que reside en la conclusión es un nuevo dato. Un dato que será explicado subsumiéndolo a la hipótesis general junto con alguna condición antecedente particular. En los dos primeros casos, la inferencia no es explicativa porque las premisas no explican la conclusión. No corresponde a los datos explicar la teoría, sino a la teoría explicar los datos. Del mismo modo, en un esquema causal, la hipótesis general explicativa se correspondería con el mecanismo causal, y los datos con los efectos. Las premisas consisten en datos que, si se interpretan como efectos, podrían justificar la creencia en un posible mecanismo causal (o elemento explicativo) en la conclusión³⁴. Justifican esa creencia, pero no explican su contenido, que es ampliativo. La cuestión es, nuevamente, que son las causas las que explican los efectos, y no al revés.

2.2. De la observación a la generalización: formación y justificación de creencias

2.2.1. De los datos específicos a las creencias y conocimientos generales

Uno puede albergar creencias respecto de cuestiones singulares, particulares, enmarcadas en un espacio y un tiempo concretos, relativas a la experiencia cotidiana y a las expectativas personales. Por otro lado, también puede sostener creencias acerca de cuestiones genéricas, descontextualizadas, más o menos abstractas e impersonales. Ambas clases de creencias son falibles, pese a que dudar de la experiencia personal sea menos común que dudar acerca de cuestiones genéricas. El primer tipo de creencias se almacenaría en la memoria episódica (e.g. Tulving, 1983; Tulving y Craik, 2000) y en la llamada memoria semántica personal (e.g. Grilli y Verfaellie, 2014; La Corte y Piolino, 2016; Renoult, Davidson, Palombo, Moscovitch, y Levine, 2012), mientras que el segundo tipo quedaría almacenado en la memoria semántica o social (McRae y Jones, 2013; Tulving, 1972; Van Dijk, 2014). Lo que generalmente se entiende por conocimiento haría referencia a las creencias de este segundo tipo.

Al observar un objeto individual y analizar su estructura, podemos formular una primera idea, una hipótesis esquemática sobre su concepto general. Por ejemplo, al observar un pino y notar sus

³⁴ En cualquier caso, la *creencia* en el contenido de las premisas justifica la *creencia* en el contenido de su conclusión: la primera creencia explica la segunda. Explica *porqué* creer el contenido de la conclusión; es decir, la creencia en los datos / efectos explica la credibilidad en la causa hipotética presentada en la conclusión del argumento inductivo.

características distintivas como raíces, tronco, ramas, y hojas verdes puntiagudas, se puede empezar a esbozar un concepto esquemático de ‘pino’. Sin embargo, para refinar y consolidar esta hipótesis conceptual, es necesario aumentar tanto la cantidad como la variedad de pinos observados. Este proceso de ampliación y diversificación de ejemplos mejora nuestra comprensión y justificación del concepto, permitiéndonos clasificar correctamente nuevas instancias como ‘pinos’. La transición entre la formación de la hipótesis y su justificación mediante la experiencia de más observaciones no siempre es evidente, lo que plantea dudas sobre cuándo una idea esquemática alcanza el estatus de conocimiento confirmado sobre ‘el pino’.

2.2.2. Evidencias y construcción de hipótesis

En las páginas que siguen, se hablará de algunas relaciones entre las hipótesis y las evidencias en los tres papeles de la inducción descritos más arriba. En estas coordenadas, las creencias sobre cuestiones particulares relativas a la experiencia mencionadas antes se corresponderán con las llamadas ‘evidencias’, mientras que las relativas a cuestiones generales y esquemáticas se corresponderán con las llamadas ‘hipótesis’. Así pues, en esta interpretación, las evidencias no se corresponden necesariamente con los hechos, sino con las experiencias de los hechos.

Tal y como quedó ejemplificado en la sec. 9.1., cap. 2, pt. I, tener la experiencia de observar la extracción de una bola negra es una evidencia que contribuye a hacerse una idea hipotética acerca de la composición de una urna. Los nombres de ‘evidencia’ e ‘hipótesis’ indican que lo primero se ha dado en mi experiencia mientras que lo segundo no. En ese sentido, lo primero es más creído que lo segundo, y puede ejercer de base para construirlo y justificarlo (o refutarlo) por inferencia. Sin embargo, la evidencia no está, a su vez, justificada por inferencia. En este sentido, las evidencias se identifican con las creencias básicas o fundacionales. En tanto que se emplean en la inducción para inferir hipótesis, toda hipótesis consiste entonces en una creencia derivada.

2.2.3. Calidad de la evidencia y su impacto en las hipótesis

Anteriormente, vimos que había tres tipos de creencias básicas: las intuitivas, las perceptivas y las pragmáticas. De estas, por lo menos, las dos últimas son *prima facie*, es decir, falibles. La calidad de la evidencia es variable. *Prima facie*, la credibilidad que se le otorgue a la experiencia particular será superior a la que le otorgue a un esquema o generalización con el que ésta se relacione. Su fiabilidad y su falibilidad dependerán de las condiciones de observación. En este sentido, unas evidencias de baja calidad, obtenidas en pobres condiciones de observación, podrían perder su capacidad para construir, justificar o refutar hipótesis. Por ejemplo, un ornitólogo observa unas aves de pico rojo en un paraje en el que no se habían observado antes. Esto le podría llevar a la hipótesis de que esas aves han cambiado su ruta migratoria y ahora pasan por esa zona. Pero si ha realizado la observación un día nublado y a mucha distancia, las pobres condiciones de observación disminuirán la solidez de la hipótesis.

Más falibles aún son las evidencias obtenidas por medio del discurso. Dependiendo de la honestidad y experticia que se atribuye a la fuente, sus emisiones podrán servir para construir, justificar o refutar hipótesis, o bien podrán ser ignoradas, rechazadas, o su interpretación podría ser

modificada. Por ejemplo, la hipótesis subjetiva de que un medicamento dado funciona para tratar determinado problema variará en función de que la evidencia discursiva que reciba sea la de ver muchos anuncios publicitarios de ese medicamento (baja calidad de la evidencia), o la de que muchas personas que lo hayan empleado me digan que funciona³⁵ (media-alta calidad de la evidencia). En estos casos, las evidencias no siempre influyen sobre las hipótesis, dependerá de la calidad de la evidencia y, posiblemente, de su relación respecto de la calidad de la hipótesis. Caso de interactuar, siempre, lo más cierto influirá sobre lo más incierto.

Pues bien, dado que en este punto de la exposición estamos hablando de las relaciones inferenciales inductivas, nos limitaremos a las evidencias procedentes de la experiencia directa o perceptiva que, *prima facie*, tienen más credibilidad que los esquemas hipotéticos con los que se relacionan.

2.2.4. Mecanismos inductivos en la construcción de esquemas genéricos

El conocimiento consiste en generalizaciones hipotéticas sobre patrones regulares entre elementos, que a menudo incluyen excepciones. Este conocimiento generalmente se deriva de experiencias específicas, limitadas en tiempo y espacio. Por lo tanto, se construye a partir de un conjunto finito de observaciones que, a través de la inducción, conducen a conclusiones ampliables y sujetas a cuestionamiento (Díez y Moulines, 1999; Tenenbaum, Griffiths y Kemp, 2006). Si el conocimiento es un cuerpo de contenidos justificado, entonces, de algún modo, lo particular puede justificar lo general. Sin embargo, la filosofía ha destacado lo difícil que es pasar de premisas específicas a conclusiones generales amplias usando métodos formales y apriorísticos de una pretendida lógica inductiva (Díez y Moulines, 1999; Vega, 2012). La justificación inductiva es más un proceso psicológico que lógico (Gärdenfors y Stephens, 2018).

Anteriormente mencionamos cómo se construyen marcos, guiones y esquemas genéricos a partir de modelos de las experiencias cotidianas, gracias a la inducción (e.g. Arthur, 1994; Battaglia y Pennartz, 2011). La memoria episódica, rica en detalles contextuales, almacena estas experiencias únicas en un contexto espaciotemporal específico. A menudo, muchas experiencias cotidianas comparten características comunes. Las experiencias similares van quedando relacionadas en la memoria debido a la frecuente superposición de sus componentes característicos, tales como agentes, acciones, estructura física de la situación, y otras categorías (e.g. Battaglia y Pennartz, 2011; Cohen y Eichenbaum, 1993; Eichenbaum, Dudchenko, Wood, Shapiro y Tanila, 1999). Por ejemplo, en el caso de ir a restaurantes, la mayoría de las veces habremos encontrado comensales comiendo, camareros sirviendo y el intercambio de dinero. Estas similitudes entre experiencias distintas permiten que nuestra memoria relacione y generalice, creando esquemas cognitivos compartidos.

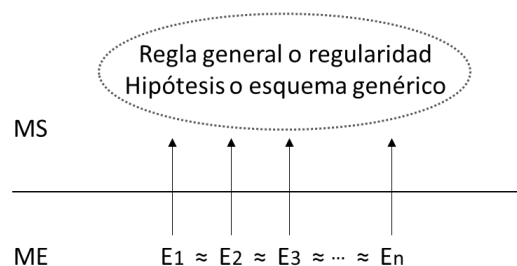
³⁵ Este es uno de los casos en los que los esquemas pueden modificar la interpretación de la información entrante. La misma información de “el medicamento X funciona” será interpretada de modo distinto si activo el esquema de “anuncio publicitario” que si no lo hago.

Cuando experiencias similares comparten componentes, se activan conjuntamente en nuestra memoria (e.g. Kumaran y McClelland, 2012), llevando a la abstracción de esos elementos comunes de sus contextos originales (Winocur et al., 2007). Este proceso de extracción de regularidades, abstracción y descontextualización culmina en la formación de esquemas genéricos, como el esquema de 'RESTAURANTE', que luego se almacenan en la memoria semántica. Allí, se consolidan y se preparan para generalizarse y aplicarse a nuevas situaciones (e.g. Sweegers et al. 2014). Estos esquemas se organizan en dominios de conocimiento más amplios, manteniendo la consistencia y la coherencia entre ellos. La memoria semántica, por tanto, actúa como un archivo de esquemas agrupados en dominios y basados en el análisis estadístico de regularidades experimentadas repetidamente (Battaglia y Pennartz, 2011), facilitando nuestra comprensión del mundo. La ilustración de la Figura 10 resume brevemente este proceso.

Figura 10

Derivación Inductiva de Hipótesis por la Experiencia

Derivación inductiva de hipótesis por la experiencia



Nota: La figura ilustra cómo experiencias similares ($E1...En$), almacenadas en la memoria episódica (ME), contribuyen al desarrollo de una hipótesis general, esquema o regla en la memoria semántica (MS). Este proceso inductivo identifica y abstrae los elementos comunes de estas experiencias, formando una estructura cohesiva de conocimiento. Estos elementos compartidos, que a menudo corresponden a categorías de situaciones típicas, se integran en una hipótesis en proceso de consolidación, indicado por una línea discontinua.

2.2.5. Evidencia, probabilidad y justificación de las representaciones hipotéticas

Según Goodman, Tenenbaum, Feldman, y Griffiths (2008) la estadística proporciona el mecanismo inductivo que mapea el paso desde los datos observados hasta las representaciones generales. Para estos autores, las representaciones genéricas consisten en reglas hipotéticas aprendidas a través de mecanismos inductivos que operan según los principios de la estadística bayesiana. La memoria almacena la información estadística (cuantitativa) acerca de las relaciones (cualitativas) entre los elementos del entorno ecológico del individuo como, por ejemplo, de las relaciones causales, de las de temporalidad, de las de semejanza o las de otras coocurrencias (Alba, Chromiak, Hasher y Attig, 1980; Battaglia y Pennartz, 2011; Cohen y Eichenbaum, 1993). Es plausible suponer que, en general, sea así con todas las relaciones que contribuyen al logro de una comprensión coherente de la realidad (Cohen y Eichenbaum, 1993) y a la transmisión de justificación entre las representaciones, permitiendo, consecuentemente, la adhesión entre las creencias.

En tanto que las relaciones mencionadas son aprendidas por inducción (Goodman et al., 2008), son probabilísticas y falibles. En este sentido, permiten formar representaciones de hipótesis generales también falibles y, por tanto, flexibles. El resultado consiste en constituir representaciones con una estructura compleja, donde sus componentes están relacionados entre sí mediante grados variables de probabilidad (Battaglia y Pennartz, 2011; Goodman et al., 2008). Esa probabilidad varía en función de las actualizaciones que se vayan produciendo sobre la estructura actual del esquema debido a las sucesivas observaciones episódicas. Esto es, en función de que la nueva evidencia confirme o refute el estado actual de la hipótesis. Esto nos conduce ya a pasar del estadio de formación de los esquemas o hipótesis, más o menos consolidados, al de su confirmación y justificación por ampliación de la muestra de evidencias o experiencias.

En esta sección, se ha explorado cómo las hipótesis y esquemas se forman principalmente mediante inducción. Sin embargo, el razonamiento abductivo y analógico también desempeñan un papel importante en este proceso, aunque algunos estudios sugieren que estos podrían ser variantes del razonamiento inductivo (e.g. Cummins, 1992; Holyoak, 2012). Además, la creatividad y la integración de conocimientos previos juegan un rol crucial en la generación de nuevas hipótesis. Este proceso es intrincadamente complejo, involucrando múltiples fuentes y mecanismos. Debido a los límites de esta tesis, este tema tan amplio no será discutido aquí.

2.2.6. Mecanismos inductivos en la confirmación y justificación de esquemas genéricos

Cuando una hipótesis o esquema está relativamente consolidado y estabilizado, la sucesiva evidencia consistente confirma y fortalece la estructura del esquema, aumentando su credibilidad. En el presente modelo, entendemos los esquemas como representaciones complejas cuyos componentes están interrelacionados mediante grados de probabilidad.

Hemos identificado que términos como “camarero”, “comensal”, “comer”, y “cobrar” son componentes clave del esquema RESTAURANTE. Este esquema, que funciona como una hipótesis, establece expectativas sobre la presencia de estos componentes en situaciones relacionadas con restaurantes, aunque admite excepciones. La probabilidad de encontrar ciertos componentes, como “comer” y “cobrar”, es alta ($P(\text{restaurante} \rightarrow \text{comer, cobrar}) = 1$). Sin embargo, otros, como “camarero”, pueden no estar siempre presentes, como en los restaurantes de autoservicio ($P(\text{restaurante} \rightarrow \text{camarero}) < 1$), lo que muestra la falibilidad de la inferencia.

El esquema implica sus componentes en cierto grado. Este esquema se fortalece o modifica a través de la acumulación de experiencias: la sucesiva coocurrencia de los componentes en diversas clases de restaurantes irá confirmando el actual esquema de RESTAURANTE. Sin embargo, cambios significativos en las experiencias, como un posible entorno en el que la mayoría de los restaurantes son de autoservicio, alterará la probabilidad asociada a ciertos componentes, produciendo así un cambio en el estado del esquema. Esto también se aplica a otros esquemas, como los causales: la expectativa de “lluvia” en un esquema de TORMENTA puede disminuir si la mayoría de las tormentas fueran secas, evidenciando la adaptabilidad de nuestros esquemas basada en la experiencia. Aunque los componentes de un esquema varían en importancia, su rol en la formación

y confirmación del esquema también difiere. A primera vista, podría parecer que la contribución del componente “camarero” es menor comparada con otros como “comer” o “pagar” dadas las probabilidades mencionadas arriba. Sin embargo, es, más bien, al revés. No solo es relevante la presencia de estos componentes dentro de un esquema, sino también su ocurrencia en contextos alternativos. Un componente será menos informativo cuanto más frecuente sea su presencia en otros esquemas del conocimiento del individuo. Aunque un restaurante pueda operar sin camareros, la presencia de “camarero” es un indicador más fuerte del esquema RESTAURANTE que “comer” o “cobrar”, debido a que estos últimos elementos son comunes a varios esquemas. Así, la presencia de camareros, siendo específica de restaurantes en contraste con otros contextos, contribuye más significativamente a la confirmación del esquema. Dicho de otro modo, los camareros están *sólo* o *típicamente* en restaurantes, pero los comensales y la acción de cobrar son compatibles múltiples esquemas o *hipótesis alternativas*.

Así pues, en la confirmación de hipótesis no sólo es relevante la probabilidad con la que se dan las evidencias dada la verdad de la hipótesis general (i.e., $P(\text{camarero} \mid \text{restaurante})$), sino también la probabilidad de la hipótesis general dada la verdad de las evidencias observadas (i.e., $P(\text{restaurante} \mid \text{camarero})$). Así pues, del mismo modo que las experiencias particulares pueden servir como premisas para dar lugar a la conformación y confirmación de hipótesis o reglas esquemáticas genéricas, también pueden dar lugar a su refutación y desaparición en su transcurso. Esto es una consecuencia del carácter no monótono de la inducción, derivado del aumento de la información observada. Actualmente, muchos psicólogos (e.g. Goodman et al., 2008; Heit y Bott, 2000; Oaksford y Chater, 2007; Tenenbaum et al., 2006) consideran que este dinamismo refleja la ejecución de unas operaciones bayesianas de actualización de probabilidades condicionadas. La probabilidad otorgada a la hipótesis varía en función de la sucesión continua de evidencia. Según Tenenbaum et al. (2006), la inducción es una forma de inferencia estadística bayesiana aplicada sobre esquemas o modelos probabilísticos estructurados del mundo. Habrá que ver, entonces, en qué consiste la justificación de hipótesis mediante la inferencia bayesiana.

2.2.7. Inducción bayesiana en la confirmación y justificación de hipótesis

Esta sección está muy relacionada con todo lo visto en el cap. 2, pt. I, en particular, a partir de la sec. 9 en adelante (i.e., probabilidad matemática y bayesiana).

Las hipótesis (i.e., h) implican la ocurrencia de unos datos o evidencias (i.e., e). Por ejemplo, si ha llovido, entonces el suelo se habrá mojado. Si la hipótesis es cierta, entonces esperaríamos observar esas evidencias (i.e., $h \rightarrow e$). Así, observar estas evidencias (suelo mojado), *confirma* la hipótesis (ha llovido), aunque no de manera definitiva, sino solo en cierto grado. El caso es que la ocurrencia de esos datos favorables no implica necesariamente la verdad de la hipótesis (i.e., $\neg (e \rightarrow h)$). Esto se debe a que otras hipótesis alternativas (i.e., $\neg h$) también podrían explicar la misma evidencia. Por ejemplo, podría haber pasado el camión de limpieza. Cuanto menos probable es la evidencia bajo otras explicaciones alternativas, más credibilidad o confirmación se le otorga a la hipótesis original. Si e sólo pudiese ocurrir por la verdad de mi hipótesis h , y resulta que ocurre e ,

entonces mi hipótesis se vería altamente confirmada y justificada³⁶. Para cuantificar el grado en que la verdad de la evidencia hace probable la verdad de la hipótesis, se recurre al concepto de *probabilidad condicionada* $P(h | e)$. De este modo podemos medir el grado en el que estamos justificados para creer en la conclusión (i.e., hipótesis inductiva) si las premisas (i.e., la evidencia) son verdaderas.

2.2.7.1. *Cómo opera el teorema de Bayes en la justificación*

Dado que una e favorable puede ocurrir tanto debido a h como a otras alternativas $\neg h$, resulta que $P(e) > P(h)$. Pero, en todo caso, dado que h implica e , la ocurrencia de una e favorable confirma en cierta proporción a h , haciendo aumentar la $P(h)$ de tal modo que $P(h | e) > P(h)$. Es decir, *ahora*, teniendo en cuenta la ocurrencia u observación de la evidencia favorable, h es más probable de lo que lo era *antes*. En este sentido, a medida que recabamos nueva información consistente con el estado actual del esquema, ese esquema se va viendo confirmado y fortalecido. Como se expuso ya en la sec. 9.2., cap. 2, pt. I, el teorema de Bayes formaliza la intuición de que es razonable que aumente el grado de creencia en una hipótesis en función de la observación de evidencia favorable. Recordemos la fórmula:

$$P(h|e) = \frac{P(e|h) \cdot P(h)}{P(e)}$$

Como se puede observar, este teorema tiene otra consecuencia satisfactoria. Dado que e podría darse tanto por h como por sus alternativas $\neg h$, se obtiene el siguiente desarrollo del denominador:

$$P(h|e) = \frac{P(e|h) \cdot P(h)}{(P(e|h) \cdot P(h)) + (P(e|\neg h) \cdot P(\neg h))}$$

Entonces, obsérvese que el numerador y el primer paréntesis del denominador son idénticos (i.e., $P(e | h) \times P(h)$), es decir, son el mismo valor. Cualquier valor dividido entre sí mismo es 1. Por tanto, ahora es fácil comprobar que, si el valor del segundo paréntesis del denominador (i.e., $P(e | \neg h) \times P(\neg h)$) es 0, entonces el resultado de la fórmula (i.e., $P(h | e)$) es 1. Y resulta que ese segundo paréntesis expresa la probabilidad conjunta de que se dé la evidencia favorable y las hipótesis alternativas³⁷. Así pues, dada la evidencia e , a medida que disminuye la probabilidad de obtenerla debido a las hipótesis alternativas $\neg h$, aumenta la probabilidad de la hipótesis target h . En ese mismo sentido, McDonald, Samuels, y Rispoli (1996) mostraron que la credibilidad de una hipótesis

³⁶ En la línea del ejemplo del restaurante, la hipótesis de que “X es un restaurante” se verá más confirmada por la observación de la evidencia “camarero” que por las de “comer” o “pagar”, puesto que estas últimas evidencias se dan también en una gran cantidad de hipótesis alternativas (p.ej. comer en casa, en el trabajo, pagar en una tienda, en un bar, etc.).

³⁷ En un esquema causal, esto equivaldría a decir que, dado un efecto, cuanto mayor sea el número de *causas alternativas* a las que acceda el individuo, menor será la probabilidad de que lo atribuya a una causa en particular. Menos necesaria será esa causa para la ocurrencia del efecto; o, lo que es lo mismo, el efecto o dato justifica tanto la causa target como las otras causas alternativas y, por tanto, a ninguna en particular. Pero, como vimos en el capítulo 4 (Parte I), esto puede verse modulado por el grado de asociación que exista entre el efecto y esas posibles causas alternativas. En este punto, son relevantes todos los hallazgos de los estudios de Markovits y colegas comentados a partir de la sec. 15, cap. 4, pt. I.

disminuía en la medida en que los participantes pudieran generar hipótesis alternativas a partir de sus conocimientos previos para la explicación de unas mismas evidencias. Una evidencia justifica una hipótesis, no sólo por derivarse de ella, sino por no derivarse de otras. De igual manera, la evidencia favorable e confirma tanto más la hipótesis h cuanto más improbable sea e si h es falsa. El que la predicción sea improbable si la hipótesis es falsa es otra condición importante para la justificación (Díez y Moulines, 1999). En una conversación cotidiana, podría expresarse como “no habríamos observado esto (e) si tal o cual cosa no fuera cierta (h)”, porque e confirma h y justifica creer en h . Esta es, a grandes rasgos, la solución bayesiana para cuantificar el grado de justificación de una hipótesis de manera inductiva, en función de la observación de nuevas evidencias.

2.2.7.2. *Determinación de probabilidades a priori*

Tal y como se expuso en el cap. 2, pt. I, el teorema de Bayes permite calcular cómo la evidencia e modifica la probabilidad anterior de la hipótesis h , pero nada hay en la teoría que permita establecer las probabilidades anteriores o *a priori* $P(h)$ y $P(e)$, necesarias para el cómputo bayesiano. Además, tal como se dijo en ese capítulo y puede verse en la fórmula, $P(e)$ depende de $P(h)$. En general, se ha optado por una de dos soluciones aparentemente distintas.

La primera de ellas consiste en aplicar el *principio de indiferencia*. Para poder hacerlo hay que identificar todas las hipótesis *posibles*, y entonces, se distribuye la probabilidad uniformemente entre todas esas posibilidades. Así, si el espacio total de hipótesis contiene dos hipótesis (como los lados de una moneda), entonces se le aplica una probabilidad *a priori* de $P(h) = 0.5$ y $P(\neg h) = 0.5$; y si fueran tres, entonces sería $P(h) = 0.33$, $P(\neg h_1) = 0.33$, $P(\neg h_2) = 0.33$, y así sucesivamente. Pero esto sólo sirve si, efectivamente, se puede determinar el número total de hipótesis posibles para poder hacer el reparto. No es este el caso en muchas de las situaciones cotidianas, donde el espacio total de hipótesis es desconocido. Por otro lado, la ley de la indiferencia se aplica por el hecho de no tener ninguna razón para favorecer una hipótesis sobre otra. Como ya se dijo en el cap. 2, pt. I, es relevante resaltar que el motivo de esta aplicación es epistémico y subjetivo: se debe a la ignorancia absoluta, es decir, a la máxima incertidumbre.

Otra solución consiste en asignar estimaciones de probabilidad a las hipótesis bajo consideración basándose en *apreciaciones subjetivas*, lo que suele ser más adecuado para situaciones cotidianas donde nuestras creencias y conocimientos previos nos dan razones iniciales para creer en algo. Aplicamos nuestros conocimientos o creencias siempre que podemos, y muchas veces no partimos de la ignorancia absoluta, sino parcial (Heit y Bott, 2000). Conforme a lo expuesto las sec. 9.2. y 11.2., cap. 2, pt. I, y a los trabajos de Goodman et al. (2007), Oaksford y Chater (2007) y Tenenbaum et al. (2006), la probabilidad *a priori* de una hipótesis ($P(h)$) refleja el grado de creencia en ella antes de considerar nuevos datos, basándose en el conocimiento previo del individuo, almacenado en la memoria a largo plazo. Esto implica que, ante la misma evidencia, distintas personas pueden asignar diferentes probabilidades *a priori* a una hipótesis debido a la variabilidad en sus sistemas de creencias y conocimientos. Así, la probabilidad *a priori* $P(h)$ no es absoluta, sino que es subjetiva y está condicionada al sistema de creencias del individuo, lo que indica su naturaleza hipotética.

Para Tenenbaum et al. (2006), la probabilidad *a priori* $P(h)$ se establece en función de una serie de principios abstractos que rigen la estructura de un *dominio de conocimiento* con el que h está relacionada. Estos dominios contienen teorías intuitivas (T) constituidas por macrounidades conceptuales junto con ciertas leyes causales y otros principios explicativos que dotan de unidad a dicho dominio (McDonald et al., 1996; Tenenbaum et al., 2006). De este modo, esa probabilidad *a priori* queda expresada como $P(h | T)$. Por otro lado, estos dominios son los que delimitan el número de posibles hipótesis que el individuo puede entrar a considerar. No cualquier hipótesis vale en cualquier dominio de conocimiento. La hipótesis debe ser consistente con los principios que rigen el dominio. Esto genera una matriz finita de posibilidades que permite llevar a cabo la distribución de probabilidades. Según Goodman et al. (2008) y Lombrozo y Rutstein (2004), en esa distribución de probabilidades, el sistema favorece de forma natural la hipótesis más simple.

Así, se hace posible la asignación variable de las probabilidades subjetivas *a priori* de h . A partir de esta distribución, ya se puede pasar a valorar el grado de justificación que nuevos datos otorgan a la hipótesis. La credibilidad de las posibles hipótesis acerca de algún dominio dependerá tanto de su adecuación a las constricciones impuestas por los principios rectores del dominio como del apoyo recibido por las evidencias disponibles (i.e., $P(h | e, T)$). De este modo, los mecanismos inductivos implicados en la confirmación de hipótesis o esquemas a partir de los datos de la experiencia derivan en cierto grado de justificación para creer en dichas hipótesis. El modelo que se presenta en esta tesis, con sus elementos de N_c , N_j , SC y SLC , incorpora los supuestos de esta explicación.

2.2.8. La ejemplificación en la justificación de hipótesis

La experiencia proporciona *ejemplos* que validan la corrección de un esquema. Una de las funciones de la ejemplificación, tanto en la argumentación dialéctica y retórica como en la cotidiana, es precisamente esa: otorgar credibilidad a hipótesis o afirmaciones generales mediante ejemplos que sirven como evidencia. Estos deben ser casos paradigmáticos, sencillos, claros, muy conocidos y estar estrechamente relacionados con el núcleo de la hipótesis que respaldan.

Ante afirmaciones generales sorprendentes, como que “viajar en avión es peligroso”, el hablante puede citar ejemplos específicos para fundamentar su argumento, tales como el accidente de Los Rodeos en 1977, por su impacto; el de Spanair en 2008, por su proximidad espacial; o el de Germanwings en 2015, por ser el más reciente. Aunque, hasta donde sabemos, no existen estudios empíricos que evalúen directamente cómo la simplicidad, familiaridad y relevancia de un ejemplo afectan la credibilidad de una hipótesis, investigaciones relacionadas han examinado cómo la diversidad y tipicidad de los ejemplos afectan la percepción de validez de las conclusiones inductivas (e.g. Osherson, Smith, Wilkie, López y Shafir, 1990; Sloman, 1993). Estos estudios sugieren que, a mayor variabilidad entre ejemplos *típicos* aducidos en las premisas, mayor será la credibilidad otorgada al alcance de la hipótesis general. Para más detalles, véanse Heit (2000), Hayes, Heit, y Swendsen, (2010) y Medin, Coley, Storms y Hayes (2003).

Los ejemplos, primero, fundamentan y después, sostienen una regularidad ya establecida pudiendo reforzar la adhesión a esa regla aumentando su presencia en la mente del individuo

(Perelman y Olbrechts-Tyteca, 1958). El rol básico de la inducción es inducir al individuo a adoptar razonablemente una creencia en base a la evidencia.

2.3. Mecanismos inductivos al emplear esquemas generales como explicación de casos particulares.

En las dos secciones anteriores, las evidencias particulares servían para construir, fundamentar y consolidar por confirmación un esquema o hipótesis general en la memoria del individuo. En este caso, sin embargo, se establece una *relación explicativa* entre dos instancias particulares mediante el recurso (normalmente implícito) a esa hipótesis general probabilística obtenida inductivamente que ahora forma parte del *explanans*.

La hipótesis general es una regularidad estadística falible de un dominio de conocimiento que aporta la cobertura legal necesaria para la explicación de un fenómeno. La estructura general interna de una hipótesis desvinculada de cualquier dominio es habitualmente la de “la mayoría de las veces que pasa A, también pasa B” o similares. Cuando quedan enmarcadas en un dominio, estas regularidades estadísticas pasan a funcionar como esquemas. Y como tales, pasan a ejercer una influencia *top-down* tanto en la interpretación de una información entrante como en las predicciones acerca de ocurrencias futuras (Sweegers, Coleman, van Poppel, Cox y Talamini, 2015; Tenenbaum et al., 2006). Dado que la experiencia pasada muestra que un caso A ha coocurrido con un caso B la mayoría de las veces, se tiende a inferir que, probablemente, hay *algo* que garantiza hasta cierto punto que nuevos casos o instancias de A vayan a coocurrir con B en el mismo dominio. Aunque no es seguro. El que varios elementos sean instancias de una misma clase implica que son, en algún aspecto, relativamente *parecidos*. Por ejemplo, uno puede haber observado que la mayoría de las instancias de una clase comparte una característica: que la mayoría de las instancias de la clase ‘perro’ comparten la característica de ladrar. Si uno establece una relación de semejanza entre una nueva entidad y las instancias de esa clase, entonces se podrá explicar o predecir que la nueva entidad también tenga o vaya a tener *probablemente* esa característica. Por otro lado, uno puede haber observado que la mayoría de las entidades que compartían cierta característica resultaron pertenecer a una misma clase. Por ejemplo, que la mayoría de las entidades que ladran resultaron pertenecer a la clase ‘perro’. Así, el que una nueva entidad se parezca a las anteriores por compartir esa característica se puede explicar suponiendo que *probablemente* pertenezca a la misma clase.

Procedamos ahora a ver un poco más de cerca la explicación por inferencia inductiva y su relación con la justificación y la credibilidad. Para ello, nos serviremos del análisis de tres ejemplos que ilustran los aspectos acabados de mencionar.

2.3.1. Ejemplos de adecuación y proyección de la explicación inductiva

En los apartados precedentes hemos visto cómo la aducción de ejemplos sirve para formar y consolidar una hipótesis y aumentar su credibilidad. Ahora, examinaremos otros ejemplos que ilustran otra función de la inducción sobre la credibilidad, esta vez, en el marco de las relaciones explicativas. Los dos primeros ejemplos pretenden mostrar cómo aumenta la credibilidad en la adecuación de una hipótesis para explicar un *hecho conocido*. Esta adecuación emerge de la

semejanza entre el hecho en cuestión y otros hechos particulares que fundamentaron la hipótesis. En el tercer y último de los ejemplos, veremos cómo las propiedades del esquema general hipotético se proyectan sobre entidades *aún no conocidas*, también por semejanza.

Para explorar los dos primeros ejemplos, debe tenerse en cuenta que una hipótesis puede estructurarse como un condicional, diferenciando entre antecedente y consecuente. Aunque ambos ejemplos aplican hipótesis causales, el primero se basa en una causalidad directa, y el segundo, en una causalidad teleológica (orientada hacia una finalidad). Además, ambos ejemplos buscan explicar por qué un sujeto exhibe cierta característica debido a su inclusión en una clase específica. Sin embargo, en el primer ejemplo, la clase de referencia forma parte del antecedente de la hipótesis, mientras que, en el segundo, se sitúa en el consecuente.

Hay que destacar que lo que gana en credibilidad no es el hecho explicado (p. ej., que β posea la propiedad P), pues ya es conocido y se acepta como verdadero. Lo que adquiere credibilidad a través de este proceso es la explicación misma de ese hecho (i.e., el *explanans* de que el hecho sea *debido a Z*). El *explanans* Z está constituido por hipótesis generales basadas en regularidades observadas y de hechos particulares adicionales. Al aceptar estos hechos particulares y las hipótesis generales, se hace esperable que β posea P . Así, se pasa a creer que, dado que Z explica adecuadamente por qué β tiene P , entonces Z debe ser cierto.

2.3.1.1. Explicación inductiva por inclusión en el antecedente

Consideremos el caso de Luís, quien ha sido diagnosticado de cáncer de pulmón. Este hecho particular nos lleva a buscar una explicación, que, para este ejemplo (tomado de Díez y Moulines, 1999), está directamente relacionada con otro hecho también particular y conocido: su historial de fumar tres paquetes diarios durante más de cuarenta años.

El *explanandum*, lo que queremos explicar, es la enfermedad de Luís. Conocemos el diagnóstico y buscamos comprender el porqué. Para hacerlo, recurrimos a su consumo intensivo de tabaco. No obstante, la conexión entre el hábito de fumar y el cáncer no es inmediata. La establecemos mediante la siguiente ley general estadística: la observación de que la gran mayoría de fumadores *así* de intensivos acaba desarrollando cáncer de pulmón. Esta regularidad, derivada de la observación de casos similares, nos proporciona una *explicación creíble*, basada en la semejanza de Luís con otros individuos con hábitos similares de tabaquismo. Aceptando esto, ahora el *explanandum* se vuelve esperable, pues se infiere del *explanans*. En terminología formal, si β es Luís, C es la propiedad de tener cáncer, y Γ designa la clase de fumadores intensos, diríamos que habiendo establecido que $\beta \in \Gamma$, $C(\beta)$ se explica por una relación causal entre Γ y C (i.e., $\Gamma \rightarrow C$). Así, la explicación adquiere credibilidad al incluir a Luís dentro de este grupo más amplio, utilizando la ley estadística general como puente para conectar su consumo de tabaco intensivo con la aparición de la enfermedad. Además, la selección de esta hipótesis sobre otras alternativas depende de su capacidad para explicar el *explanandum* de manera más coherente y completa, y aumentando su credibilidad en la medida en que se descarten otras explicaciones posibles (como, por ejemplo, que Luís trabaje en la minería). Es importante destacar que lo que gana en credibilidad no es el

explanandum (el diagnóstico de cáncer de Luís, ya conocido y aceptado), sino el *explanans*, es decir, la validez de la explicación propuesta.

Por otra parte, el razonamiento aquí es ampliativo, pues la conclusión va más allá de lo que las premisas permiten. O, dicho de otro modo, las premisas tienen menos alcance que el hecho conocido afirmado en la conclusión.

2.3.1.2. *Explicación inductiva por inclusión en el consecuente*

Este segundo ejemplo proviene de la *Retórica* de Aristóteles y muestra una estructura causal distinta a la vista anteriormente. Si el primer caso trataba sobre hechos sin una función o meta específica, este ejemplo se centrará en las intenciones y motivaciones detrás de una acción, adoptando un enfoque teleológico: se busca explicar una conducta por su propósito. En este caso, se hace un uso explícito de la ejemplificación como recurso retórico. Así dice el texto:

“...como cuando Dionisio, si pide una guardia, es porque pretende la tiranía. Porque, en efecto, como con anterioridad también Pisístrato solicitó una guardia cuando tramaba esto mismo y, después que la obtuvo, se convirtió en tirano, e igual hicieron Teágenes en Mégara y otros que se conocen, todos estos casos sirven de ejemplo en relación con Dioniso, del que todavía no se sabe si la pide por eso. Por consiguiente, todos estos casos quedan bajo la misma proposición universal de que quien pretende la tiranía, pide una guardia” (*Ret.* 1357b, 30-35).

Lo que queremos explicar (*explanandum*) es el hecho particular conocido y aceptado de que Dionisio ha pedido una guardia personal. Para explicarlo, recurrimos a una colección de otros hechos particulares, también conocidos y aceptados: que Pisístrato, Teágenes y otros también pidieron una guardia personal. En este caso, esa colección de hechos particulares constituye el grupo de referencia en el que incluir a Dionisio. A través de la inclusión de Dionisio en este grupo se llega a la explicación: quien pretende la tiranía, forma parte de este grupo. De este modo, se infiere inductivamente el hecho particular de que Dionisio pretende la tiranía como condición antecedente del *explanans*.

La ley general tomada como universal de que “quien pretende la tiranía, pide una guardia” ha sido construida inductivamente mediante casos particulares de personas que pretendían la tiranía y que, previamente, habían pedido una guardia. Esta regularidad nos proporciona una *explicación creíble*, basada en la semejanza de Dionisio con los otros ejemplares del grupo de referencia en el criterio relevante. Entonces, aceptando la ley universal (aunque construida inductivamente) y el supuesto (también inductivo) de que Dionisio pretende la tiranía, se explica por qué ha pedido una guardia personal. Nuevamente, el *explanans* vuelve esperable al *explanandum*³⁸. En terminología formal, β es Dionisio, G es la propiedad de solicitar una guardia, que coincide con la clase de referencia; a su vez, T es la consecuencia de pretender la tiranía, por lo que $T \rightarrow G$. Así, $G(\beta)$ se

³⁸ Si trasladásemos este razonamiento al ejemplo anterior, diríamos que Luís tiene cáncer de pulmón, y otros que tuvieron cáncer de pulmón eran fumadores intensos, por lo que se infiere que Luís es un fumador intenso. Pero en su caso, eso ya lo sabíamos. Por eso, una conclusión como esta no habría tenido sentido.

explica postulando que $T(\beta)$, ya que $T \rightarrow G$. Por tanto, aunque este ejemplo difiera estructuralmente del caso de Luís, también aquí la explicación gana credibilidad al incluir a Dionisio en el grupo de referencia. Esta inclusión, al conectarlo con la ley general, nos lleva a inducir el supuesto de que la intención de Dionisio es instaurar una tiranía, y esto funciona como la causa del *explanandum*. Aún más creíble puede ser si se descartan explicaciones alternativas, como que Dionisio teme ser víctima de una conspiración o un magnicidio.

Tanto en el ejemplo de Luís como el de Dionisio, la justificación de la hipótesis causal se realiza a través de la inferencia a la mejor explicación, apoyándose en la observación de regularidades estadísticas y en la exclusión de alternativas. Inferir hipótesis explicativas que se adecúen a determinados hechos dados en función de su semejanza con otros es una característica de los procesos inductivos implicados en el sistema de creencias, y está muy presente en los discursos argumentativos. Otros ejemplos claros los encontramos en la argumentación jurídica y la argumentación histórica (Collingwood, 1946; Dutilh Novaes, 2022; Lamond, 2016). Un mismo conjunto de datos o de hechos puede ser explicado por un sinnúmero de hipótesis alternativas. En ambos terrenos, se trata de construir unos modelos completos y causalmente coherentes de las situaciones planteadas a fin de lograr inducir esa creencia en los oyentes o lectores (Voss, Wiley y Sandak, 1999).

2.3.1.3. Proyección del esquema o hipótesis general sobre nuevas instancias particulares

Pasamos, finalmente, al tercer ejemplo. En los dos casos anteriores, se pretendía explicar un hecho ya conocido haciendo una inferencia sobre su historia causal (ya fuera normal o teleológica). La estructura implícita era la de “el hecho de que A sea B, probablemente se deba a la explicación Z”. Sin embargo, ahora se pretende explicar y/o predecir un hecho aún desconocido, o aún no ocurrido, o una afirmación o preferencia que requiere de justificación explicativa para ser aceptada. Ahora, la explicación inductiva sirve para justificar la creencia de que una *nueva entidad* debe poseer determinada característica. Y esto se hace recurriendo, otra vez, a la semejanza de la nueva entidad con instancias conocidas representativas del dominio del que participa la hipótesis general establecida. De este modo, la hipótesis ya aceptada se hace extensiva a la nueva entidad. Así, se explica y/o predice (la preferencia de) que tal entidad, *probablemente*, posee tal característica. En este sentido, la estructura implícita aquí sería algo como “A probablemente ‘sea o será’ B, puesto que sabemos que Z”.

En el ejemplo de Dionisio, explicábamos el hecho conocido de su solicitud de una guardia, aduciendo que tenía pretensiones tiránicas. Y para ello, nos basábamos en el esquema de que quien pretende la tiranía solicita una guardia, ejemplificado por otros casos semejantes de personas conocidas que también pidieron una guardia y acabaron instaurando una tiranía. Sin embargo, en este caso, cambia el enfoque. Si Dionisio aún no hubiese solicitado una guardia, se trataría ahora de justificar la creencia o la afirmación de que lo hará. En ese caso, emplearíamos el mismo esquema, pero deberíamos saber desde el principio que Dionisio tiene intenciones tiránicas. De este modo, a partir de saber que otros con intenciones tiránicas solicitaron una guardia, podríamos predecir o explicar la afirmación de que “Dionisio solicitará una guardia”. Podemos advertir que eso nos acerca mucho al ejemplo del cáncer de Luís. Sin embargo, para que el caso de Luís acabara de encajar con

este caso, el hecho de que “tiene cáncer” no sería un hecho conocido, sino una suposición que estaríamos tratando de justificar.

En el presente caso, lo que se pretende explicar coincide con aquello que gana en credibilidad: si antes lo que ganaba en credibilidad era el *explanans*, ahora lo es el *explanandum*. Por lo demás, al igual que en los ejemplos previos, inferimos el *explanandum* del *explanans* debido a su semejanza con otras instancias dentro de una clase. Esta clase es un componente esencial de una hipótesis, la cual se entiende como una regla, ley o esquema general en un dominio de conocimiento específico. Así pues, se podría decir que “A probablemente sea o será B, puesto que sabemos que A se parece a las instancias C en el aspecto D, y las instancias C suelen coocurrir con B”. Aplicado a los ejemplos anteriores:

- a) Dionisio probablemente haya pedido o pedirá una guardia, puesto que sabemos que Dionisio se parece a un grupo de políticos en el aspecto de pretender la tiranía, y esas instancias suelen coocurrir con la solicitud de una guardia.
- b) Luís probablemente tenga o tendrá cáncer de pulmón, puesto que sabemos que Luís se parece a un grupo de personas en el aspecto de fumar mucho, y las instancias de ese grupo suelen coocurrir con cáncer de pulmón.

Y, por poner un ejemplo distinto, con un *explanandum* deliberadamente erróneo:

- c) Los delfines probablemente son ovíparos, puesto que sabemos que los delfines se parecen a los grandes peces (como los tiburones, o los atunes) en el aspecto de vivir en el mar, ser vertebrados, tener aletas caudal, dorsal y pectoral, etc., y tales peces son ovíparos.

Volviendo a la notación anterior, igual que antes, a la hora de explicar *por qué* se dice que determinada entidad β posee determinada propiedad P , se puede recurrir a una explicación basada en la semejanza. En tanto que $\beta \approx \alpha_i$ y $\alpha_i \in \Gamma$ (donde Γ es la clase de referencia), y $\Gamma \rightarrow P$, se justifica o explica la creencia de que $P(\beta)$. Como en los casos anteriores, al mostrar la semejanza de β con α_i , se puede incluir a β en el grupo de referencia Γ . Pero en este caso, esto sirve para generalizar la regla desde los casos conocidos hacia los desconocidos. De este modo, por generalización, si la mayoría de (o todos) los miembros pertenecientes a Γ poseen la característica P , entonces se infiere inductivamente que β probablemente posee la propiedad P debido a su semejanza con α_i . Así, junto con unas condiciones iniciales, los esquemas hipotéticos explican una eventual nueva información subsumiéndola en su patrón general (Lombrozo, 2006; Lombrozo y Carey, 2006).

La relación de implicación contenida en la regla hipotética del *explanans* $\Gamma \rightarrow P$ es, más precisamente, una probabilidad condicionada $P(P \mid \Gamma)$. Además, hay otra fuente de probabilidad que depende del grado de semejanza establecido entre β y α_i que afecta al grado de pertenencia de β en Γ , repercutiendo así también en $P(P \mid \beta)$. Si la semejanza de las características relevantes entre β y α_i es total, entonces todo lo que se predique de Γ se predicará de β en ese mismo grado. Pero no siempre la semejanza es propiamente identidad. La semejanza es, también, una cuestión

de grado y permite la transferencia de propiedades en ese grado. En efecto, en este caso, la conclusión de que los delfines son ovíparos es falsa.

2.3.2. El rol de la semejanza y la tipicidad categorial en la ejemplificación

Evaluar un nuevo caso a partir de otros anteriores similares es un modo de proceder habitual en algunos sistemas de jurisprudencia (Ashley, 1990). Consiste en tratar de resolver problemas cotidianos mediante la extracción de reglas teóricas generales a partir de los casos episódicos particulares pasados similares. Del pasado se extrae una regla que se hace extensiva a un caso nuevo parecido. De este modo, la selección de unos casos previos concretos, distintivos, muy conocidos y accesibles puede guiar el razonamiento ante casos nuevos. Tales ejemplares deben estar claramente relacionados con aquello que pretenden ejemplificar, por lo que deben consistir en *miembros típicos de la clase de referencia en función de los conocimientos del individuo*. Y, en su uso retórico, además, deben ser *aceptados como tales por la comunidad epistémica* donde se emplean.

Los ejemplares típicos reúnen rasgos y propiedades compartidas con la mayoría de los otros miembros (conocidos) de la clase. De este modo, en la mente del individuo, se hacen semejantes a la mayoría de las demás instancias a la vez (Rosch y Mervis, 1975). Esto es, tienen menos rasgos distintivos entre los miembros de su clase, y a su vez, menos rasgos compartidos con miembros de otras clases (Heit, 2000). Por tanto, los ejemplares típicos son más representativos de su clase que los atípicos. Por eso, tienen más probabilidades de activar el esquema de toda su clase de referencia. En consecuencia, la tipicidad del ejemplar empleado facilita la generalización de propiedades del esquema a nuevas observaciones (Osherson et al., 1990).

Los ejemplares típicos tienen una mayor *cobertura*: en nuestra comunidad epistémica, afirmar que los osos tienen huesos sesamoideos para respaldar la idea de que todos los mamíferos, o que particulares como los caballos o los gatos, los poseen, es más eficaz que usar a las ballenas como ejemplo (Heit y Rubenstein, 1994). Estos ejemplares típicos son representativos por su similitud con el resto de los miembros conocidos de su clase, lo que los convierte en los más asociados y accesibles mentalmente dentro de la comunidad. Esta accesibilidad significa que recuperarlos de la memoria tiene un menor coste cognitivo. Dado su bajo coste de recuperación y su alta representatividad, son los ejemplares más relevantes (Medin, Coley, Storms y Hayes, 2003). Además, varios estudios han demostrado que compartir una propiedad entre dos ejemplares típicos, aunque muy diferentes dentro de su clase, facilita la generalización de esta propiedad a toda la clase (p. ej., emplear a un oso y una vaca). Esto se debe a la amplia cobertura que proporcionan estos ejemplos (Osherson et al., 1990; Rhodes, Brickman y Gelman, 2008). Así, unos pocos ejemplos cuidadosamente seleccionados pueden tener un impacto significativo en la credibilidad de la transferencia de propiedades a toda la clase.

Como se expuso anteriormente, muchas hipótesis explicativas se han formado inductivamente a partir de relaciones de semejanza entre los objetos, situaciones o fenómenos concretos. Las relaciones de semejanza no se limitan a la semejanza física. Más arriba, ya se dijo que el hecho de que varias entidades sean miembros de una misma clase implica que son, en algún sentido,

similares. Como mínimo, similares en aquello que les hace pertenecer a una misma clase. Tal y como Goodman (1972) puso de manifiesto, decir simplemente que “X es semejante a Y” es una expresión sin restricción alguna; debe completarse con la restricción “con respecto a la propiedad P”. Esta última cláusula es la que tiene todo el poder explicativo (para una discusión, ver Goldstone, 2003).

En una situación dada, la hipótesis de un esquema genérico aporta la información sobre qué aspectos son relevantes para juzgar cómo se asemejan dos entidades entre sí. Por ejemplo, los humanos pueden parecerse más a los cuervos que a las cabras en términos de inteligencia, pero más a las cabras que a los cuervos en aspectos físicos o en etapas tempranas del desarrollo.

El contexto situacional o discursivo en el que se desarrolla un razonamiento dado puede activar el dominio de conocimiento específico en función del cual juzgar la semejanza entre entidades. Dentro de ese dominio, se pueden priorizar unas creencias previas u otras que constituyan explicaciones implícitas. Esas explicaciones activadas en un dominio pueden servir de guías para determinar qué aspectos son relevantes para la generalización de propiedades entre dos entidades y cómo se deben interpretar las similitudes entre casos nuevos y conocidos. El dominio de conocimiento activado proporciona el marco conceptual que influye en cómo se lleva a cabo la inducción y la generalización de propiedades (Heit y Rubenstein, 1994; Holyoak, 1985; Lombrozo y Gwynne, 2014).

Las inferencias inductivas que se realizan están limitadas por las hipótesis genéricas disponibles, que a su vez están influidas por los principios abstractos que rigen las teorías intuitivas dentro de los dominios de conocimiento relacionados (Lombrozo, 2006; Tenenbaum et al., 2006).

Algunos estudios han revelado que cuando las personas cuentan con conocimiento previo adecuado, tienden a preferir explicaciones causales sobre la semejanza para realizar inferencias (Rheder, 2006; Sloman, 1997). Así, el conocimiento de que los delfines son mamíferos impide que se les atribuya la propiedad de poner huevos, a pesar de su apariencia similar a la de los peces. Este tipo de conocimiento, adquirido a través de la educación y la comunicación, forma una red de explicaciones causales que a menudo no guarda relación con la semejanza. Y es que, finalmente, las reglas o hipótesis generales que intervienen en las explicaciones inductivas revisten, ulteriormente, cierta causalidad. No son regularidades accidentales (Díez y Moulines, 1999).

3. La deducción

3.1. Concepto clásico de explicación deductiva

La deducción es el último de los mecanismos inferenciales que comentaremos aquí acerca del establecimiento de relaciones explicativas. Para ello, tomaremos como punto de partida el sentido clásico de explicación (científica) propuesto inicialmente por Hempel (1965) y Hempel y Oppenheim (1948).

Como se mencionó páginas atrás, se aducen explicaciones para hechos que son, en cierto sentido, sorprendentes. Por lo menos, lo son en tanto que no se dispone aún de motivos para *justificar* la creencia en su ocurrencia. La explicación de la ocurrencia de un hecho *particular* consiste en establecer o mostrar que se dan otros hechos que lo hacen esperable. Estos otros hechos actúan entonces como razones para creer en la ocurrencia del que está bajo cuestión. Y consisten, esencialmente, en unas leyes naturales universales y necesarias en sentido estricto, y en unas condiciones antecedentes constituidas por hechos particulares. De este modo, los hechos a explicar pasarán a ser esperables en tanto que *ya estén contenidos* en las leyes y condiciones antecedentes tomadas conjuntamente. El *explanandum* está contenido en el *explanans*. Si la naturaleza de la explicación es argumentativa, entonces las premisas se corresponden con el *explanans* y la conclusión con el *explanandum*. Así, si la conclusión ya está contenida en las premisas, la relación que se establece entre ambas partes del argumento es la de inferencia lógica deductiva: el *explanandum* se infiere deductivamente del *explanans*. De este modo, lo que estaba *implicado* en aquellos hechos se hace explícito, quedando ahora *explicado* por deducción. Pero este modelo de explicación no acaba de ajustarse a la realidad de las explicaciones científicas distintas de la física (Salmon, 1989) y, menos aún, a las explicaciones cotidianas que se producen en la justificación de creencias de las personas (Keil, 2006).

Muchas veces, los individuos parecen basar su comprensión de los eventos en teorías explicativas implícitas e intuitivas, regidas por unos principios abstractos difíciles de explicitar (Kozhevnikov y Hegarty 2001; Tenenbaum et al., 2006) y conectadas a unas hipótesis generales que están lejos de ser leyes universales. En muchos casos, esas leyes, reglas o regularidades no accidentales sobre las que se apoyan las explicaciones cotidianas son del tipo visto en el apartado anterior. Por tanto, son regularidades conformadas, entre otros procesos, por inducción a partir de los datos proporcionados por la experiencia ya sea directa o vicaria. Son, pues, regularidades falibles sostenidas con cierta probabilidad. Por otro lado, dadas las limitaciones epistémicas del sistema cognitivo, se puede decir algo parecido acerca de las condiciones antecedentes. La probabilidad y la incertidumbre no están únicamente presentes en esas reglas generales, sino también en los hechos particulares que permiten la aplicación de la regla al caso en cuestión.

3.2. Concepto de explicación deductiva probabilística o psicológica

3.2.1. Representaciones y probabilidad subjetiva en la explicación deductiva

El modelo que se describe en esta tesis asume que los elementos o categorías de cualquier representación compuesta o divisible se mantienen juntos mediante relaciones de probabilidad subjetiva. En este sentido, las representaciones pueden quedar expresadas como matrices direccionales de distribuciones de probabilidad condicionada entre los distintos componentes de la representación³⁹. En estas matrices, el peso asignado a una celda que relaciona los componentes x

³⁹ La idea de representación compuesta o compleja como una matriz de distribución de probabilidades subjetivas entre los distintos componentes no se desarrollará en este trabajo. Estas matrices pueden referirse

e y en la dirección P ($y \mid x$) no tiene por qué ser equivalente al peso asignado a otra celda que relaciona esos mismos componentes en la dirección P ($x \mid y$). La variabilidad en la distribución de pesos de la matriz refleja la interpretación subjetiva del individuo acerca de la representación global en cuestión. En ese mismo sentido, diferencias individuales en estos pesos conllevan diferentes interpretaciones del objeto representacional. Esto vale, entonces, tanto para la regla, hipótesis o esquema genérico como para las condiciones antecedentes relativas a la representación de los hechos particulares, e incluso para los conceptos que conforman cada uno de estos componentes.

Así pues, retomando la notación anterior, en las *condiciones antecedentes* se establece que $\beta \in \Gamma$ con cierta probabilidad subjetiva. Si todo β conocido perteneciese a la clase de referencia Γ , entonces la probabilidad subjetiva del condicional $\beta \rightarrow \Gamma$ sería máxima. Es decir, que $P(\Gamma \mid \beta) = 1$. Por el contrario, si ningún β perteneciese a la clase de referencia Γ , entonces $P(\Gamma \mid \beta) = 0^{40}$. Retomemos ahora aquel ejemplo acerca del papel de los medios de transporte en la extinción de los osos polares. La idea de que “los medios de transporte emiten gases invernadero” equivale a que el elemento “los medios de transporte” pertenece a la clase de cosas que “emiten gases invernadero” (aquí, lo trataremos como un hecho particular). De este modo, creer que todo medio de transporte pertenece a esa clase equivale a una probabilidad condicionada $P(\text{emite gases invernadero} \mid \text{medios de transporte}) = 1$. Sin embargo, si el individuo recupera refutadores de la memoria a largo plazo, como la bicicleta o el tren, entonces esa probabilidad deberá ser inferior a 1. Por otro lado, también se pueden recuperar refutadores para la *regularidad* de que “si continúa la emisión de gases invernadero (Γ), entonces se contribuirá a la extinción de los osos polares (P)”. Al recuperarse esos refutadores, esa regularidad se sostendría también con una $P(P \mid \Gamma) < 1$.

Por tanto, a las representaciones que componen el *explanans* les corresponden las probabilidades $P(\Gamma \mid \beta)$ para las condiciones antecedentes y $P(P \mid \Gamma)$ para la regularidad o regla general. Ambas son esenciales para *deducir* la probabilidad $P(P \mid \beta)$ de que “los medios de transporte contribuirán a la extinción de los osos polares”. Esto significa que *todo* lo que hay en la conclusión, tanto su contenido como su probabilidad, se obtiene a partir de operaciones deductivas en las que intervienen conjuntamente las dos premisas. De este modo, y a diferencia de lo que ocurría con la inducción en la formación y confirmación de hipótesis, las conclusiones *no son ampliativas*: no van más allá de lo que permiten las premisas. Nótese, en todo caso, que se trata de deducciones sobre materiales inciertos, siempre condicionados a los conocimientos y a las creencias de un individuo en un momento dado.

a situaciones o a esquemas. El que la matriz sea direccional se refiere a lo siguiente. Si, por ejemplo, el esquema de RESTAURANTE es una representación compuesta, las probabilidades subjetivas condicionadas entre los elementos “traer platos” (tp) y “camarero” (c) pueden cambiar en función de la dirección del condicional. Si bien el agente “camarero” siempre realiza la acción de “traer platos”, la acción de “traer platos” no siempre es realizada por el agente “camarero”. Por tanto, el peso de la probabilidad subjetiva asignado a la celda $P(tp \mid c)$ puede no ser igual al asignado a la celda $P(c \mid tp)$.

⁴⁰ Esto se verifica porque la definición de la probabilidad condicionada establece que $P(b \mid a) = ab/(ab + a-b)$; por tanto, si el valor correspondiente a ab es 0, entonces el resultado de la fórmula es también 0.

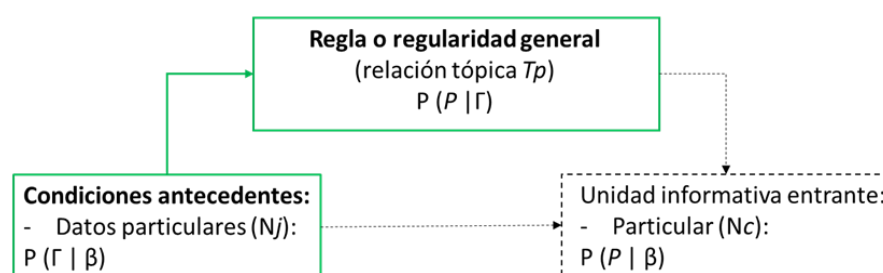
3.2.2. Nodos de información y estructura argumentativa en la explicación deductiva

Retomemos ahora los conceptos relativos a los nodos de información que intervienen en un esquema de argumentación. Las condiciones antecedentes del *explanans* se corresponden con los datos representados por el nodo justificado N_j que funciona como razón para creer en la información candidata entrante N_c . Por otro lado, la ley, regla o regularidad de la hipótesis general del *explanans* integra una relación tópica T_p regida por un criterio determinado que permite garantizar la transmisión de cierto grado de justificación desde la condición antecedente N_j hasta la información candidata entrante N_c . Esta regularidad está conectada con la teoría intuitiva implícita del dominio general de conocimiento que se activa en el foco local relevante. En función de ello, el tópico integrado en la regla o regularidad muestra la pertinencia de N_j como razón para N_c . N_j y T_p constituyen el *explanans* y N_c el *explanandum*.

Así pues, el proceso consiste en explicar una representación candidata entrante mediante otras representaciones ya aceptadas en el sistema de creencias. De este modo, el fenómeno a explicar, N_c , queda como conclusión deductiva (i.e., no ampliativa) de los datos presentados en las premisas N_j y de las reglas presentadas en las premisas T . La figura 11 representa el esquema de la explicación deductiva descrito aquí.

Figura 11

Estructura Argumental de la Relación Explicativa Deductivo-Psicológica



Nota: Este esquema ilustra los elementos discutidos anteriormente, destacando cómo las relaciones probabilísticas unen los distintos componentes de reglas y representaciones. Las líneas continuas verdes representan elementos del sistema de creencias, es decir, contenidos que ya son conocidos o aceptados. Las líneas discontinuas, por otro lado, indican la representación bajo cuestión, a la cual se transfiere cierto grado de credibilidad debido a su conexión con las creencias previas.

Desde esta perspectiva, una de las formas por las que una representación puede volverse aceptable es mediante la explicación, mostrando los datos y las reglas que conducen a ella. Las premisas explican el evento al convertirlo en esperable, permitiendo su predicción, volviéndolo probable. O por lo menos, más probable de lo que parecería sin tomarlas en consideración. En este sentido, la explicación ofrecida en las premisas modifica la probabilidad que se le otorga a la conclusión.

3.3. Dinámicas de la credibilidad y justificación en la inferencia deductiva

Muchas páginas atrás, se había representado la conclusión como una oración con un sujeto y un predicado oracionales (SP), y su credibilidad quedaba representada como $P(P | S)$. Ahora ya se

puede ver que $P(P | S) = P(P | \beta)$. Por otro lado, también se dijo que la credibilidad que se otorgue a una representación cualquiera siempre estará condicionada a la organización de las creencias y conocimientos del individuo en un momento dado. Tales conocimientos y creencias consisten en el foco local de información previa relevante para contrastar la representación bajo cuestión. Así, en un primer momento, uno podría no creer determinada información puesto que, al contrastarla con el estado actual de su sistema de creencias, no encuentra elementos que le transmitan justificación. La explicación cambia ese foco local de contraste, poniendo de relieve *otra* información también aceptada en el sistema de creencias que ahora permite predecir la unidad candidata, volviéndola más probable que antes. Anteriormente, al foco de información de contraste le habíamos llamado sistema local de creencias.

Pues bien, la probabilidad que se le otorgue a SP en un momento o tiempo t_0 no es igual a la que se le otorgue en un tiempo t_1 en el que se dispone de una explicación. En t_1 ha cambiado el SLC en función del cual se evalúa la probabilidad. Por tanto, $P_{t_0}(SP | SLC) \neq P_{t_1}(SP | SLC)$ puesto que $SLC_{t_0} \neq SLC_{t_1}$, donde $P_{t_1}(SP | SLC) = P(\textit{explanandum} | \textit{explanans})$. La deducción así planteada es compatible con la llamada inferencia dinámica que se produce en los procesos de revisión de creencias (Oaksford y Chater, 2013). Crucialmente, en este planteamiento deductivo, la probabilidad de la conclusión SP (o *explanandum*) en t_1 depende *exclusivamente* de las probabilidades de las premisas del *explanans*, y no de la probabilidad que se le otorgaba a SP en t_0 . Esto es una diferencia importante respecto del método inductivo del teorema de Bayes.

De este modo se produce la transmisión de la justificación por inferencia deductiva dada la información almacenada en la memoria a largo plazo del individuo. Tal y como afirma Lombrozo (2006), generar explicaciones de por qué una afirmación podría ser cierta proporciona un modo de evaluar la probabilidad de esa afirmación a la luz de las creencias previas. Así, si en mi sistema de creencias albergo una teoría intuitiva relacionada con la hipótesis de que “si se siguen emitiendo gases invernadero, se extinguirán los osos polares”, entonces a partir del dato (creído o aceptado) de que “los medios de transporte emiten gases invernadero” podré justificar la creencia de que “los medios de transporte contribuirán a la extinción de los osos polares”. La justificación que hay en esas creencias previas se transmite a la predicción porque ya estaba implicada en ellas. Según Moretti y Piazza (2023), estos procesos son los que permiten al razonamiento deductivo propagar la justificación de unas creencias a otras.

3.4. Simetría entre explicación y predicción

Que las premisas del *explanans* modifiquen la probabilidad otorgada al *explanandum*, volviéndolo esperable implica la existencia de cierta simetría entre explicación y predicción. De hecho, según Hempel (1965), tienen la misma estructura lógica; la única diferencia es de tipo pragmático y tiene que ver con la relación temporal entre la construcción del argumento y el evento representado por la unidad informativa entrante:

“... en un caso, se sabe que ya se ha producido el suceso descrito en la conclusión, y se buscan enunciados adecuados que expresen leyes generales y hechos particulares para

explicarlo; en otro, se dispone ya de estos enunciados y de ellos se deduce el correspondiente al suceso en cuestión antes del momento de su presunta aparición. [...Ésta es] la tesis de la identidad estructural (o simetría) de la explicación y de la predicción” (Hempel, 1965; § 2.4).

Así pues, a decir de Díez y Moulines (1999), la tesis de Hempel concibe las explicaciones como *retrodicciones*, i.e., “predicciones” de hechos conocidos; por otro lado, las predicciones, si llegan a confirmarse, son explicaciones “avanzadas”.

3.5. Comparación entre explicación inductiva y deducción probabilística

Esta es una de las características que hace que la explicación inductiva dada en el ejemplo del cáncer de Luís sea muy parecida a esta predicción deductiva dada para el ejemplo de la extinción de los osos polares. Hay que advertir, en todo caso, que la explicación deductiva basada en probabilidades no está presente en el pensamiento de Hempel. Como se verá más adelante, la explicación deductiva probabilística aquí presentada depende del desarrollo de la lógica probabilística, muy posterior a Hempel (e.g., Nilsson, 1986, 1993). De hecho, él excluyó explícitamente inferencias deductivas con conclusión particular y alguna premisa general probabilista. Para Hempel, la diferencia básica entre sus conceptos de explicación inductiva y deductiva es que, en el primer caso, el *explanandum* es altamente esperable pero no se sigue necesariamente, pues constituye una conclusión ampliativa. Sin embargo, en el segundo caso, es totalmente esperable por estar contenido de facto en las premisas. En el modelo que se presenta en esta tesis, no obstante, tanto en la inducción como en la deducción, el *explanandum* es esperable en cierto grado. La cuestión es que, mientras que en la inducción ese grado no se sigue necesariamente de las premisas, en la deducción sí lo hace.

En el siguiente apartado volveremos sobre la comparación de estos dos ejemplos. En cualquier caso, la discusión sobre las posibles diferencias entre una deducción de tipo probabilista y la inducción sigue estando abierta, por lo menos, en psicología. Unos son partidarios de mantener la distinción (e.g., Evans y Over, 2013). Sin embargo, existe la posibilidad de que la deducción se reduzca a la inducción o, incluso, de que tanto la deducción como la inducción se reduzcan a la probabilidad⁴¹. Por mi parte, me inclino a mantener esta distinción y, consecuentemente, a seguir considerando la ejemplificación como el dispositivo retórico equivalente a la inducción y al entimema como silogismo retórico correspondiente a la deducción; ambos consisten en procedimientos mediante los cuales puede operar el cambio de credibilidades. Y los experimentos que se presentan en esta tesis emplean entimemas. Por tanto, es pertinente precisar algo más la caracterización psicológica de este tipo de deducción probabilística y tratar, a su vez, de distinguirla de la inducción.

⁴¹ Aunque, tal como se estableció en la sec. 9, cap. 2, pt. I, la probabilidad también tiene una vertiente deductiva (teórica y a priori) y otra inductiva (empírica y a posteriori).

CAPÍTULO 5

Deducción en el sistema de creencias I: fundamentos probabilísticos

1. Características de la deducción en la lógica clásica

En orden a caracterizar la deducción de tipo probabilista que se asume en esta tesis habrá que partir de algunas características que tiene la deducción de la lógica clásica. En primer lugar, la lógica clásica es un sistema basado en una semántica extensional bivalente por la cual a cada variable se le asigna uno de dos valores de verdad posibles. Las sentencias que intervienen en el razonamiento lógico son verdaderas o son falsas en función de que se correspondan adecuadamente o no con aquello a lo cual se aplican. No admite una tercera posibilidad. Una proposición que es verdadera no puede ser más o menos verdadera; y tampoco puede serlo más que otra también verdadera. En segundo lugar, como ya se ha dicho en repetidas ocasiones, las conclusiones deductivas no son ampliativas. Toda la información que da la conclusión estaba ya implícitamente contenida en las premisas tomadas conjuntamente. La validez de la forma asegura esto. El argumento explicita esta información en la conclusión haciendo evidente lo que era menos evidente. En este sentido, la conclusión no dice nada realmente nuevo, no añade *otra* información. Esto trae una serie de consecuencias.

Por un lado, en la conclusión no habrá conceptos que no hayan aparecido ya en las premisas. Por otro, si las premisas eran verdaderas, dado que en la conclusión no se dice nada que no estuviera ya en las premisas, entonces la conclusión también es necesariamente verdadera. La verdad de la conclusión estaba ya implícita en la verdad de las premisas enlazadas de forma válida. Esta consecuencia está relacionada con dos características importantes de la deducción: la validez y la consistencia.

Un argumento deductivo es *válido* si y sólo si no puede ser que las premisas sean todas verdaderas y la conclusión sea falsa. Pero, como se dijo en la sec. 1.1., cap. 1, pt. I, la validez no depende del contenido del argumento, sino de su forma o estructura lógica. Esta forma se configura mediante las partículas lógicas (e.g. *y*, *o*, *no*, *si*, *entonces*, *todos*, *ninguno*, *algunos*, etc.). Cuando esta configuración resulta en esquemas deductivos válidos, estos esquemas permiten la realización de

inferencias vàlidas. Éstas consisten en una serie de operaciones mecánicas que se aplican a los términos del argumento de tal modo que, si las premisas *fueran* verdaderas, entonces la conclusión resultante también lo *sería*.

Por otro lado, tal y como se discutió en los cap. 2 y 3 de la pt. I, la *consistencia* es una propiedad que deben satisfacer los lenguajes o sistemas formales deductivos, sus fórmulas o conjuntos de fórmulas y sus cálculos derivativos. A nivel semántico, la consistencia de un argumento deductivo se cumple si, en una tabla de verdad, existe una interpretación lógica en la que todas las proposiciones del sistema son simultáneamente verdaderas. A nivel sintáctico, se satisface si no hay contradicción entre ninguna de las proposiciones del sistema; es decir, cuando no es posible deducir una fórmula y su negación a partir de las proposiciones del sistema mediante las reglas de inferencia vàlidas. Dicho de otro modo, un sistema es consistente si y sólo si no es inconsistente; y un sistema es inconsistente si alberga una contradicción, esto es, si se reduce al absurdo. De un sistema que se reduce al absurdo se sigue cualquier fórmula. Por tanto, de un sistema inconsistente se deducen todas las fórmulas. No discrimina, no distingue nada y, en consecuencia, es inútil. Esto implica que un sistema consistente es útil porque discrimina, habiendo fórmulas que no se deducen de él. Lo que no es consistente con sus fórmulas, no tiene cabida en el sistema. Así pues, si el sistema de creencias fuera inconsistente, no permitiría realizar explicaciones ni predicciones fiables (i.e., cualquier cosa se seguiría de él). No contribuiría a la reducción de la incertidumbre ni, consecuentemente, a la adaptación del organismo a su medio ambiente. Las estimaciones de probabilidad de un evento en función de nuestras creencias y conocimientos no tendrían sentido.

Así pues, puesto que el contenido factual de la conclusión está incluido en las premisas, la validez y la consistencia aseguran que no pueda suceder que las premisas sean verdaderas y la conclusión sea falsa. Los argumentos deductivos trasladan la verdad de las premisas a la conclusión. Se dice entonces que *la deducción es preservadora de la verdad*. En los argumentos deductivos, la verdad de las premisas garantiza plenamente la de la conclusión. Sin embargo, en los inductivos las premisas apoyan la conclusión sólo en cierto grado, pudiendo ser que las premisas fueran verdaderas pero la conclusión falsa. En todo caso, la verdad de las premisas hace probable la conclusión. Volveremos sobre esto de aquí a un momento.

2. Deducción probabilística para la justificación de las creencias

Todo lo anterior, en todo caso, está dicho en términos semánticos relativos a la correspondencia entre los enunciados y aquello a lo que se aplican; es decir, en términos de la *verdad* de las afirmaciones. Esta discusión se puede trasladar fácilmente al ámbito epistemológico en términos de la *justificación*. En efecto, por ‘justificación de una afirmación’ se entiende “justificación de la creencia en su verdad” (Díez, 2002). De este modo, los razonamientos deductivos transmiten entonces la justificación (de la creencia en el contenido) de las premisas a (la creencia en el contenido de) la conclusión, siendo, por tanto, *preservadores de la justificación*.

Ahora bien, no es lo mismo la justificación de una creencia que la verdad de una afirmación. Por un lado, la verdad de una afirmación no depende de ningún tipo de proceso psicológico. La justificación de una creencia, sin embargo, se da en un sujeto que cree; es el individuo quien está justificado para creer algo. Además, la justificación es algo que puede darse y, de hecho, se da en grados. Y eso no convierte a la inferencia en inductiva necesariamente, puesto que la preservación de la justificación no tiene por qué ser ampliativa.

Es concebible, como se mostrará seguidamente, que *toda* la justificación que se obtenga en una conclusión ya estuviera contenida en la justificación de sus premisas. De este modo, si la creencia en la verdad de las premisas está justificada, entonces también lo estaría la creencia en la verdad de la conclusión⁴². Llevando este supuesto al extremo, a diferencia de la inducción, *aquí no puede pasar que uno crea totalmente que las premisas son verdad sin creer totalmente que la conclusión también lo es*. En cuanto a los casos no extremos, puesto que estas credibilidades se dan en cierto grado, esta preservación de la credibilidad se debe producir proporcionalmente teniendo en cuenta esos grados, de forma consistente. Así, para la justificación de una representación que inicialmente no se cree o de la cual no se sabe nada, si una o más premisas son inciertas, entonces la conclusión también debe serlo. Los grados de credibilidad fluyen de las premisas a la conclusión. El modo en que lo hacen será razonado más adelante, y constituirá una de las principales hipótesis a comprobar experimentalmente en esta tesis.

3. Un concepto subjetivo de la deducción

Desde hace ya algo más de una década, los psicólogos del razonamiento han asumido que los procesos del razonamiento humano no son independientes del contenido sobre el que se aplican. Tratar de estudiarlo desligado del contenido produce unos resultados que más bien parecen un artefacto de la situación de laboratorio que unos datos producidos por un proceso natural de la mente humana (Evans y Thompson, 2004). Los resultados mostraban persistentemente que los individuos no suelen ajustarse a las prescripciones de la lógica clásica extensional bivalente. Eso se interpretaba como una falta de racionalidad de los seres humanos. Sólo unos pocos individuos, con altas capacidades cognitivas y en unas condiciones muy delimitadas parecían dar respuestas lógicas y, según ese criterio, racionales (e.g. Evans, Handley, Neilens, y Over, 2010; Stanovich y West, 1998). Que solo unos pocos sean racionales parece contrastar con el nivel de desarrollo científico y tecnológico alcanzado por la especie, y con muchas conductas cotidianas habituales que la mayoría de la gente realiza, como operaciones aritméticas básicas, o el mero hecho de argumentar una posición en una conversación (Oaksford y Chater, 1998, 2007).

Tal y como se expuso en apartados anteriores de este capítulo, los procesos de comprensión y de razonamiento están dirigidos a que los individuos puedan alcanzar unos objetivos que les

⁴² En el siguiente apartado se matizará esta afirmación en relación con el concepto de *grado* de justificación.

permitan la adaptación óptima a un entorno cambiante. Lo que se requiere para lograr esos objetivos en el mundo real son *grados de creencia bien justificados* y la *realización de inferencias fiables con esas creencias* (Evans y Over, 2013). Los cálculos y procesos del razonamiento están íntimamente vinculados al sistema de creencias sobre el que operan. La justificación inicial de esos grados de creencia se puede obtener, por un lado, mediante la formación de creencias fundacionales y, por otro, mediante procesos inductivos.

3.1. Grados de creencia bien justificados

Como ya quedó explicado, las creencias fundacionales del sistema son el resultado de las tres fuentes de formación de creencias. Los mecanismos por los cuales formamos creencias son altamente fiables de tal modo que han facilitado esa adaptación del individuo a su entorno. Desde una perspectiva evolucionista, este sería el motivo por el cual no han desaparecido en la selección natural.

3.1.1. Fiabilidad de las creencias fundacionales

La mayoría de nuestras creencias están adecuadamente justificadas sobre la base del correcto funcionamiento del sistema cognitivo. Sin embargo, también se mostró que ninguna fuente de formación de creencias era infalible. Si acaso, las menos susceptibles a la falibilidad parecían ser las creencias fundacionales procedentes de la intuición, como los principios de identidad o de no contradicción, o la creencia de que ‘pienso’.

Las creencias particulares acerca del mundo empírico obtenidas mediante la percepción y la sensación, pese a ser altamente fiables, eran algo más susceptibles al error. Por ‘error’ se entiende la falta de una correspondencia adecuada entre la representación mental formada en un momento dado y las características reales del objeto representado. En este sentido, las creencias fundacionales perceptivas se consideran evidencias *prima facie*, estando justificado otorgarles un máximo grado de credibilidad hasta que se demuestre lo contrario. Son falibles, derrotables o no monótonas.

Finalmente, las creencias derivadas del discurso y del testimonio del medio social eran las más falibles de las tres. Dependían de una serie de factores como un juicio subjetivo de ‘percepción’ social acerca de la honestidad y la experticia del interlocutor, de ciertos principios pragmáticos, emocionales y de la interpretación lingüística. Todos estos factores son potencialmente falibles, y pueden derivar en distorsiones relevantes de los contenidos informativos representados. Son interpretaciones personales acerca de hechos previamente interpretados por otros. La representación final de esos contenidos podría llegar a quedar muy alejada del objeto representado. Es una vía de entrada de información que no es ni racionalista (i.e., puramente racional, abstracta, general y desconectada del mundo) ni empirista (i.e., puramente perceptivo-sensitiva, concreta, particular y ligada al mundo inmediato del individuo). Sin embargo, es el medio por el cual adquirimos la mayor parte de nuestras creencias y conocimientos. Si bien en el primer caso, la confianza recae sobre la razón y en el segundo, sobre los sentidos, en el caso de la comunicación recae sobre una tercera persona. No obstante, las distorsiones graves parecen ser más la excepción

que la norma. Según Sperber (2001), de no ser así, la comunicación no se habría seleccionado naturalmente. Es por esto por lo que también conforma una vía de obtención de un gran número de creencias *prima facie* justificadas hasta que se demuestre lo contrario.

3.1.2. Formación de creencias mediante inducción

Por otro lado, hemos visto algunos aspectos de la formación de creencias esquemáticas genéricas derivadas por inferencia inductiva a partir de la acumulación de creencias fundacionales. Al abordar la inducción, en este trabajo me he centrado en las experiencias particulares de primera persona, pese a que también se pueden realizar inducciones sobre la base de textos y discursos (e.g. Landauer y Dumais, 1997); pero eso sobrepasa los límites de esta tesis. Estas creencias obtenidas por inducción constituyen parte del cuerpo de representaciones al que llamamos conocimiento genérico. Ya se comentó que, en general, en la medida en que se parte de *evidencias* para generar y confirmar *hipótesis* que las expliquen, las premisas tienen más credibilidad que las conclusiones inducidas. Y es que las premisas inductivas están conformadas por creencias fundacionales particulares procedentes de la experiencia, mientras que las conclusiones son creencias derivadas relativas a hipótesis generales.

Las generalizaciones hipotéticas que encontramos en las conclusiones inductivas van más allá de lo que pueden garantizar sus premisas. Los procesos inductivos toman una muestra de datos de la experiencia e identifican regularidades del tipo “la mayoría de las veces que pasa A, también pasa B”. Esto es, relacionan distintos elementos en función de la ratio de coocurrencias en la experiencia, entre otros parámetros. Cuando esa regularidad se ha consolidado como un patrón relativamente estable se fija como una ley o hipótesis general del tipo “(probablemente) si A, entonces B”. De este modo, pueden servir de base fiable, aunque falible, para la realización de inferencias. De hecho, muchos esquemas de argumentación son, precisamente, reglas de relación falibles constituidas sobre relaciones tópicas acerca de lo que pasa la mayoría de las veces.

En tanto que una de las formas en las que se manifiesta la inferencia inductiva es yendo de la experiencia particular al establecimiento de regularidades generales, es la encargada de proporcionar avances en el conocimiento del mundo. La inducción genera nuevas creencias en su búsqueda de la verdad (Evans y Over, 2013). Una verdad, si acaso, escurridiza, cuya búsqueda acaba derivando siempre en grados de confianza sobre la probabilidad de que la representación se corresponda efectivamente con su referente.

3.2. Evaluación de la credibilidad y consistencia en la inferencia inductiva y deductiva

El hecho de que las creencias del sistema estén sostenidas con grados de justificación, de credibilidad o de probabilidad ha sido el origen de ciertas confusiones entre la inducción y la nueva forma de entender la deducción en psicología. Como ambos mecanismos tratan ahora sobre la incertidumbre, aparentan ser muy similares. En estas coordenadas, algunos consideran que ya no son distintos y que deberían fusionarse. Esta es la perspectiva de los psicólogos bayesianos Oaksford, Chater y Hahn (e.g., Chater y Oaksford, 2000; Oaksford y Chater, 1994; Oaksford y Hahn, 2007). Según Heit (2008), esencialmente, consideran que la gente resuelve problemas deductivos

mediante procesos inductivos. Consideran que esto es una consecuencia del carácter no monótono y derrotable del razonamiento humano.

Pero no tiene por qué verse como una consecuencia necesaria de estas características. La investigación ha mostrado que el razonamiento cotidiano se realiza sobre contenidos relativos a los conocimientos del individuo. Y, muchas veces, esos conocimientos se han obtenido por inducción a partir de una muestra limitada de experiencias. En ese sentido, como mínimo, los conocimientos así obtenidos serán siempre rebatibles o derrotables y actualizables. Esto le da esa apariencia de inducción, por su componente empírico, muestral y cambiante. Pero nada de esto le hace ser necesariamente ampliativo. Se pueden hacer deducciones probabilísticas sobre los materiales obtenidos por inducción o por cualquiera de las tres fuentes de formación de creencias. Estas deducciones siempre llegan a conclusiones con cierto grado de probabilidad. Pero, desde la perspectiva del presente modelo, ese grado de probabilidad está, en cierto modo, determinado por los grados de credibilidad de las premisas.

Así pues, me inclino a sostener que existen tanto inducciones como deducciones con grados de creencias. Las primeras siempre derivarán en conclusiones con mayor incertidumbre que en las premisas debido a que la inferencia es ampliativa, mientras que eso no deberá suceder así en las deducciones (Evans y Over, 2013). El resultado de los procesos primarios de formación de creencias y de los procesos inductivos consiste en la construcción de unas representaciones que llevan adscritos unos grados de justificación variables, aunque suficientes para su aceptación en el sistema. A partir de ahí, la deducción las toma como premisas para extraer implicaciones de lo que se cree.

Si una característica de la deducción clásica es preservar la verdad cuando ya se tiene, en la deducción probabilística cotidiana será la de preservar la justificación o credibilidad con la que ya vienen las representaciones originadas por los procesos anteriores. Es decir, lo que hace la deducción es operar con el producto de esos procesos previos y con sus combinaciones. Por tanto, a partir de este punto, la verdad deja de importar. Se computan grados de credibilidad subjetivos proporcionados por la representación subjetiva de los hechos y de las hipótesis generales (Poltzer y Baratgin, 2016). En ese sentido, uno puede estar adherido a representaciones incorrectas en tanto que no se correspondan con la realidad. Y las puede tomar como premisas en un razonamiento deductivo válido. Si cree en premisas que son falsas, la credibilidad de estas premisas le justificará para creer implicaciones deductivas que también pueden ser falsas. Esa credibilidad puede propagarse de una representación a otra y generar una teoría explicativa falsa. Crucialmente, los mecanismos inferenciales pueden ser totalmente fiables y el individuo puede acabar estando fuertemente adherido a esa teoría. Probablemente, esto sea parte de lo que pasa en ciertas enfermedades mentales graves como el delirio paranoide (Evans y Over, 1996) o en la creencia en teorías conspirativas (a lo que se le añaden factores comunicativos).

En este mismo sentido, queda claro que no es necesario que los elementos del *explanans* sean verdaderos, puesto que cualquier explicación es una interpretación de los hechos. Tampoco es necesario que el *explanandum* sea verdadero, pues podría no haber ocurrido aún. Basta con creer

en el *explanans* para mostrar que una afirmación sorprendente tiene sentido. Y, de forma consistente con todo esto, si se creen premisas falsas, entonces se rechazan sus contrarias, que podrían ser verdaderas. Consecuentemente, también se rechazan las implicaciones de sus contrarias, que también podrían ser verdaderas. Los mecanismos inferenciales no están conectados con el mundo. Son reglas mecánicas de cómputo de credibilidades.

Las creencias básicas y las inducciones adquieren justificación mediante procesos directamente conectados con el mundo. Sin embargo, las propiedades deductivas de validez y consistencia no lo están. Los procesos conectados con el mundo están orientados a la producción de representaciones que reflejen cómo son las cosas, proporcionando contenidos representacionales con grados de creencia. La deducción, por su parte, está orientada a la preservación de esos grados de creencia a través de las combinaciones y transformaciones en los sucesivos pasos inferenciales. En la deducción, los contenidos son importantes en tanto que los distintos grados de credibilidad están, en un momento dado, inherentemente adscritos a ellos (Pfeifer, 2014). Pero la deducción no contempla la relación entre esos contenidos y el mundo. La deducción se limita a computar grados de creencia. Ahora bien, dado que la validez y la consistencia se aplican sobre grados de justificación en lugar de valores de verdad bivalentes, será necesario especificar cómo se definen estos conceptos ahora.

Superficialmente, se podría suponer que un argumento es válido si y solo si no permite derivar conclusiones injustificadas a partir de premisas justificadas, es decir, que la adhesión a las premisas por parte del sujeto implica su adhesión a la conclusión. De modo similar, se podría suponer que un argumento será semánticamente consistente si existe al menos una interpretación por la que todas las representaciones estén justificadas simultáneamente. O que será sintácticamente consistente si no permite sostener una representación y su negación al mismo tiempo. Sin embargo, como se discutirá más adelante, la realidad es más compleja. Esta complejidad surge porque la validez y la consistencia ahora se aplican a grados de creencia variables, es decir, a la medida en que uno está comprometido con la verdad de una representación hasta cierto grado. Antes de detallar estos conceptos de validez y consistencia aplicados a grados de justificación, es necesario explorar detalladamente las características métricas de estos grados, tratados como probabilidades subjetivas. Este análisis constituirá el enfoque de los primeros subapartados del siguiente capítulo.

3.3. Deducción cotidiana con grados de credibilidad

Por todo lo que llevamos dicho hasta aquí, se perfila con claridad la idea de que el sistema de creencias consiste en una red de representaciones sostenidas con grados de confianza o justificación variables pero suficientes para ser aceptables. Los razonamientos deductivos se aplican sobre esos grados de confianza adscritos a los contenidos. Se asume aquí que las personas hacen buenas deducciones con facilidad cuando no se les pide que supongan como verdaderas unas premisas que creen falsas o que no comprenden. Cuando el razonamiento discurre en el terreno de unas representaciones plausibles que encajan con las creencias previas de los individuos, entonces éstos

pueden realizar suposiciones con facilidad (Evans y Over, 2004; 2013; Nichols y Stich, 2003). De este modo, pueden inferir lo que es probable a partir de lo que consideran también probable.

Lo que interesa entonces es investigar cómo se produce esa transmisión de las credibilidades de las premisas a la conclusión. En particular, si parte de estas relaciones de justificación se produce mediante mecanismos deductivos, entonces habrá que suponer que la transmisión se da de forma válida y consistente. Y si esa transmisión ya no se ajusta al concepto de la preservación de los valores de verdad sino de los grados de justificación o credibilidad, habrá que dotarle de una interpretación cuantitativa. Parte del trabajo realizado a lo largo de todo el cap. 2, pt. I, fue mostrar, por un lado, que el terreno en el que se pueden mover esos grados es el de la posibilidad contingente, con sus límites en lo imposible y lo necesario; todo ello, desde una perspectiva epistémica. Y, por otro lado, que la teoría matemática de la probabilidad cuantifica ese terreno de manera exitosa. En ese sentido, el objetivo general de dicho capítulo fue mostrar que es lícito y legítimo asumir que la credibilidad, entendida como una oscilación entre certidumbre (positiva y negativa) e incertidumbre, se comporta como lo hace la probabilidad. Y, por tanto, debe respetar los axiomas de la probabilidad (vistos también en ese capítulo).

En el siguiente apartado se presentarán las características relevantes de la escala probabilística y, muy superficialmente, los conceptos de validez y consistencia probabilística. En función de esos conceptos, se pasará a explicar cómo considero que se combinan las credibilidades de las premisas en las deducciones *con entimemas*, lo cual constituirá una de las hipótesis centrales de este trabajo.

CAPÍTULO 6

Midiendo credibilidades como probabilidades

Las creencias son susceptibles de gradación o, cuando menos, de ser sostenidas con mayor o menor intensidad, y la probabilidad subjetiva es una medida de dicho grado, mide la confianza del sujeto en la ocurrencia del hecho en cuestión.

Díez y Moulines, 1999, p. 162

En la propuesta teórica de esta tesis se considera que la credibilidad es una propiedad relacional consistente en la adhesión de una representación al sistema de creencias del individuo. Esa adhesión puede variar en intensidad (Perelman y Olbrechts-Tyteca, 1958). La intensidad de adhesión es la magnitud psicológica que expresa la *cantidad de justificación* que adquiere una creencia fundacional de manera primitiva, o que se le transfiere a una creencia derivada por parte del sistema de creencias mediante una relación tópica. La intensidad de adhesión se mide aquí en una escala más amplia de probabilidad subjetiva que también abarca la disensión. En este sentido, la dinámica del continuo ‘disensión – adhesión’ se comporta respetando los axiomas de la probabilidad.

Hans Rott (2008) ha argumentado que, pese a que la flexibilidad y expresividad de la modelización probabilística de las creencias es una gran ventaja del método cuantitativo para captar la incertidumbre del razonamiento cotidiano, tiene la desventaja de que no está claro de dónde salen esos números, ni qué se supone que representan. Esta es una crítica importante a este modo de entender la credibilidad. Así pues, como se anunció en el apartado anterior, antes de entrar en el papel de la validez y de la consistencia probabilísticas en la dinámica y preservación de credibilidades, habrá que decir algo a este respecto. En primer lugar, trataré de explicar qué se supone que representan esos números en la presente propuesta teórica. Por otro lado, también comentaré la única propuesta que, por el momento, hay sobre la mesa acerca de la generación de esos números. Una propuesta originada en el seno de la teoría de los modelos mentales. Al hilo de todo esto, aprovecharé para hacer algunas consideraciones sobre varios aspectos concernientes a la escala de medición.

Los conceptos modales de imposibilidad, necesidad y posibilidad, así como los de probabilidad e incertidumbre, muchas veces se encuentran entremezclados en la bibliografía sobre las creencias. Sin duda, están íntimamente relacionados, pero no son lo mismo. A lo largo del cap. 2, pt. I, se abordaron las relaciones de la probabilidad preteórica con los términos modales, así como con la teoría matemática de la probabilidad. También se presentaron las distintas interpretaciones contemporáneas del concepto de probabilidad. Además, tanto la probabilidad preteórica como la probabilidad subjetiva contemporánea aparecieron relacionadas con el concepto de incertidumbre, pero no se entró en el detalle de esas relaciones. La incertidumbre ha recibido diversos tratamientos e interpretaciones y es importante aclarar cual se empleará aquí. Por su parte, el concepto de adhesión hizo su aparición en la sec. 4.1., cap. 1, pt. I, al presentar la Nueva Retórica de Perelman y Olbrechts-Tyteca (1958) y, desde entonces, ha ido asomando con relativa frecuencia hasta ser plenamente desarrollado en la sec. 1.3., cap. 2, pt. II. Pese a ello, aún no se han establecido unas relaciones sistemáticas entre la adhesión y estos otros conceptos. A continuación, se expondrán todas estas relaciones de manera expresa. Esto contribuirá a mostrar qué es lo que representan los números del tratamiento probabilístico adoptado en esta tesis.

1. Relaciones entre términos modales y probabilidades

La probabilidad cuantifica la posibilidad contingente, que tiene como límites naturales lo imposible y lo necesario. En el límite inferior, el de lo imposible, su probabilidad es de magnitud 0. No hay posibilidad alguna de ser. A partir de ahí, puede surgir la posibilidad contingente e ir aumentando progresivamente hasta llegar a ser, no solo muy posible, sino necesaria. En este límite superior su probabilidad es de magnitud 1. Dados estos dos valores límite, encontramos un nivel de *máxima contingencia* en su punto medio, i.e., en el 0.5. En este punto, la ocurrencia no es ni imposible ni necesaria (o tan imposible como necesaria). Así pues, la probabilidad asigna valores numéricos a todo el rango de la posibilidad contingente, donde el 0 indica una auténtica ausencia de posibilidad y el 1 una presencia total.

Trasladando esto al ámbito de la confianza o certidumbre, esa imposibilidad, posibilidad y necesidad serían declaradamente *subjetivas*. La imposibilidad se correspondería con la certeza de que algo no puede ser u ocurrir (certeza negativa), la necesidad se correspondería con la certeza de que algo tiene que ser u ocurrir (certeza positiva) y la máxima contingencia se correspondería con la *máxima incertidumbre*, donde *no se sabe* si algo puede o no ser u ocurrir. No se dispone de ninguna *razón* y por lo tanto se aplica la ley de la indiferencia (introducida en el cap. 2, pt. I). Así pues, con probabilidades de 0 y de 1 hay certeza total; la máxima incertidumbre se corresponde con una probabilidad de 0.5. Entre otras cosas, esto implica que a valores como, por ejemplo, 0.2 y 0.8 les corresponde la misma magnitud de incertidumbre. Así, en la línea de autores como Leibniz hasta Ramsey o De Finetti (ver cap. 2, pt. I), la probabilidad numérica consiste en una noción epistemológica acerca de los grados de certeza. Unos grados de certeza condicionados a las razones

de las que se disponga en un momento dado. De este modo, según todos estos autores, la probabilidad es un atributo de las creencias y de la opinión.

2. De los términos modales a las relaciones de adhesión y disensión

En relación con las credibilidades, creer que algo es imposible implica estar subjetivamente justificado para creer que algo no puede ocurrir; es decir, que se tienen razones para considerar que su *no* ocurrencia es segura. Según los conocimientos del individuo, la verdad de aquellas razones implica la falsedad de esta ocurrencia. Dicho de otro modo, esta razón y esta ocurrencia son incompatibles, inconsistentes, por lo cual se establece una relación máximamente negativa entre ambas. Se correspondería con un efecto de *máxima disensión* acerca de la posibilidad de ocurrencia del evento representado dada la consideración de verdad de las razones. En este sentido, según SLC, el predicado P es incompatible con el sujeto S. Por otro lado, creer que algo es necesario conlleva estar subjetivamente justificado mediante razones para creer que la ocurrencia positiva del evento representado es segura. Se trataría de un caso de relación máximamente positiva, correspondiéndose con un efecto de *máxima adhesión* entre el contenido representacional (SP) y el sistema de creencias del individuo (SLC), así como entre el predicado P y el sujeto S dentro de la propia representación. Finalmente, la máxima contingencia o incertidumbre se correspondería con los niveles mínimos de adhesión y disensión; sería una ausencia de relación. La posible ocurrencia del evento sería *irrelevante* en tanto que no podría vincularse ni favorable ni desfavorablemente con contenidos previos del sistema de creencias. Así pues, el efecto de máxima disensión se correspondería con una probabilidad de 0, y el de máxima adhesión, con una probabilidad de 1 y el de irrelevancia con una probabilidad de 0.5.

Por otro lado, esta distinción ha mostrado tener una realidad psicológica y neuronal (Harris et al., 2008). A nivel neuronal, los estados de disensión o incredulidad, adhesión o creencia e incertidumbre activan áreas diferenciadas del cerebro. A nivel psicológico, no solo derivan en disposiciones a conductas distintas, sino que también producen diferentes respuestas hedónicas. Aquello considerado verdadero provoca placer y es asociado con la belleza en un sentido literal. Sin embargo, aquello considerado falso provoca auténtica repulsión y desagrado.

3. Relaciones conceptuales entre escalas y la aceptación o el rechazo

En lo que resta del análisis que se presenta en este apartado general se hablará las más de las veces de probabilidad puesto que es la escala de medición que se emplea a lo largo de esta tesis y recoge, en un mismo continuo, las intensidades de disensión y adhesión. Pero conviene no olvidar que es una medición y que el aspecto interesante del que nos habla es, finalmente, de la magnitud de adhesión, esto es, de credibilidades. La siguiente tabla muestra las equivalencias entre las

magnitudes acabadas de comentar. Se asignará el valor 0 a la ausencia de magnitud, y un valor de 1 a su presencia total.

	Imposible	Contingente	Necesario
Probabilidad	0	0.5	1
Incertidumbre	0	1	0
Adhesión	-	0	1
Disensión	1	0	-
	Rechazo	Aceptación	

Se observa que los grados de probabilidad cubren todo el espectro de la progresión que va desde el rechazo hasta la aceptación de una representación mediante el empleo de valores *positivos*. Por otro lado, se observa cómo la máxima incertidumbre coincide con una probabilidad de 0.5 mientras que en los extremos no hay incertidumbre, sino certeza negativa o positiva. Además, se puede ver cómo cuando la probabilidad es de 0.5 y la incertidumbre es máxima, la adhesión es nula. Ahí no existe aún un grado de justificación satisfactorio para aceptar algo como una creencia.

La adhesión se produce con probabilidades comprendidas entre el 0.5 y el 1. Esto supone que el sistema de creencias no alberga nada con probabilidades inferiores al 0.5. Este resultado va en la línea del pensamiento de Lewis (1946) visto en la sec. 3.1.2., cap.3, pt. I. Por otro lado, el rango comprendido entre el 0.5 y el 1 se corresponde con los conceptos premodernos de lo probable, adoptable, afirmable o admisible (Schuessler, 2019) tratados en el cap. 2, pt. I. Así pues, es consistente con una larga tradición filosófica sobre las relaciones entre creencias y probabilidades. Por otro lado, el sistema de creencias está albergado en la memoria. El sistema cognitivo tiende a ser parsimonioso y no almacena información que considere falsa.

4. Adhesión y disensión reflejadas en la probabilidad y su complementaria

Podría parecer que al otorgar a una representación x una probabilidad tal que $0.5 < P(x) \leq 1$, uno está adherido también a su contradicción en una magnitud equivalente a su probabilidad complementaria (i.e., $1 - P(x)$). Por ejemplo, si sostengo la creencia particular de que “esta tarde va a llover” con un 0.7, entonces creo que “esta tarde *no* va a llover” en un 0.3. Pero los valores inferiores al 0.5 indican que se tienen razones para suponer que algo *no* es el caso, alejándose de la adhesión para adentrarse en la disensión. Por tanto, no es que esté adherido a esa representación en un 0.3; no estoy adherido en absoluto. Más bien, *disiento* hasta cierto punto de la idea de que “esta tarde *no* va a llover”. Ese 0.3 significa que acepto la *posibilidad* de que no vaya a llover, *pero no lo creo*. Desde una perspectiva disposicional se diría que no me dispondré a adoptar conductas consistentes con que “esta tarde no va a llover”. No se trata, por tanto, de creer x y $\neg x$ en diferentes

proporciones. Se trata de que, al *creer* x en una probabilidad $0.5 < P(x) \leq 1$, se *acepta la posibilidad* de $\neg x$ en una magnitud complementaria, a saber, $1 - P(x)$. Todo esto se ve aún más claro con creencias genéricas. Yo puedo creer que “si hay nubarrones, entonces llueve” en un 0.7. Su contradicción sería que “si hay nubarrones, entonces *no* llueve”, correspondiéndole una probabilidad de 0.3. Esta contradicción afirma que los nubarrones son un motivo para (creer en) la ausencia de lluvia. Es evidente que no creo eso, tal como indica esa probabilidad de 0.3. No estoy adherido a esa idea. Sin embargo, sí que lo estoy a las excepciones que causan ese 0.3: en efecto, creo totalmente que hay veces en las que, aun habiendo nubarrones, al final no llueve (y creo totalmente que esas veces son pocas). Esos refutadores sí que están albergados en mi sistema de creencias.

Finalmente, cuando la probabilidad es del 0.5 tampoco hay disensión, pero a medida que aquella descende, la disensión aumenta hasta su nivel máximo cuando la probabilidad es 0, puesto que se considera que un evento es imposible y se rechaza totalmente. Se tiene la certeza subjetiva de que no puede ser. Nada acerca de lo cual disentimos quedaría albergado en el sistema de creencias, pues son cosas que *no* se creen. Es cierto que para aquello a lo que asignemos probabilidades comprendidas en $0 < P(x) \leq 0.5$ admitimos la posibilidad de su ocurrencia. Pero, como se razonó arriba, no lo consideramos probable en su sentido preteórico: no creemos que vaya a ocurrir. Cuando uno cree que “no- p ”, ese “no- p ” no está albergado como tal en su sistema de creencias y/o conocimientos. En tal sistema, si acaso, podrá tener almacenado el concepto p , pero la adhesión a su negación es algo que se computa on-line cuando lo demande el contexto, en la forma de una creencia disposicional.

5. Cantidades que representan cualidades

Una crítica de Johnson-Laird (1994) a la concepción probabilística fue la de que no queda claro en qué sentido es psicológica, pues la gente no otorga mentalmente valores numéricos a la fiabilidad que le merecen sus pensamientos. De ser así, añade, la gente no ilustrada en números no podría representarse las relaciones de fuerza de las representaciones mentales. Para el autor, el nivel de competencia que una teoría psicológica debería aspirar a explicar es el de la probabilidad de las creencias entendida en su sentido premoderno y cualitativo.

Sin embargo, más que una crítica a la concepción probabilística en general, parece ser una crítica al supuesto de que las creencias o su intensidad se representen como números a nivel psicológico. No sé si hay alguna teoría que sostenga eso, pero, desde luego, el presente modelo no lo hace. Medir la credibilidad como probabilidades no implica asumir que las credibilidades sean números. Cualquier medición implica la cuantificación numérica de alguna magnitud, esto es, de alguna propiedad cualitativa del objeto que se dé en grados. Sin embargo, ello no supone que deba reducirse el objeto ni la propiedad medida al número, ni tampoco a su unidad de medida. Los

objetos tienen propiedades cualitativas, algunas de las cuales se dan en grados. Esos grados son cantidades o magnitudes mensurables. Los números representan esas magnitudes.

La cuantificación numérica es un modo de *expresar* la magnitud de manera precisa y ordenada, permitiendo realizar comparaciones entre, por ejemplo, la credibilidad relativa a dos representaciones distintas, o a la misma representación en dos momentos distintos, así como realizar operaciones como la adición, la sustracción, la multiplicación o la división de tales magnitudes. Permite *combinarlas* de un modo informativo y sistemático. Pero lo que hay debajo de los números es cualitativo. La adhesión o la disensión son propiedades cualitativas esenciales a la credibilidad. En la cuantificación, *esas propiedades empíricas quedan representadas por números reales* que indican el grado en el que los objetos las exhiben. Los números se asignan sobre la base de una correspondencia o similitud entre la estructura del sistema numérico y la estructura del sistema cualitativo. De este modo, aún sin ser números, se puede operar con las magnitudes de las propiedades como si lo fueran. Así, el relato cuantitativo de las probabilidades parece el adecuado para ofrecer una *descripción* de los cómputos cognitivos de credibilidad. Sin embargo, el nivel correspondiente a la *explicación* psicológica de la dinámica de credibilidades se encuentra en el terreno cualitativo en el sentido apuntado por Johnson-Laird (1994). El nivel explicativo debe incluir al descriptivo. En este sentido, la concepción premoderna y moderna de la probabilidad se hacen compatibles en la interpretación subjetivista de la probabilidad.

La metrización de las credibilidades en probabilidades puede resultar de gran ayuda para la descripción de los cómputos dinámicos implicados en su actualización, ya que permite ordenar, comparar y, crucialmente, combinar los grados de credibilidad según unas leyes sobre su comportamiento. En cualquier caso, pese a ello, su unidad de medida requiere todavía de una mayor especificación. No hay dudas sobre la correspondencia numérica de los grados de probabilidad subjetiva con sus extremos y el medio de la escala. Pero no está claro que esos grados de probabilidad, entendidos como su unidad de medida, sean constantes y uniformes entre el 0 y el 0.5, ni entre el 0.5 y el 1. Es decir, podría suceder que la distancia entre 0.2 y 0.3 no fuera la misma que entre 0.7 y 0.8. Por otro lado, también podría ser que una probabilidad de, digamos, 0.2 no significara lo mismo para distintos individuos. Ahora bien, si 0, 0.5 y 1 significan lo mismo para todos, no parece haber razones para suponer que deba ser diferente con los otros valores. Pero hay que admitir que es posible que el “grado” de probabilidad subjetiva no esté aun claramente delimitado. A diferencia de lo que ocurre con la masa o la temperatura, no hay un estándar externo u objetivo que lo determine para poder establecer comparaciones precisas. Por lo menos, más allá de los extremos y el punto medio (i.e., 0, 0.5 y 1). Así pues, es un sistema comparativo al que aún podría faltarle una mayor precisión. Saber si es lícito exigirla o no, dado su carácter subjetivo, es otra cuestión que no será abordada aquí.

Por último, tampoco hay datos que muestren si los individuos son consistentes o no con las asignaciones que hacen en la escala. Podría ser que las mismas medidas sobre unos mismos ítems con unos mismos sujetos tomadas en dos momentos extremadamente cercanos en el tiempo

mostrasen variaciones. Podría ser que *etiquetasen* de forma ligeramente distinta una misma magnitud. O bien, que cambiasen la magnitud de un mismo ítem por entrar a considerar un dominio de conocimiento distinto en el segundo momento. No está claro hasta qué punto las personas tienen un acceso directo y fiable a sus conocimientos y creencias reflexivas (Van Dijk, 2014). En este sentido, es posible que la propia asignación de probabilidades conlleve incertidumbre, por lo que, en lugar de asignar un valor debería asignarse un intervalo.

6. Origen de las estimaciones de probabilidad: una propuesta sobre el paso de lo cualitativo a lo cuantitativo⁴³

La probabilidad que interesa aquí es la subjetiva. En los experimentos que se han realizado desde esta perspectiva se ha solido pedir estimaciones directas de probabilidad a los participantes acerca del contenido de los ítems experimentales (e.g., Evans, Thompson, Over, 2015; Over, Hadjichristidis, Evans, Handley, y Sloman, 2007; Khemlani, Lotstein, y Johnson-Laird, 2012). Los participantes no tienen muchos problemas para hacer tales estimaciones, independientemente de que las situaciones que se les planteen se presten a un análisis de frecuencias (p.ej., “si hay nubarrones, entonces llueve”) o que no se presten a ello (p.ej., “la empresa Apple seguirá siendo rentable en los próximos años”). En el primer caso, esas estimaciones pueden surgir de un proceso de calibración donde el sistema cognitivo ha aprendido implícitamente las frecuencias relativas de coocurrencias y las ha almacenado en forma de esquema en la memoria a largo plazo. Pero cabe preguntarse, como hacen algunos autores (e.g., Khemlani, et al., 2012; Khemlani, Lotstein, y Johnson-Laird, 2015), de dónde surgen las estimaciones para los otros tipos de casos. Según Johnson-Laird, Khemlani, y Goodwin (2015), las estimaciones de probabilidad para eventos únicos, ulteriormente, pueden estar basadas en esquemas que han sido construidos también por frecuencias. Por ejemplo, que la mayoría de las veces las empresas con *x*, *y* o *z* características (compartidas por la empresa Apple) continúan siendo rentables de un año para el otro. Se asignan mediante la activación de determinados dominios de conocimiento que contienen teorías intuitivas *T* regidos por ciertos principios explicativos abstractos.

Pero, de forma más relevante, la pregunta que se plantean estos autores es la de ‘*de dónde salen esos números*’. Como he comentado antes, Johnson-Laird (1994) apuntaba que las personas no asignan mentalmente valores numéricos de fiabilidad a sus pensamientos, ni siquiera de manera implícita; del mismo modo, decía que esto sería un impedimento para que las personas sin conocimientos numéricos pudiesen creer cosas con más o menos fuerza. A esto, añadía, ahora junto a Khemlani et al. (2015), que el nivel de competencia que la psicología del razonamiento cotidiano debe explicar es el de la probabilidad preteórica de sentido común desarrollada en la *Retórica* y en

⁴³ Dejo fuera del siguiente análisis la asignación numérica de credibilidad para *creencias fundacionales*. En particular, para las creencias empíricas procedentes de la experiencia personal singular. En cualquier caso, merece la pena apuntar que consistiría en una asignación directa de probabilidad 1 *prima facie*.

los *Analíticos de Aristóteles* (sec. 2 y 3, cap.2, pt. I). Un nivel cualitativo que se expresa, según ellos, en términos ordinales de cualificadores modales de probabilidad (tal y como también propuso Toulmin, 1958).

6.1. De lo cualitativo a lo cuantitativo: propuesta de los Modelos Mentales

En particular, Khemlani et al. (2015) proponen que las intuiciones de las personas sobre probabilidades reflejan la idea de que los eventos varían de lo imposible a lo necesario, con varios casos intermedios. Según ellos, tales grados pueden expresarse informalmente en una escala ordinal mediante los ‘términos aristotélicos’ de *imposible*, *casi imposible*, *muy improbable*, *poco probable*, *tan probable como improbable*, *probable*, *muy probable*, *casi seguro* y *cierto*. Según Khemlani et al. (2015), la granularidad con que la gente se representa los niveles (o grados) de esta escala es relativamente gruesa y variable entre sujetos, aunque no lo demuestran. Escogen estos ocho cualificadores como una aproximación satisfactoria.

Proponen un algoritmo en términos de modelos mentales para explicar el proceso que media entre esta información cualitativa y las estimaciones numéricas que dan los participantes de los experimentos. Para estimar la probabilidad de una aserción, el sistema cognitivo activa los conocimientos acerca del dominio relevante, lo simula en modelos mentales y emplea esa simulación de posibilidades para formar una representación icónica, analógica y primitiva de la magnitud, al modo en que lo hacen los niños y las personas sin formación numérica. Los iconos que plantean Khemlani et al. son lineales, con lo imposible y lo necesario como extremos o topes fijos y una línea discontinua entre ellos que indica los grados discretos de credibilidad o probabilidad subjetiva. Esa línea discontinua se presenta con un número de puntos variable según la magnitud a representar, pudiendo llegar hasta el extremo de la necesidad (8) o no llegar a él (< 8). Estos iconos no son numéricos y representan magnitudes por analogía directa con el tamaño de la línea discontinua: a mayor magnitud, más puntos en la línea discontinua hasta llegar a la consideración de necesidad. Esos iconos lineales *pueden combinarse* para realizar *operaciones de razonamiento deductivo con las magnitudes representadas*. Esas operaciones serían análogas no numéricas al promedio, la adición y la multiplicación; serían muy primitivas y cuasi-aritméticas. El resultado de tales operaciones es, nuevamente, un icono que representa una probabilidad intuitiva. La cuestión crucial es que esos iconos pueden ser entonces descritos verbalmente empleando descripciones cualitativas con facilidad, usando los términos aristotélicos; por otro lado, con algo más de esfuerzo, los individuos con formación numérica también pueden expresarlos ya numéricamente. Esta explicación es, en general, compatible con las asunciones que se adoptan en el modelo teórico que se describe en esta tesis.

6.2. Integración de la cuantificación de los modelos mentales en nuestro enfoque

En el modelo que presentamos, asumimos que tanto el acceso a los nodos informativos de teorías explicativas implícitas almacenadas en la memoria a largo plazo, como los procesos de simulación mental de posibilidades, son relevantes para comprender y razonar sobre contenidos cotidianos. Durante esos procesos se accede, aunque sea implícitamente, a cierto cuerpo de respaldos y de

refutadores en SLC (e.g. Cummins, 1995; Cummins, Lubart, Alksnis, y Rist, 1991; De Neys 2010; Geiger y Oberauer, 2007). Ello fija la coherencia con el sistema general de creencias y modula la magnitud de la probabilidad subjetiva otorgada a la situación representada.

En ese sentido, parece bastante plausible que, como producto de estos procesos, los individuos se representen las magnitudes de probabilidad de un modo no numérico parecido al que describe la teoría de los modelos mentales. Asimismo, también parece razonable que puedan traducir estas intuiciones a un formato verbal explícito en términos de cualificadores modales ordinales, y que las personas conocedoras de los números puedan, a su vez, asignarles valores concretos, ahora, *en una escala métrica de razón*, donde el 0 indica ausencia de magnitud. En cualquier caso, el hecho de que las personas no versadas en números no puedan realizar conscientemente tales asignaciones numéricas no implica que no le sean aplicables a esas magnitudes; tampoco implica que la escala deba ser ordinal. De hecho, la posibilidad de realizar operaciones de combinación de esas magnitudes implica que no puede tratarse de una escala meramente ordinal. Por otro lado, el supuesto de que las operaciones deductivas con esas magnitudes se comporten como probabilidades numéricas no significa que los individuos operen mentalmente con números en la dinámica de sus creencias. Tampoco lo hace el agua de un recipiente cuando, estando un litro a 40 ° C, se combina con otro litro que está a 20 ° C y queda finalmente a 30 ° C. La operación responde en este ejemplo a una media aritmética, pero es evidente que el agua no lo ha calculado. La probabilidad es a las creencias como la teoría fractal a las hojas de un helecho.

7. Sobre la naturaleza instrumental de la escala de medición

Cuántos niveles deba contener dicha escala es otra cuestión que, pese a su relevancia, no será abordada en esta tesis. En todo caso, cabe hacer algunos comentarios. Khemlani et al. (2012, 2015) arguyen que la representación mental de la escala no deberá ser continua, sino discreta. Lo continuo es infinito y, por tanto, no es computable. Así, la teoría de los modelos mentales propone estos ocho niveles listados más arriba como una aproximación razonable.

Por otro lado, es posible que en un proceso natural de asignación explícita de probabilidades sea necesario acceder a la memoria de trabajo. Si fuera así, la escala debería atenerse a las diferencias individuales en las limitaciones de la memoria de trabajo. En ese caso, podría ajustarse al 7 ± 2 (Miller, 1956) oscilando, entonces, entre los cinco y los nueve niveles (Van Dijk, comunicación personal). Existen algunos trabajos (e.g. Ferrando, 2003a) que han estudiado cómo las personas distribuyen sus respuestas en escalas gráficas continuas, situando su respuesta a lo largo de una línea recta horizontal comprendida entre dos polos. Estos estudios han mostrado que, pese a la gran capacidad expresiva de la escala debido a su naturaleza continua, las respuestas de los participantes tendían a agruparse en un número limitado de puntos, colapsando en siete niveles. Parece, pues, que hay indicios de la existencia de una escala natural a nivel psicológico (Miller, 1956; Saaty y Ozdemir, 2003). Es decir, puede que haya un número correcto de particiones a realizar en el

continuo que va de la consideración de imposibilidad a la de necesidad. Si eso fuera así, podría argumentarse que la medición óptima debería establecerse sobre la base de esos siete puntos. Por otra parte, el sistema decimal que se emplea habitualmente es una convención basada en el número de dedos que tenemos para contar, y han existido otros sistemas en otros ámbitos, culturas y tiempos. En ese sentido, podría parecer que la medición de las probabilidades subjetivas sobre una base decimal no acabe de ser lo que mejor se ajuste a la naturaleza del objeto medido.

Pero, salvo excepciones, la regla de medir no suele ser intrínseca al objeto medido; se aplica desde fuera. A veces, es conveniente separar la medición del objeto medido. Gran parte de la matemática actual está basada en el sistema decimal, y dentro de tal parte, la probabilidad. En este sentido, los cálculos probabilísticos dependen, hasta cierto punto al menos, del número 0 y del número 1. Aquí se asume que las operaciones deductivas que se lleven a cabo sobre las magnitudes de credibilidad se atienen a los axiomas de la teoría de la probabilidad. Los cálculos con credibilidades se comportan como cálculos probabilísticos. En estas operaciones, no significa lo mismo el resultado de, por ejemplo, multiplicar cualquier número por 1 que por 7. Por otro lado, imaginemos que el número de particiones correctas fuera siete. En ese supuesto, es cierto que medir esa gradación con una escala inferior a siete podría resultar en un empobrecimiento de la información. Sin embargo, no parece que medirla con una escala superior deba resultar problemático. Cuanto mayor sea la escala, mayor capacidad expresiva, más fino será el grano de la información recogida. Así, aun admitiendo la posibilidad de que la medición de la credibilidad en una escala de once puntos (del 0 al 1) pueda no ser lo más descriptivo del objeto, considero que es adecuado en un sentido instrumental; permite que las respuestas puedan distribuirse en 9, 7 o 5 puntos sin imponer ninguna restricción. También lo sería en una escala gráfica continua.

CAPÍTULO 7

Deducción en el sistema de creencias II: validez y consistencia probabilísticas

A lo largo del cap. 5, pt. II se especificaron algunos motivos para sostener la distinción entre los procesos de razonamiento deductivo e inductivo, pese a que ahora ambos traten con grados de justificación. Uno de ellos consistía en que la deducción no va más allá de las premisas, no es ampliativa, mientras que la inducción sí lo es. En la deducción, el grado de justificación de las premisas garantiza plenamente el grado de justificación de la conclusión. Si uno cree totalmente las premisas de un argumento deductivo, entonces también creerá totalmente la conclusión. Nada de esto sucede así en la inducción. Se propuso que la forma más adecuada de caracterizar la validez deductiva en el razonamiento cotidiano (y, en particular, con entimemas) era en términos de la *preservación de los grados de justificación* a través del paso inferencial de las premisas a la conclusión. En el siguiente capítulo (cap. 6, pt. II) se razonó lo adecuado de medir la justificación en términos de probabilidades subjetivas, cuyos grados oscilan entre el 0 y el 1. La deducción opera ahora, no con verdades y falsedades, sino con credibilidades. Así, el razonamiento deductivo cotidiano consiste en combinar las credibilidades de las premisas para obtener la credibilidad de la conclusión. El resultado de la combinación de dos objetos que tienen cierto grado de magnitud es un nuevo objeto que también tendrá la magnitud en cierto grado (Díez y Moulines, 1999).

El presente modelo sostiene que esa combinación se realiza de una manera sistemática, siguiendo cierto tipo de ley, patrón o fórmula que relaciona las magnitudes en juego. En la línea de la interpretación subjetivista de la probabilidad (ver sec. 11.2., cap. 2, pt. I), las personas pueden diferir en sus grados de creencia sobre algo, pero no pueden hacerlo en el modo de combinarlos. En esa combinación, la información ya conocida y justificada actúa a modo de premisas de un argumento. Tal información posee cierto grado de credibilidad. A partir de la credibilidad de esas premisas, el sistema “calcula” deductivamente la que le corresponde a la nueva información (en tanto que se presenta como derivada de ellas). Se asume aquí que la consistencia probabilística

construye ese “cálculo”; esto es, construye la forma en la que las conclusiones heredan los grados de justificación de las premisas. Ahora se trata pues de ver específicamente cómo se produce esa transferencia de magnitudes. Para ello, nos adentraremos primero en el concepto de validez probabilística para, en el apartado siguiente, desarrollar la idea de consistencia probabilística y el modo en que ésta construye ese cálculo.

En orden a abordar estos conceptos, tomaremos como punto de partida el formalismo llamado probabilidad lógica (Pfeiffer, 2019; Williamson, 2002), lógica probabilística (Nilsson, 1986; Oaksford y Chater, 2007) o argumentación probabilística (Haenni, Kohlas y Lehmann, 2000). En esta lógica se combina el manejo de la incertidumbre que realiza la teoría de la probabilidad con la estructura de la argumentación de la lógica deductiva. Según algunos autores (e.g., Adams y Levine, 1975; Evans, Thompson y Over, 2015; Leblanc, 1983; Reichenbach, 1949), la lógica probabilística es una extensión natural de las tablas de verdad de la lógica deductiva, donde los valores de verdad se sustituyen por probabilidades. Así, los valores V-F de la lógica clásica bivalente pasan a corresponderse con los valores 1 y 0 respectivamente. De este modo, las premisas y conclusiones de los argumentos podrán tomar cualquier valor de probabilidad comprendido en el intervalo delimitado por el 0 y el 1. Es por esto por lo que se considera la lógica probabilística como una generalización de la lógica clásica (o, esta última, como un caso extremo de la primera).

En cualquier caso, conviene no perder de vista que, pese a que 1 y 0 puedan funcionar de forma equivalente a V y F, no son lo mismo. Tal como se explicó más arriba, una probabilidad subjetiva extrema sólo conlleva certeza subjetiva. Uno puede estar totalmente seguro de que algo es verdadero o falso y estar, a su vez, equivocado. En este modelo empleamos la probabilidad subjetiva para representar cuantitativamente la justificación. Y la justificación es de naturaleza epistémica y no semántica. Las probabilidades están definidas sobre un universo epistémico relativo al estado actual del conocimiento del mundo de un individuo (que, en el presente modelo, se corresponde con SC o SLC). Y el conocimiento suele ser limitado.

1. Validez probabilística de los argumentos deductivos cotidianos

1.1. Introducción a la validez probabilística

La idea central es que es posible realizar inferencias deductivas válidas usando argumentos basados en premisas que el individuo considera probables, de acuerdo con sus conocimientos. Las inferencias válidas preservarán la probabilidad. Si las premisas se consideran totalmente probables, entonces también se considerará así la conclusión. Pero si una o más premisas del argumento se considera incierta y participa esencialmente en la inferencia, entonces también se considerará incierta la conclusión⁴⁴.

⁴⁴ Como argumentaré más detalladamente en el próximo apartado, esto será así siempre y cuando la conclusión sea inicialmente desconocida o no creída (i.e., si no se dispone de razones alternativas para

Aunque la conclusión se infiera efectivamente de las premisas, puede carecer de justificación completa si alguna de las premisas también carece de ella. En el clásico ejemplo silogístico de “todo hombre es mortal – Sócrates es hombre – Por tanto, Sócrates es mortal”, la justificación de la conclusión menguaría si uno guardase reservas acerca de la verdad de las dos o una de las premisas.

1.2. Axiomas de probabilidad y sus consecuencias

La certidumbre y la incertidumbre se transmiten en la inferencia deductiva probabilística. Esa transmisión debe respetar los axiomas de la probabilidad de Kolmogorov (1933) vistos en la sec. X, cap. 2, pt. I. Brevemente, eran los siguientes (expresados aquí más sucintamente):

- a) No-negatividad: $P(\alpha) \geq 0$
- b) Tautologías o eventos seguros: $P(\Omega) = 1$, donde Ω es un suceso seguro, como el conjunto universal o una tautología (p. ej. inferencias del tipo $\alpha \rightarrow \alpha$; $\neg(\neg\alpha) \rightarrow \alpha$).
- c) Adición finita: sean $S_1, S_2 \dots S_n$ sucesos mutuamente excluyentes que conjuntamente conforman la totalidad del espacio muestral; entonces, resulta que $P(S_1 \cup S_2 \cup \dots \cup S_n) = \sum P(S_i) = P(\Omega) = 1$

De estos axiomas, se deducen una serie de consecuencias:

- a) Si la probabilidad del conjunto universal es $P(\Omega) = 1$, entonces la probabilidad del conjunto vacío es $P(\emptyset) = 0$, donde el conjunto vacío $\emptyset$ representa el suceso imposible.
- b) Si un evento o suceso es una partición perteneciente al conjunto universal (i.e., $\alpha \in \Omega$), entonces $P(\alpha) \leq 1$
- c) Si un evento o suceso $\alpha \in \Omega$ y $P(\Omega) = 1$, entonces la probabilidad complementaria de α es $\bar{P} = 1 - P(\alpha)$

1.3. Concepto de validez probabilística (*p*-validez)

Una vez más, en un razonamiento deductivo válido, si el individuo adscribe una probabilidad total de 1 a todas las premisas que lo componen, dada la preservación de la probabilidad, deberá otorgarle una probabilidad de 1 a la conclusión. Esto se debe a que en la conclusión no hay nada que no estuviera ya en las premisas. Dicho de otro modo, si no hay incertidumbre en las premisas, entonces tampoco la habrá en la conclusión. No debe agregarse incertidumbre en el proceso inferencial (Evans y Over, 2013). Sin embargo, si la hubiera en las premisas, podría o no verse reflejada en la conclusión. Para muchos autores (e.g. Adams, 1998; Evans y Over, 2013), la preservación del grado de probabilidad en una deducción válida supone conservar el grado de confianza de las premisas en sus conclusiones. En este sentido, estos autores consideran que una conclusión deductiva probabilísticamente válida (*p*-válida) no puede ser más improbable que la improbabilidad conjunta de sus premisas. A continuación, se explica con más detalle esta última afirmación.

creerla). Esto es consistente con el ya comentado estándar argumentativo que asume que la conclusión inicialmente debe ser menos creída que las premisas.

Siguiendo a Adams (1975, 1998), estos autores llaman “incertidumbre” a $\bar{P}$, donde $\overline{P(\alpha)} = 1$ se correspondería con la máxima “incertidumbre”. Así pues, según ellos, cuando la $P(\alpha) = 0$, la incertidumbre es máxima. Así, dicen que *un argumento es p-válido si y solo si la incertidumbre de la conclusión no puede ser coherentemente mayor que la suma de las incertidumbres de las premisas* (e.g. Adams, 1998; Evans et al., 2015; Oaksford y Chater, 2020).

A diferencia de los autores citados, en el presente modelo, la máxima incertidumbre se ha concebido como $P(\alpha) = 0.5$, de tal modo que no coincide con $\bar{P}$. Recordemos, por ejemplo, que una $P(\alpha) = 1$ es una certeza *tan absoluta como* una $P(\alpha) = 0$. En este sentido, cuando $P(\alpha) = 1$, $\overline{P(\alpha)} = 0$; y cuando $P(\alpha) = 0$, $\overline{P(\alpha)} = 1$. Si ambos extremos de probabilidad (i.e., $P(\alpha) = 1$ y $P(\alpha) = 0$) son certezas, entonces también lo son sus correspondientes complementarias $\bar{P}$. Y en la certeza no hay incertidumbre. Baratgin y Politzer (2016), entre otros autores, han puesto de manifiesto lo confuso de la nomenclatura de Adams (1998) al llamar incertidumbre a lo que no lo es. Lo que para Adams y otros autores se correspondía con la máxima incertidumbre, para el presente modelo se corresponde con la máxima disensión, según se comentó más arriba. Esa ‘incertidumbre’ (i.e., esta $\bar{P}$) se refiere al restante no creído. En la bibliografía sobre creencias y probabilidades, encontramos que también se le ha llamado “grado de descreencia” (e.g. Harris et al., 2008; Jøssang, 2002; Spohn, 2012) y, en el ap. C, le llamo “incredulidad” e “incredibilidad”. En lo que sigue, me referiré a ese concepto indistintamente como $\bar{P}$ (probabilidad complementaria) o incredibilidad⁴⁵. Pero, en este contexto, trataré de evitar en lo posible el término ‘incertidumbre’ excepto cuando sea necesario por estar citando a los autores que siguen la terminología de Adams.

Ateniéndonos a ello, el criterio de *p*-validez que comentan los autores anteriores quedaría definido como: un argumento es *p*-válido si y solo si la $\bar{P}$ (o incredibilidad) de la conclusión no puede ser coherentemente mayor a la suma de las $\bar{P}$ (o incredibilidades) de las premisas:

$$\text{Validez-}p: \bar{P}(\text{conclusión}) \leq \sum \bar{P}(\text{premisas})$$

Esto, a su vez, implica que los argumentos inválidos, como las falacias, sólo se producen cuando la $\bar{P}$ (o incredibilidad) de la conclusión es mayor que la suma de la incredibilidad de las premisas.

1.4. Implicaciones de la validez probabilística

En tanto que la suma de las incredibilidades de las premisas oscila entre 0 y 2, la *p*-validez es un requisito muy fácil de satisfacer por cualquier modo de combinación de probabilidades. Esa facilidad aumenta cuanto más bajas son las probabilidades de las premisas. Cuando el resultado del sumatorio de la $\bar{P}$ de las premisas sea superior a 1 (i.e., $\sum \bar{P}(\text{premisas}) > 1$), cualquier resultado en la conclusión será *p*-válido. Evans et al. (2015) consideran que esto es análogo a lo que ocurre con la validez clásica en un Modus Ponens con ambas premisas falsas: la validez sólo dice que, si las

⁴⁵ Pese a lo deseable de adoptar el término “descreencia” al mantener cierta unidad con el “disbelief” empleado por los autores citados, en nuestro idioma guarda relativa connotación religiosa que quiero evitar.

premisas son verdaderas, entonces la conclusión también; no se puede violar la validez si se parte de premisas falsas.

Por otro lado, ya vimos (cap.5, Pt II; ap. C) que, en la validez clásica, la relación de consecuencia es preservadora sólo de la verdad y va sólo de las premisas a la conclusión. En este sentido, vimos que es perfectamente posible que se den argumentos válidos con premisas falsas y conclusión verdadera. Así pues, como se puede observar, la *p*-validez respeta esta regla: cuando las premisas son poco o nada creídas, la conclusión puede ser nada, poco, moderadamente, muy o totalmente creída.

1.5. Límites y extensión de la *p*-validez

Una virtud de la *p*-validez es la de que permite establecer un *límite superior de incredibilidad* de la conclusión que es informativo, por lo menos, en todos aquellos casos donde no suceda $\sum \bar{P}(\text{premisas}) > 1$. En su correlato *directo*, la *p*-validez permite establecer un *límite inferior para la credibilidad o probabilidad* de las conclusiones a partir de conocer las probabilidades de las premisas. Por ejemplo, partiendo de un argumento con dos premisas, si otorgamos una credibilidad de 0.7 a cada una de ellas, nos queda un grado de incredibilidad $\bar{P} = 0.3$ para cada una. De este modo, $\sum \bar{P}(\text{premisas}) = 0.6$. Entonces, pasándolo a probabilidades directas, nos queda que $1 - 0.6 = 0.4$. Lo que significa que el individuo deberá dar una probabilidad a la conclusión de, como mínimo, 0.4 (una forma más directa y sencilla de calcular este límite inferior es $\sum P(\text{premisas}) - 1$; siguiendo el ejemplo: $(0.7 + 0.7) - 1 = 0.4$). Según Baratgin y Politzer (2016), si un individuo juzga la probabilidad de la conclusión como siendo igual o superior al límite inferior establecido por sus premisas, se puede asumir que dicho individuo tiene conocimientos implícitos sobre la preservación de probabilidades en el argumento, y que no los tiene en caso contrario. Como se verá en breve, el tipo de cálculo que se propone en el presente modelo respeta la *p*-validez, derivando en predicciones sobre las probabilidades de la conclusión siempre superiores a ese límite inferior.

1.6. Dinámica de la acumulación de incredibilidad en argumentos *p*-válidos

Por otro lado, en los argumentos *p*-válidos, la preservación de la probabilidad se da de tal modo que, si un argumento válido tiene una pequeña cantidad de premisas, y cada una de las cuales tiene una pequeña ‘incertidumbre’ (en el sentido de Adams) o incredibilidad (es decir, si tiene una probabilidad alta), entonces su conclusión también tendrá una ‘incertidumbre’ o incredibilidad razonablemente pequeña (es decir, una probabilidad razonablemente alta). Por el contrario, si un argumento válido tiene un gran número de premisas con pequeñas ‘incertidumbres’ o incredibilidades, entonces su conclusión sólo puede ser muy ‘incierta’ o muy poco creída (Demey, Kooi y Sack, 2019). Es decir, *una conclusión derivada de una larga cadena de premisas, de algún modo, habrá acumulado las incredibilidades de todas ellas*. Más adelante, veremos que esto también es consistente con el tipo de cálculo de credibilidades que se propone aquí. Su correlato no inverso sería que un argumento es *p*-válido si y solo si la credibilidad de la conclusión puede acercarse arbitrariamente a 1 si las probabilidades de las premisas son suficientemente altas.

2. Consistencia probabilística de los argumentos deductivos cotidianos: una propuesta de cálculo para la dinámica de credibilidades

En este apartado expondré con detalle la hipótesis principal de esta tesis. Como hemos visto, el presente modelo de sistema de creencias sostiene que la justificación se transmite desde unas representaciones ya justificadas hacia otras que no lo están (o que lo están menos) mediante cierto tipo de relaciones; a saber, relaciones tópicas. Asimismo, el modelo defiende que estas relaciones tópicas son parte de la noción más general de relaciones de coherencia. Y, como también se dijo, las relaciones de coherencia son tipos de relaciones explicativas. Las explicaciones adoptan tres formatos en función de los mecanismos inferenciales empleados en ellas: el abductivo, el inductivo y el deductivo. Por tanto, la función de adhesión producida por la transmisión de justificación se lleva a cabo a través de estos tres mecanismos inferenciales.

Por otro lado, este modelo sostiene que la justificación es una magnitud que se da en grados. Esos grados de justificación (o, finalmente, de adhesión o de creencia) se miden como probabilidades subjetivas. Y en tanto que tales, su cómputo debe respetar los axiomas de la probabilidad. Así pues, cuando la adhesión se produce por medio de relaciones explicativas de tipo deductivo, se hace necesario abrazar una concepción nueva de razonamiento deductivo, un razonamiento que opera ahora con probabilidades subjetivas, con esa magnitud graduada, en lugar de con la dicotomía de lo verdadero y lo falso.

En el apartado anterior se ha abordado la validez de la deducción probabilística examinando su propiedad de ser preservadora de la probabilidad. La validez permite la transmisión de los grados de justificación de las premisas a la conclusión. Es, pues, necesario que esos grados de justificación de las premisas se combinen de algún modo a la hora de ser transmitidos a la conclusión. Según la acepción más aceptada de validez probabilística, un argumento es probabilísticamente válido si la incredibilidad de la conclusión es inferior a la suma de las incredibilidades de las premisas. Esto ya está proporcionando alguna indicación acerca de la credibilidad de la conclusión dadas las credibilidades de las premisas. Aun así, no parece una condición muy restrictiva ni tampoco una descripción suficiente del modo en que las credibilidades de las premisas se combinan.

Como ya se dijo páginas atrás, el modo en el que el sistema cognitivo ejecuta esa combinación debe ser sistemático e igual para todas las personas. Crucialmente, el presente modelo sostiene que la consistencia probabilística impone las restricciones críticas en el modo en el que se combinan las credibilidades de las premisas. La credibilidad de la conclusión deberá ser consistente con la de las premisas. Y eso significa que no debe entrar en contradicción con ellas. Para desarrollar esta idea, veamos un poco más de cerca en qué consiste la consistencia probabilística.

2.1. Consistencia probabilística subjetiva

Se considera que un conjunto de creencias es consistente para un individuo si éste puede atribuirles a todas ellas simultáneamente una alta probabilidad subjetiva, es decir, si no se dan contradicciones entre ellas en un momento dado. Dicho de otro modo, si del estado actual de tal conjunto no puede deducirse una idea y su negación. Este es el único requisito. Así, por ejemplo, una creencia como la de que “la distancia que separa a Tarragona de Barcelona es de exactamente 100 km” es perfectamente consistente con otra como “hoy he desayunado cereales”. Puedo creer sin contradicción en ambas a la vez. De hecho, las creencias que no guardan absolutamente ninguna relación *para el individuo* siempre serán consistentes, pues para contradecirse tienen que estar relacionadas de algún modo (y conscientemente) en la mente de la persona. Por ejemplo, si en mi estado epistémico actual creo que de Tarragona a Barcelona hay exactamente 100 km, no puedo creer que sólo hay 5, ni que hay 200. Estas otras posibles creencias sí que están relacionadas con la inicial y entrarían en contradicción con ella porque implican su negación siendo así inconsistentes. Si dos creencias se contradicen, como mínimo una debe ser incorrecta. No puedo sostener las dos a la vez.

Pasando del nivel de las creencias al del lenguaje, es perfectamente admisible que una afirmación sea verdadera en un sentido y falsa en otro, pero no puede ser verdadera y falsa en el mismo sentido o interpretación. Del mismo modo y volviendo a las representaciones mentales, un individuo no puede creer que una cosa es y no es a la vez y bajo el mismo aspecto. Esto responde al *principio de no contradicción*. Este es, según el presente modelo, el principio rector del sistema de creencias. El sistema de creencias buscará eliminar la contradicción allá donde pueda detectarla, esto es, en el sistema local de creencias activado por la información entrante.

Ahora bien, la mayoría de las veces creemos que las cosas son de tal manera hasta cierto punto. En este sentido, podría parecer que creemos en algo y su negación simultáneamente y, por tanto, parecería que la inconsistencia es la norma general en el sistema de creencias. Pero esto no es así. Las inconsistencias que haya en el sistema de creencias no son de este tipo. Veamos.

2.2. El Concepto de contradicción en un sistema probabilístico

En primer lugar, “creer algo hasta cierto punto” implica que ese “algo” tiene cabida en el sistema de creencias. Y el sistema de creencias solo alberga eso, creencias; no alberga ideas que no creamos. Al creer algo hasta cierto punto, en tanto que es creído, su contrario no es creído. Debido a la acción de los refutadores activados, se acepta la posibilidad de su contrario. Pero esto es diferente de creer en su ocurrencia. Es como cuando uno dice o piensa “puede que mañana llueva, pero por las razones x, y, z, no lo creo”.

En segundo lugar, teniendo en cuenta que la credibilidad es una magnitud con límites mínimos y máximos claros, ese “cierto punto” es mensurable. Así pues, uno puede creer totalmente que “mañana hará sol”, asignándole implícitamente una probabilidad subjetiva de 1 y a su contrario de 0. Por otro lado, uno podría guardar ciertas reservas, asignándole implícitamente una probabilidad de, digamos, 0.7. En este último caso, está aceptando la posibilidad de su contrario en un 0.3. Pero

eso no significa que crea que mañana estará nublado. En este sentido, este creer en un 0.7 y aceptar la posibilidad de su contrario en la proporción complementaria de 0.3 es perfectamente consistente. No hay contradicción. Por otra parte, se podría decir que a estos 0.7 y 0.3 les corresponden dos *proposiciones* relacionadas y contrarias, pero sólo una *creencia* propiamente dicha, que se corresponde con el 0.7.

Como vemos, se respeta el axioma de probabilidad de la adición finita. En cambio, cuando no se respeta este axioma sí se cae en contradicción. No puedo creer algo en 0.7 y su contrario también en 0.7. La probabilidad subjetiva del contrario deberá ser siempre la complementaria. Por tanto, sólo puedo creer en algo y su contrario en la misma medida cuando soy absolutamente ignorante a su respecto, porque no tengo ninguna razón para otorgarle más credibilidad a una cosa que a otra, aplicándole el principio de indiferencia. En este sentido, si el universo conocido a su respecto consta sólo de dos posibilidades (e.g., ese algo y su contrario), entonces puedo atribuirles a ambas posibilidades una probabilidad subjetiva de 0.5, puesto que siguen sumando 1.

2.3. Consistencia y contradicción probabilística en argumentos

Ahora explicaré de qué manera se aplica la consistencia en argumentos entimemáticos, así como en otras variantes de argumentos no entimemáticos con una o más premisas no creídas. A la hora de abordar esta explicación nos serviremos de cierta variedad terminológica que se expone en la Tabla 2 para referirnos a las premisas y a la conclusión de los argumentos.

Tabla 2

Relación de Correspondencias de Nomenclatura para las Partes del Argumento

Nomenclatura Lógica Estándar	Nomenclatura funcional inspirada en el Modelo de Toulmin		Nomenclatura en SLC	Nomenclatura en Teoría de la Explicación
Premisa Mayor	∅	Garante	<i>Tp</i>	Regla o ley general
Premisa Menor	∅	Dato	<i>Nj</i>	Condiciones antecedentes
Conclusión	Aserción Prior	Aserción Post	<i>Nc</i>	Unidad informativa entrante

Nota: Esta tabla distingue entre aserciones no precedidas y precedidas por premisas. Una aserción no precedida por premisas se denomina "Aserción Prior", mientras que una que sí lo está se considera una "conclusión" o "Aserción Post" (posterior a las premisas). En el nivel psicológico del presente modelo, la conclusión se correspondería con el *Nc*. Además, la tabla incorpora términos de la teoría de la explicación para facilitar la integración de los conceptos discutidos, donde, la unidad informativa entrante hace referencia al *explanandum*.

2.3.1. Las premisas

Primero, centrémonos en la consistencia de las premisas. Vamos a ver que las premisas de los entimemas y de los argumentos no entimemáticos pueden ser consistentes o no en varios sentidos.

En los Entimemas: Primero, las premisas deben ser consistentes entre sí. En los entimemas, las premisas incluyen la representación de un hecho o dato del mundo y una regla de relación tópica aplicada a un contenido específico. Estos elementos son considerados *sucesos independientes*, lo

que asegura su consistencia. La independencia entre dos sucesos implica que la probabilidad asignada a uno no depende de la verificación del otro. Es decir, puedo creer en cada premisa en cualquier grado, ya que la creencia en una no afecta mi creencia en la otra. Por ejemplo, mi creencia en que “los medios de transporte emiten gases de efecto invernadero” es independiente de mi creencia en que “la continuación de estas emisiones causará la extinción de los osos polares”. Puedo creer en cada una a cualquier nivel entre 0 y 1, lo que permite una creencia total en ambas. Al ser sucesos independientes, no existe contradicción entre ellos.

En segundo lugar, las premisas del entimema son consistentes con las representaciones mentales almacenadas en nuestro SLC, es decir, no contradicen dichas representaciones. Además, su credibilidad *depende* crucialmente de los conocimientos y creencias activados en el SLC. La probabilidad subjetiva asignada a los contenidos de las premisas *está condicionada* por esta base de conocimientos y creencias activada. Los contenidos de las premisas están relacionados con los del SLC de manera que este último justifica un alto grado de creencia en aquellos⁴⁶. Dependiendo de los refutadores que recuperemos de nuestra memoria a largo plazo, las premisas serán creídas en un rango entre 0.5 y 1. Así, dado que los contenidos de las premisas se relacionan con los del SLC, los consideramos *relevantes* y, dado que también son consistentes, los consideramos *coherentes* con ellos. No solo son compatibles con lo almacenado en el sistema local de creencias, sino que son efectivamente creídos por los oyentes.

En otros Argumentos no Entimemáticos: Por contraposición a los entimemas, los no entimemáticos constarán de, al menos, una premisa no creída por los oyentes. Para la consistencia subjetiva de las premisas entre sí, al igual que en los entimemas, basta con que éstas sean consideradas como sucesos independientes. Por tanto, también es de aplicación para el siguiente par de premisas: “si se incrementan las ventas de telefonía móvil, entonces se contribuirá a la extinción de los osos polares” y “los medios de transporte contribuyen al aumento de ventas de telefonía móvil”. Nuevamente, estas dos premisas se refieren a sucesos independientes entre sí y, por tanto, el grado de credibilidad otorgado a cualquiera de ellas no tiene consecuencias para la otra. Son también consistentes entre sí.

Sin embargo, estas premisas pueden no parecer tan naturales como las anteriores. Anteriormente se mencionó que las premisas de los entimemas suelen ser creídas por los oyentes, ya que forman parte del sistema local de creencias. No obstante, en este caso, las premisas no son creídas en nuestra comunidad epistémica. La sensación de extrañeza que provocan podría deberse a que son inconsistentes o irrelevantes respecto a nuestro SLC. Al procesar estas premisas, se activan en nuestro sistema de creencias conceptos y relaciones que pueden ser directamente inconsistentes con los contenidos de las premisas, o consistentes pero irrelevantes. En el primer

⁴⁶ Esto será así para las premisas que activen creencias derivadas. Probablemente, la mayoría de los contenidos de premisas que presente un orador a un interlocutor sean de este tipo. La excepción a esta regla se daría cuando los contenidos de las premisas se identifiquen con una creencia fundacional de cualquiera de las tres tipologías descritas, pues no reciben justificación de otras creencias anteriores, sino que lo hacen por el propio modo de funcionar del sistema cognitivo.

caso, tenemos razones para desconfiar de las premisas, otorgándoles baja credibilidad. En el segundo, la falta de razones para creerlas nos lleva a una incertidumbre, resultando en una credibilidad intermedia. En ninguno de los dos casos se creerán esos contenidos.

2.3.2. La conclusión

Hemos visto que la consistencia de las premisas se puede dar en dos niveles: por un lado, pueden ser o no mutuamente consistentes entre ellas y, por otro lado, cada una de ellas puede ser o no consistente con los conocimientos y creencias previas del individuo.

Los entimemas operan de manera diferente. Como se mencionó anteriormente, las premisas de un entimema suelen ser aceptadas por los oyentes de antemano, mientras que la conclusión inicialmente no lo es. Antes de presentar el argumento, los contenidos que serán parte de la futura conclusión pueden ser inconsistentes o irrelevantes para el sistema de creencias del individuo. Sin embargo, una vez que el argumento es presentado y las premisas son aceptadas, el sistema busca reconciliar la credibilidad de la conclusión con la de las premisas para lograr la consistencia en SLC. Si las premisas son aceptadas en cierta medida, la conclusión también debe ser aceptada en cierta medida evitando caer en contradicciones. Así, la credibilidad de la conclusión *depende* completamente de la credibilidad de las premisas, las cuales, a su vez, están condicionadas por el sistema local de creencias.

La conclusión incorpora elementos semánticos de las premisas: la representación del referente, que aparece en la premisa menor, y del contenido del predicado, presente en la premisa mayor. Además, la credibilidad de la conclusión debe derivarse sistemáticamente de los grados de credibilidad de las premisas, las cuales son igualmente importantes en términos de relevancia y pertinencia para la inferencia. Sin N_j , no existe una base de justificación que transmitir, y sin T_p , no se puede transferir la credibilidad de N_j , sin importar su magnitud. Por lo tanto, este modelo propone que, en los entimemas, la consistencia rige el modo en el que la credibilidad de las premisas determina la estimación de la probabilidad subjetiva asignada a la conclusión.

2.4. Evaluación de la credibilidad como magnitud en función de la combinación de premisas

Como se mencionó anteriormente, siguiendo a Díez y Moulines (1999), la combinación de dos objetos que poseen una magnitud en cierto grado resulta en un tercer objeto que también presenta la magnitud en cierto grado. En este contexto, los objetos que se combinan son las credibilidades de las premisas de un mismo argumento y el objeto resultante es la credibilidad de la conclusión. Los objetos y sus magnitudes se pueden combinar de diferentes maneras, como la agregación o la mezcla, entre otras. A cada manera le corresponde una operación distinta. Dependiendo del método de combinación, el objeto resultante puede exhibir una magnitud que es:

- a) Mayor que la de los objetos originales; se les denomina magnitudes *crecientes*. Un ejemplo sería la combinación de masas, longitudes o tiempos.

- b) Menor que la de los objetos originales; se les conoce como magnitudes *decrecientes*. Por ejemplo, la combinación de resistencias eléctricas en paralelo.
- c) Intermedia entre las magnitudes de los objetos originales, referidas como magnitudes *internas*. Un ejemplo sería la mezcla de temperaturas o densidades.

Dado que la credibilidad de la conclusión debe estar limitada entre 0 y 1, es evidente que no puede ser una magnitud creciente; de lo contrario, podría superar el 1 incluso con solo dos premisas, y este exceso se agravaría con la adición de más premisas en argumentos más complejos. Por lo tanto, solo consideraremos las magnitudes decrecientes e intermedias.

Como se ha mencionado antes (sec. 1.6., cap. 7, pt. II), nuestro modelo sostiene que la credibilidad de la conclusión disminuye a medida que aumenta el número de premisas en un argumento. Por ejemplo, un argumento con una premisa de probabilidad subjetiva de 1 y otra de 0.9 será más creíble que uno donde ambas premisas tienen 0.9. Esta credibilidad se reduce aún más en argumentos con tres premisas al 0.9, y continúa disminuyendo con cada premisa adicional. Esto se debe a que la conclusión acumula la incredulidad de cada premisa. Por lo tanto, la credibilidad de la conclusión solo puede ser de tipo *decreciente*, ya que las alternativas crecientes e intermedias resultarían en magnitudes para la conclusión que son inconsistentes con las premisas.

2.5. Formulación de la hipótesis multiplicativa para la combinación de credibilidad en entimemas

Ahora exploraremos qué operación representa la combinación decreciente de credibilidad en la deducción mediante entimemas. Aunque pueda parecer redundante, es crucial reafirmar un punto ya mencionado: la credibilidad de cada premisa es esencialmente igual de importante. Sin N_j , no existe una base fáctica de credibilidad para transmitir, y sin Tp , falta la regla general que facilita esta transmisión.

Así pues, consideremos una situación idealizada de argumentación donde las premisas tengan valores extremos de credibilidad (1) o incredulidad (0). Si partimos de valores extremos de incredulidad en la conclusión (0), obtendríamos la Tabla 3 de relaciones de credibilidad. En la presentación de esta tabla, tomaré la terminología de Toulmin (1958): Dato, Garante y Aserción crítica. Nótese que sólo la primera columna de valores se correspondería con un entimema:

Tabla 3

Relaciones de Credibilidad en Argumentos con Valores Extremos en las Premisas

	Valores de Credibilidad			
	1	1	0	0
Garante	1	1	0	0
Dato	1	0	1	0
Aserción Crítica	1	0	0	0

Nota: Esta tabla muestra las credibilidades resultantes en la conclusión basadas en la combinación de las credibilidades extremas de las premisas, según el modelo. Se asume que la credibilidad inicial o a priori de la conclusión es 0.

Recordemos entonces que las restricciones básicas impuestas por la consistencia en la deducción probabilística de este modelo son:

- Adherencia a los axiomas de la teoría de la probabilidad, especialmente la no-negatividad y la adición finita, que limitan los valores entre 0 y 1.
- Tratamiento de la credibilidad en la conclusión como una magnitud decreciente, debido a la acumulación de incredulidad en la combinación de premisas.
- Roles equivalentes de Dato y Garante en la transmisión de justificación o credibilidad a la aserción crítica.

Por tanto, la operación de combinación que mejor representa cómo la credibilidad de la conclusión está condicionada y determinada por la credibilidad de las premisas es la *multiplicación aritmética*, expresada de la siguiente manera:

$$P(\text{aserción}) = P(\text{Dato}) \cdot P(\text{Garante})$$

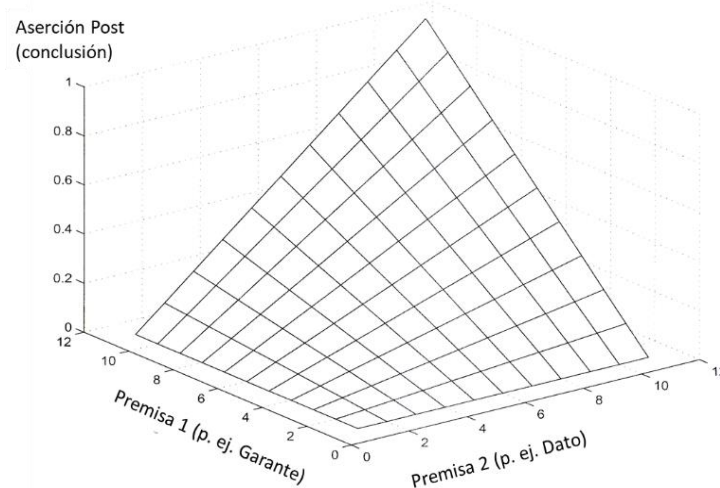
O, de un modo resumido, como (cualquier nomenclatura equivalente es válida):

$$P(\text{conclusión}) = \prod P(\text{premisas})$$

El gráfico de la Figura 12 muestra la distribución esperada de credibilidades en función de la combinación de los valores de las premisas:

Figura 12

Credibilidad de la Conclusión en Argumentos con Valores No Extremos en las Premisas



Nota: En esta figura podemos observar la credibilidad que se espera desde nuestro modelo en las aserciones críticas o conclusiones en función de todo el rango de valores de credibilidad que pueden adoptar las premisas.

3. Consecuencias de la multiplicación:

3.1. Regla epistémica de la parte más débil

3.1.1. En condiciones ideales de argumentación

Una consecuencia de este cómputo multiplicativo es que la credibilidad de la conclusión deberá ser igual o menor a la credibilidad de la premisa menos creída. En efecto, si de dos premisas, una de ellas tiene el valor máximo de 1 y la otra tiene un valor inferior j , entonces la credibilidad de la conclusión se corresponderá con ese valor j (i.e., $1 \times j = j$). Sin embargo, si ninguna premisa tiene un valor máximo, sino que ambas están por debajo de 1, entonces la credibilidad de la conclusión será inferior a la de la premisa menos creída⁴⁷ (i.e., $0.9 \times 0.8 = 0.7$). Así, al distinguir premisas con una credibilidad alta (A) y con una credibilidad baja (B) nos queda que:

$$P(\text{aserción}) \leq P(B)$$

Por tanto, en estas condiciones idealizadas de incredulidad total inicial en la conclusión, la multiplicación establece un *límite superior* para su credibilidad posterior a la presentación del argumento. La credibilidad de la conclusión será, como máximo, tan alta como la credibilidad más baja de las premisas. En los experimentos que serán expuestos más adelante, me referiré a esta propiedad como “regla *epistémica* de la parte más débil”, siguiendo la estela de la regla silogística de la parte más débil que se presenta con cierto detalle en el ap. C (sección de “reglas complementarias”).

3.1.2. En condiciones realistas de argumentación

Consideremos ahora condiciones más realistas. Generalmente, es probable que las personas tengan diversas razones y refutadores en su SLC al evaluar tanto las aserciones críticas aisladas como las premisas. Por lo tanto, es plausible que en general evalúen las premisas y las aserciones críticas aisladas (aquí denominadas 'Aserciones Prior' y futuras 'Aserciones Post' o conclusiones) con una probabilidad subjetiva baja pero superior a 0. Esto indica que posiblemente tengan algunas razones que justifiquen mínimamente la posibilidad de que el contenido de estas aserciones pueda ser cierto, asignándoles una probabilidad subjetiva de hasta 0.5. En condiciones cotidianas, la credibilidad inicial de una futura conclusión será, por lo tanto, algo mayor que 0. Del mismo modo, es probable que las premisas B también tengan una credibilidad inicial mayor que 0, mientras que las premisas A, no serán totalmente creídas en su mayoría, sino que sólo es seguro que se les asignará una probabilidad subjetiva mayor a 0.5.

Imaginemos entonces que un individuo cree que “los medios de transporte contribuirán a la extinción de los osos polares” en un 0.4 (i.e., considera que, hasta cierto punto, “podría ser, pero no lo creo”). Entonces, si un interlocutor le presenta un argumento con $\prod P(\text{premisas}) > 0.4$, al estar proporcionando nuevas razones para creer, la multiplicación será efectiva. Sin embargo, si

⁴⁷ Nótese que esta consecuencia teórica del presente modelo permite explicar y predecir los resultados empíricos obtenidos en el estudio de George (1997) expuestos en el capítulo 4.

ΠP (*premisas*) < 0.4 , el resultado puede variar según las diferencias individuales. El individuo podría:

- Desestimar la información de las premisas y mantener su credibilidad inicial basada en sus razones previas.
- Inhibir sus creencias previas sobre la conclusión y aceptar únicamente la evaluación de las premisas, asignando la credibilidad de ΠP (*premisas*) a la conclusión.
- Intentar un equilibrio, sin descartar completamente las premisas (atendiendo pues a la multiplicación) ni ignorar sus creencias previas (tomándolas como límite superior de la credibilidad).

Podemos formalizar esto de la siguiente manera:

Si la combinación de las credibilidades de las premisas (ΠP (*premisas*)) es mayor o igual que la credibilidad inicial de la aserción (P (*Aserción Prior*)), entonces la credibilidad de la aserción posterior (P (*Aserción Post*)) será igual a la combinación de las credibilidades de las premisas:

- Si (ΠP (*premisas*) $\geq P$ (*Aserción Prior*)), entonces P (*Aserción Post*) = ΠP (*premisas*).

Si la combinación de las credibilidades de las premisas es menor que la credibilidad inicial de la aserción, entonces la credibilidad de la aserción posterior estará entre la combinación de las premisas y la credibilidad inicial:

- Si ΠP (*premisas*) $< P$ (*Aserción Prior*), entonces P (*Aserción Prior*) $\geq P$ (*Aserción Post*) $\geq \Pi P$ (*premisas*).

Veamos esto con ejemplos numéricos en una tabla igual que la anterior, pero no con valores extremos, sino más realistas, altos (A) y bajos (B), y suponiendo que el individuo le otorga a la aserción prior un 0.4:

Tabla 4

Relaciones de Credibilidad con Valores no-Extremos

	Valores de Credibilidad			
	0.9 (A)	0.8 (A)	0.2 (B)	0.3 (B)
Garante	0.9 (A)	0.8 (A)	0.2 (B)	0.3 (B)
Dato	0.7 (A)	0.3 (B)	0.9 (A)	0.2 (B)
Aserción	0.63	$0.4 \geq P(x) \geq 0.24$	$0.4 \geq P(x) \geq 0.11$	$0.4 \geq P(x) \geq 0.06$

Nota: En esta tabla se muestran las credibilidades resultantes en la conclusión (i.e., Aserción Post) en función de la combinación de unas credibilidades no extremas en las premisas y asumiendo una credibilidad inicial en la aserción superior a 0 (aquí, de 0.4).

De forma más general, siempre que la multiplicación resulte en un valor *superior* al de la credibilidad a priori funcionará adecuadamente como predictor en tanto que representa el modo en el que se combinan las credibilidades de las premisas (teniendo como límite superior el valor de credibilidad de la premisa menos creída). Sin embargo, siempre que la multiplicación resulte en un valor *inferior* a la credibilidad a priori, la credibilidad que finalmente se le otorgue a la conclusión tendrá como

límite inferior el resultado de la multiplicación, y como límite superior, el valor de la credibilidad a priori.

Y esto será así siempre. Hemos visto un ejemplo con una credibilidad a priori baja. Pero también será de aplicación cuando la credibilidad a priori sea alta. Por ejemplo, supongamos que nuestra credibilidad a priori es de 0.7. Un interlocutor podría querer que la aumentásemos aún más. Pero el argumento que construye sólo logra un producto del 0.5. En ese caso, o bien nos quedamos con nuestro 0.7 inicial o, basándonos solo en las premisas, bajamos al 0.5, o nos quedamos en algún punto entre esos dos valores.

3.2. Estándar de argumentación revisado y la función de fuerza argumental

Para mover un mínimo de un punto más arriba nuestro 0.7 debería encontrar premisas con alguno de los siguientes valores: 1×1 ; 1×0.9 ; 1×0.8 ; 0.9×0.9 . Es decir, premisas con un valor máximo conjunto de incredibilidad de 0.2.

De hecho, esto revela otra implicación de la multiplicación: para aumentar significativamente la credibilidad a priori de una aserción, es necesario seleccionar premisas cuya incredibilidad combinada máxima sea igual al complemento de la credibilidad a priori de la conclusión menos un punto (es decir, 0.1). Dado que un argumento se considera fuerte si logra aumentos significativos en la credibilidad de las aserciones críticas, podríamos denominar a este efecto como 'función de fuerza argumental' o FFA. De este modo, la FFA quedaría expresada como:

$$FFA \leq \bar{P} (\text{Aserción Prior}) - 0.1$$

Y esto supone, a su vez, una revisión del estándar de argumentación según el cual, para que un argumento sea convincente y efectivo, es necesario que las premisas sean más creíbles que la conclusión. La consecuente revisión que se deriva de la multiplicación es la de que, para que un argumento sea mínimamente fuerte será necesario que el sumatorio de la incredibilidad de las premisas sea igual o menor a la incredibilidad de la aserción a priori. Formalmente:

$$\sum \bar{P} (\text{premisas}) \leq \bar{P} (\text{Aserción Prior})$$

El FFA implica que cuanto mayor es la credibilidad inicial en la aserción prior, menos posibilidades de combinación de premisas quedan para producir un aumento significativo de un punto, de tal modo que cuando la credibilidad inicial es 1, las posibilidades de combinación de premisas son 0, mientras que cuando la credibilidad inicial es 0 las posibilidades de combinación para que la credibilidad aumente en, como mínimo, un punto, son más de 40. Por otra parte, cuando la credibilidad inicial es de 0.5, las posibilidades de combinación de credibilidades de las premisas ya se han reducido a 9.

Es decir, que cuanto menor es la credibilidad inicial, más fácil será que el argumento que construyamos sea fuerte, en el sentido de que logre un aumento significativo de al menos un punto. Ahora bien, eso no significa que un argumento fuerte definido de este modo logre aumentar la

credibilidad en la aserción inicial a un valor superior al 0.5 pasando así a ser creída por el individuo. Como vemos, para ello, solo hay 9 posibilidades de combinación.

3.3. Implicaciones de la multiplicación probabilística: cómo premisas creíbles pueden derivar en conclusiones menos creíbles

En este modelo, se considera baja una credibilidad inferior al 0.5. Proponemos que la dinámica de las credibilidades se aproxima a una multiplicación. Esta operación no solo es consistente sino también *p*-válida. Por ejemplo, si tenemos dos premisas con una credibilidad de 0.8 cada una, la *p*-validez nos indicaría que la credibilidad mínima de la conclusión sería 0.6 (0.8+0.8-1), mientras que, por multiplicación, sería de 0.64 (0.8×0.8). Esto sucede con cualquier valor.

El umbral de 0.5 separa las credibilidades altas de las bajas. Por lo tanto, si la multiplicación es una aproximación adecuada, es completamente posible que premisas con altas credibilidades deriven en una conclusión de baja credibilidad. Esto pone de relieve una diferencia importante con la validez lógica⁴⁸, donde de premisas verdaderas se deduce necesariamente una conclusión verdadera. En contraste, en la validez probabilística, altas probabilidades en las premisas no garantizan necesariamente una alta probabilidad en la conclusión. Depende del nivel de probabilidad de las premisas y de cómo se combinan.

Por ejemplo, si cada premisa tiene una probabilidad de 0.9 y 0.8, la probabilidad de la conclusión será alta, de 0.72. Sin embargo, si ambas premisas tienen una probabilidad de 0.6, la conclusión resultará en una baja probabilidad de 0.36, demostrando que dos premisas relativamente creíbles pueden resultar en una conclusión poco creíble. Esta característica de la multiplicación refleja cómo la combinación de magnitudes decrecientes afecta la credibilidad final.

Criterio Para que una Representación Entrante se Integre en el Sistema de Creencias

En este sentido, para lograr que la credibilidad de la conclusión pase a ser creída será necesario cumplir con el siguiente criterio:

$$\sum \bar{P}(\text{premisas}) < 0.5 \rightarrow \prod P(\text{premisas}) > 0.5$$

Esto es, que el sumatorio de las incredibilidades de las premisas debe ser inferior a 0.5 para que la multiplicación de las credibilidades de las premisas sea superior a 0.5, produciendo así un argumento que pase a ser creído independientemente de la credibilidad inicial de la aserción crítica.

⁴⁸ Como veremos en la discusión al experimento 2, esta diferencia será relevante a la hora de abordar las críticas a las teorías del razonamiento plausible que apoyan la regla de Teofrasto (también llamada condición de consecuencia, o regla del eslabón más débil o regla MIN).

4. Reflexiones Finales y Perspectivas Futuras

En este trabajo, hemos desarrollado un modelo comprensivo del sistema de creencias y la dinámica de la credibilidad que no solo resalta la complejidad inherente en la formación y justificación de las creencias, sino que también proporciona un marco para entender cómo las interacciones entre diferentes tipos de creencias—fundacionales (i.e., teóricas, empíricas y comunicacionales) y derivadas—pueden ser entendidas a través de un enfoque probabilístico y deductivo. En su recorrido, hemos visto una propuesta sobre sus fuentes primarias de formación, su función central de adaptación al entorno, su desarrollo individual y sociocultural, así como su función de marco a través del cual sustentar interpretaciones generales del mundo. Asimismo, también hemos abordado los distintos elementos de los que se compone una creencia, lo que nos ha llevado a realizar una propuesta general del sistema de creencias entendido como un sistema argumentativo. Hemos propuesto que, en ese sistema, las creencias están interconectadas por relaciones explicativas, prestando una especial atención a las relaciones inductivas y deductivas. Al abordar la inducción, hemos planteado una propuesta sobre la formación de creencias generales o esquemas, y sobre la aplicación inductiva de los mismos a nuevas instancias particulares que participan de una regla. También hemos realizado una serie de consideraciones acerca de la medición de la credibilidad como probabilidades subjetivas. Finalmente, hemos detallado la aplicación y dinámica de la credibilidad mediante las relaciones de explicación deductiva, proponiendo una serie de hipótesis que serán puestas a prueba en la siguiente parte de esta tesis.

Las implicaciones de este modelo teórico tienen la posibilidad de extenderse más allá de los límites de la investigación básica, pudiendo ser de interés para otros campos que requieran de una comprensión precisa de cómo las personas forman y ajustan sus creencias en respuesta a nueva información. Asimismo, este modelo abre nuevas vías para la investigación sobre los procesos mentales en la validación de la información y el cambio de opinión, sentando unas bases cognitivas para la definición y descripción de las creencias, así como del tipo de sistema en el que están albergadas. Pese a los avances que este modelo propone, se reconoce la necesidad de investigaciones futuras que puedan abordar sus limitaciones. En efecto, este modelo se limita a creencias acerca de posibles hechos, sin abordar creencias que implican juicios de valor, como las actitudes y los propios valores de las personas. Consideramos que este último tipo de creencias está directamente ligado a la toma de decisiones, a la motivación y emoción, y a la acción. En ese sentido, sería necesaria una exploración profunda de cómo los contextos culturales específicos pueden influir sobre los procesos de adhesión.

Finalmente, este trabajo invita a reconsiderar la interacción entre la lógica deductiva y la inferencia probabilística en el mantenimiento y revisión del sistema de creencias, destacando la importancia de entender no solo cómo se forma una creencia, sino también cómo una red de creencias se ajusta y recalibra en la presencia de información que ya era aceptada, pero que no estaba en el foco atencional en un primer momento al evaluar una pieza informativa. Este enfoque dinámico y relacional ofrece un marco valioso para futuras exploraciones en la psicología del

razonamiento y de procesos del discurso. En las páginas que siguen, se expondrán dos experimentos destinados a comprobar ciertos aspectos clave de este modelo. En particular, estos experimentos se centran en las hipótesis planteadas a lo largo de la exposición de la vertiente deductiva del sistema de creencias propuesto.

PARTE III: TRABAJO EXPERIMENTAL

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

Note for the members of the examination board:

In the introduction of the two experiments presented in this thesis, I summarize parts of the model detailed in the Part II of the thesis, written in Spanish. These introductions aim to ensure that all members of the examination board have a comprehensive understanding of the model. However, it is important to note that the essential information for understanding the experiments and their hypotheses is found only in a portion of each introduction. In particular, to understand Experiment 1, one could begin at Heading 3, and for Experiment 2, at Heading 5. The sections preceding these headings provide the necessary context for those who have not had access to the theoretical development in Spanish. If one has had access to it, they may start directly at the indicated headings.

EXPERIMENT 1

Theoretical Foundations

Nearly a decade ago, belief emerged as a burgeoning focus of research across various realms of psychology. While social psychology has historically explored into the study of belief, its predominant emphasis has leaned towards a socio-affective perspective, distinct from cognitive nuances. Significantly advancing our understanding of the interplay between attitude and belief, social psychologists, nonetheless, often sidestepped cognitive intricacies. In contrast, recent alternative models have surfaced such as those presented by Connors and Halligan (2015) rooted in pathological delusion research, and Angel and Seitz (2016) adopting a neurocognitive perspective. Galesic et al. (2021) proposed a multidisciplinary model drawing from social psychology, philosophy, and artificial intelligence models of belief updating formalised by statistical physics. Originating from the complex networks domain, this model stands as the most recent quantitative framework.

This thesis, in contrast, adopts an approach to understanding beliefs and the belief system based primarily on epistemology, argumentation, and rhetoric literature. These fields have long explored the nature of knowledge, opinion, belief, and the dynamics of belief systems. This approach integrates insights from the psychology of reasoning and discourse, aligning with the more recent recognition within these disciplines of the cognitive construct of belief. In previous chapters, we observed a transition in reasoning psychology from a truth-functional logicist formal approach to a broader consideration of how mental processes implement logical rules. While initial research often prioritized the computation of argument validity, recent investigations reveal that humans, in the course of reasoning, incorporate knowledge contents. This knowledge, responsible for recognizing exceptions to general rules, compels a departure from the strict binary world of logic.

Instead, humans consider assertions relying on degrees of belief— a concept central to the New Paradigm in psychology of reasoning.

The New Paradigm, replacing the logical paradigm with a probabilistic one, interprets *degrees of belief* as *subjective probabilities* governed by their own laws of probabilistic validity and coherence. This paradigm shift aligns with a broader Bayesian approach to understanding belief updating. Works in this field have probed the *computation of degrees of belief* when reasoning with conditionals under uncertainty (e.g., Evans and Over, 2004; Oaksford and Chater, 2007; Politzer, 2007; Pfeifer, 2014) and explored argumentation fallacies (e.g., Collins and Hahn, 2018; Corner and Hahn, 2007; Farmer, Warren, and Hahn, 2017; Hahn and Oaksford, 2006a, 2007). Simultaneously, the psychology of discourse processes has made substantial contributions to understanding beliefs and their dynamics. In the theory presented in this thesis which is extensively developed in Part II of this thesis, concepts from discourse psychology, such as situation models and coherence relations, play a crucial role in defining the characteristics of a *mental representation* that qualifies as a belief, as well as determining the nature of the links within the belief network. Concepts like validation processes and epistemic monitoring contribute to comprehending the *mental processes* involved in belief evaluation during argument processing.

While these psychological contributions lay the groundwork for empirical exploration of belief-related questions, the classical deductive side of belief change has yet to be directly studied. This study addresses the previously indicated gap by investigating the extent to which individuals believe something about the real world before and after undergoing manipulations designed to induce change. To comprehend these manipulations, a brief overview of some aspects of the theory developed in this thesis is provided.

1. Fundamental Assumptions of the Belief System

The theoretical framework presented in this thesis, adheres to classical philosophical principles regarding the conception of belief systems and the mechanisms governing belief change. It integrates conceptual and methodological tools derived from psychological theories of reasoning and discourse. In essence, the proposed theory outlines a foundationalist-directed network of beliefs within long-term memory. Representations within this network are interconnected by topical relations (a subset of coherence relations) and refutative relations, with foundational beliefs serving as the bedrock.

Justification flows through topical relations from more justified representations (justified nodes) to less justified representations (candidate nodes), functioning as an *explanans* does for an *explanandum*. Simultaneously, refutative relations place constraints on the justification a given node can receive. They achieve this by linking exceptions, counterexamples, and counterarguments, acting as defeaters to inferences and hindering the facilitation of justification transmission. Defeaters, in this manner, prevent a representation from attaining full justification for an individual.

Then, the amount of justification received by candidate nodes results from the interplay between justified nodes and their defeaters. This combination balances favourable and contrary values, ensuring a consistent outcome. While the belief system inherently tends towards consistency, the potentially limitless number of beliefs it can harbour makes global consistency checking unfeasible. Inevitably, inconsistencies may exist, although their awareness is limited by human cognitive constraints. The cognitive capacity to activate and scan the entire system at once is beyond human capability, given attentional resources and working memory constraints. Hence, inconsistencies are resolved within a localized focus here termed the local belief system (LBS). Routine inconsistency resolution occurs in the LBS by adjusting believability values.

Historically, approaches to belief change in artificial intelligence and psychology have emerged to address detected inconsistencies between new data and elements stored in the cognitive system. The essence of belief change lies in resolving these inconsistencies through one of the next three strategies. First, individuals may delete a previous *link* stored in long-term memory between two representations if this deletion aligns with new data and is not a key element of the system (Ellio and Pelletier, 1997; Dieussaert, Schaeken, De Neys and d'Ydewalle, 2000; Revlin, Cate and Rouss, 2001; Politzer and Carles, 2001). The second strategy involves making minimal changes to one's stock of beliefs to accept incoming information, opting to delete the least connected and incompatible *belief*. This choice ensures that the deletion has minimal consequences on the system's configuration (Alchourrón, Gärdenfors and Makinson, 1985; Johnson-Laird, Girotto and Legrenzi, 2004; Rott, 2008). The third strategy entails creating alternative explanations that *bridge* existing beliefs with seemingly inconsistent new information (Festinger, 1962; Johnson-Laird et al., 2004; Maier and Richter, 2016).

In this model, a complementary approach to belief change aimed at achieving consistency is proposed. Unlike previous strategies where incoming information was taken as undeniable and necessitated modification of the belief system, here, new incompatible information is initially deniable and unbelieved. The existing belief system is then leveraged to render the new information acceptable and believed, reflecting a persuasive-like property inherent in the argumentative model proposed in this thesis.

2. Argumentative and Probabilistic Nature of the Model

The conceptualisation of the belief system in this thesis bears a resemblance to argumentative diagrams found in argument mapping literature, featuring multiple connections to favourable elements and objections. In our model, a representation is considered rationally believed if it has received a sufficient amount of subjective justification. This aligns with the notion that beliefs are justified to an extent that meets the individual's epistemic criteria, and argumentation plays a fundamental role in justifying claims.

Since Toulmin's argumentation theory (1958), topical relations have been recognized as *warrants* facilitating justification transmission⁴⁹. This holds true in the belief system (Moretti and Piazza, 2023). These warrants, or topical relations, encompass various criteria of connection, such as causal and temporal relations, leading to transmission structures known as argumentation schemes, as introduced in Appendix C. These connectors, conceptualized as general rules, follow a conditional internal structure and function as the major premise in arguments (i.e. Dyck, 2002; Toulmin, 1985; Vega, 2004, 2012). Linking a justified node (akin to a *datum* in Toulmin's model) with a candidate node (a *claim* in Toulmin's terminology), these connectors form enthymemes when the premises are accepted and believed, but not the claim initially. Indeed, enthymemes are arguments that build their premises from an individual's beliefs, thereby allowing the modification of the credibility attributed to an idea presented as their conclusion.

The interconnected representations within this system are sufficiently justified, though not always completely due to the presence of defeaters. This implies that justification is a quantifiable magnitude, coming in degrees. A maximal degree of justification indicates absolute subjective certainty about the reality of a representation. When its reality is considered *necessary*, we refer to it as affirmative certainty, whereas when its reality is deemed *impossible*, we speak of negative certainty. Finally, the absence of any justification implies complete uncertainty, where its reality is regarded as *possible and contingent*. Importantly, these degrees of justification and certainty are inherently subjective. As discussed in Chapters 2, pt. I, and Chapter 6, pt. II, subjective probability serves as the quantification method for degrees of justification and uncertainty. In this context, affirmative certainty (considered necessary) corresponds to a subjective probability of 1, negative certainty (impossibility) to a probability of 0, and uncertainty (contingent possibility) to a probability of 0.5. The model asserts that a representation fits into the belief system only when it receives justification corresponding to a subjective probability higher than 0.5.

Reasoning within this framework involves graded probabilistic magnitudes ranging from 0, indicating impossibility, to 1, denoting necessity. Unlike traditional binary logic, this approach engages with subjective probabilities, ensuring the preservation of justification from premises to conclusion. Complete justification for the premises necessarily implies complete justification for the conclusion. However, when premises have only a limited degree of justification, the degree of justification for the conclusion may depend on the individual's prior level of certainty regarding that idea. For example, imagine an individual who initially deems the content of a statement as impossible (e.g. 'Advances in medicine will lead to an increase in people with diseases.'). Also, suppose this person strongly believes in the content of two other independent sentences (e.g. 'Advances in medicine are causing an increase in the number of elderly people' and 'if the number of elderly people increases, then the number of people with diseases increases'). If an argument can be constructed such that its conclusion is the initially rejected statement, using those two sentences as premises (i.e., an enthymeme), belief in the statement could increase. Since

⁴⁹ Note, in any case, that this is a simplification. Although topical relations act as warrants, there are differences between these relations and Toulmin's warrants.

conclusions are based on their premises, the local belief system should adjust the credibility of the conclusion to maintain coherence. Conversely, an individual uncertain about the claim may accept its possibility up to a certain degree, say 0.4. If the interlocutor introduces an argument with premises deemed impossible by the hearer, the hearer may either ignore the premises, thus maintaining their a priori credibility, or attend to the premises and decrease their credibility in the claim.

3. A Rhetorical Approach to Credibility

In the first experiment presented in this thesis, a rhetorical approach has been employed. Before delving into the details of the experiment, it is essential to outline key concepts related to rhetorical communicative intercourse.

3.1. Rhetorical Communicative Intercourse and Doxastic Plausibility

Rhetorical situations involve three primary entities: the speaker, an audience, and a context. Oral verbal communication is a cooperative activity wherein the speaker endeavours to convey an idea, or message, from his/her mind to the minds of the listeners through spoken words, sentences, and discourses, all within a specific context. During the decoding process, each listener activates relevant concepts from their long-term memory, along with highly associated features and other concepts, leading to subtle differences in interpretation. Background knowledge and the context's mental model of the communicative situation significantly influence this interpretation. To ensure effective communication, speakers tailor their discourse to the audience's knowledge, incorporating shared knowledge about the situation. They initiate the discourse with information shared and known by the audience, progressing to new information.

A rhetorical situation is one among various communicative scenarios. The primary objective of rhetoric is to persuade the audience about an idea or action, often both. In this context, the goal is not merely comprehension but belief. Therefore, speakers consider what the audience already knows and, more crucially, what they already believe. It is a fundamental criterion in argumentation that *the premises must be more believed than the conclusion initially* – a principle we refer to as the *standard of argumentation*. This criterion, attributed to Aristotle over two millennia ago, was originally articulated as the idea of being “more plausible than”. As discussed in Chapter 1, pt. I, in Aristotle's dialectical and rhetorical theories, “the plausible” denote what is considered probable, not only for an individual but for the majority or the wise. Among the wise, it encompasses the opinions of all, the most, the better known, and the most reputable. Vega (1998, 2012) refers to this Aristotelian argumentative concept of shared belief as *endoxastic* plausibility. This criterion is particularly crucial for a public speaker aiming to construct a persuasive discourse for a large audience, with a focus on delivering a one-sided presentation. Therefore, the rhetorical speaker must utilize the audience's shared beliefs as premises when constructing enthymematic arguments, ensuring the believability of the premises is transmitted to the conclusion.

3.2. The Structure of Enthymemes: Terms, Premises, and Conclusions

Enthymemes, as explained in sect. 1.2. chap. 1, and app. C, are syllogistic structures typically composed of three terms and two premises, although additional premises and terms are possible. Traditionally, these three terms are identified as the subject (S), the middle term (M), and the predicate (P). The middle term is repeated in the two premises but absent from the conclusion. Functionally, the middle term unifies S and P in the conclusion. The syllogisms can be categorized into different *figures* based on the position of the middle term. For the purposes of this thesis, only first-figure syllogisms without quantifiers or negations will be considered.

The minor premise of an enthymeme refers to a fact or data about the world known or highly believed by the general audience. This minor premise encompasses S and M, with M serving the role of representing a potential cause. The major premise, on the other hand, constitutes a general topical relational rule describing how things in the world typically occur, or at least how most people perceive them to occur. The chosen topical relation for the experiments in this thesis is causality, embodied in the “from cause to effect” argumentation scheme, presented in the form of a conditional (if C, then E). Here, the antecedent (C) signifies a potential cause, with the consequent (E) representing a proposed effect. As detailed in ap. C and is mentioned in some sections of Part II, causality holds a privileged place in human cognition, playing a pivotal role in conditional reasoning, argumentation, discourse coherence, plausibility, and explanation theory.

Table 5 (below) illustrates this first-figure structure, along with specificities and a concrete example relevant to this study. In this table, the structure of the traditional first-figure syllogism is presented on the left, alongside its adaptation/conversion to the type of arguments that will be used in this study. Below, an example of an argument is given. This argument will only be enthymematic if the reader believes in the content of the premises and, initially, did not believe in the content of the conclusion. Regarding its structure, it can be seen how the major premise consists of a conditional warrant that entails a causal relationship, while the minor premise includes data (reported in the media at that time) involving specific agents (Ag) from a specific place or context (Cx). Finally, a controversial and generally disbelieved claim or conclusion is presented, as the concept of a farmer or shepherd is not commonly associated with that of an arsonist.

For an enthymeme to be endoxastically plausible, both premises must be accepted and shared by the majority of the audience. The claim, formed by S and P, should be contentious and not readily believed as possible, or at least less so than the premises.

Table 5
Structure of Arguments of the Study and an Illustration

Syllogism 1 st Figure				Enthymematic Argument			
Major Premise	M	is	P	Warrant	Cause	→	Effect
Minor Premise	S	is	M	Data	Ag/Cx	is	Cause
Conclusion	S	is	P	Claim	Ag/Cx	is	Effect

Warrant	If the forests are set on fire, then the land will become desertified.
Data	The Asturian farmers set fire to the forests.
Claim	The Asturian farmers will desertify the land.

Note: This argument was constructed in 2016. Nowadays, this argument would not be effective, as there are more accessible alternative causes for land desertification (though the truth of the premise). This example also illustrates that the credibility attributed to the content of a sentence is entirely circumstantial.

The use of three-term syllogisms represents an advancement in the study of belief and argumentation in psychology, setting itself apart from traditional two-term arguments by enabling a clearer distinction between the cause and its agent or specific context. This methodology enhances the "situationality" of arguments, rendering statements less arbitrary by anchoring them in familiar contexts, which directly influences the credibility and subjective probability attributed by individuals. Specifying agents and contexts allows for more nuanced and accurate evaluations of claims, based on the subjects' world knowledge. This contributes to a deeper understanding of how beliefs are formed and adjusted within a coherent belief system.

4. Aims and Hypotheses of this Study

The primary goal of the first experiment in this thesis is to empirically assess the argumentative nature proposed in the belief dynamics framework. Specifically, we aim to evaluate the effectiveness of enthymemes in influencing the believability of contentious assertions, referred to as *claims* or critical assertions. This involves manipulating the believability of the conditional major premise (the *warrant*) while maintaining consistently high believability of the minor premise (the *data* premise). The objective is to confirm the assumed process of justification transmission from the data premise to the claim through the warrant, as posited by the theoretical framework. Specifically, we seek to determine whether shared and accepted warrants contribute to increase the subjective probability of the claims compared to unshared and unaccepted warrants. Therefore, in a valid argument with a widely accepted and shared data premise, its credibility will be transferred to the claim through the warrant, but only if the warrant is also highly accepted and shared.

The Present Study

1. Objectives, Variables, and Design

The experiment outlined here explores how argumentation influences the perceived credibility of critical claims by analysing the impact of two independent variables: the presence/absence of 'argumentative context' and the 'strength of association' between causes and effects in warranting premises. The argumentative context was manipulated by presenting critical claims either in isolation or preceded by premises, therefore, in the absence or presence of such context. This manipulation allows us to test if premises can cause variations in the credibility of critical claims when presented as conclusions of an argument.

The strength of association was our key variable for manipulating plausibility and, consequently, the credibility of the warrants. Plausibility was determined based on the proportion of people within an epistemic community who recognise a specific causal relationship; the higher the proportion, the more plausible the relationship is considered. For example, between two potential effects of the same cause, the effect more frequently linked by participants is deemed stronger and more plausible than the less mentioned one. The strength of association was measured based on how many people link specific causes and effects (e.g., Quinn & Markovits, 1998). This allowed us to classify effects as weakly (W) or strongly (S) associated with their cause, thus varying the credibility of the warrants in accordance with the recognition proportion of their causal connection. With this, we hope to verify if the strength of association of the warrant influences the credibility that can be transferred to the conclusion.

An experimental design with a control group and pretest-posttest stages was used. The pretest phase served as the initial evaluation of the credibility (a priori) of the critical claims presented in isolation, measuring their subjective probability before introducing any argumentative context. This is the condition of 'absence' within the argumentative context. The posttest phase was set as the core of our study: here, the claims were presented as conclusions of arguments, i.e., preceded by premises, to determine the a posteriori credibility. This is the condition of 'presence' within the argumentative context. The premises were differentiated by the strength of association of the warrant (strong or weak), while the data premise remained constant at high levels of credibility. This ensures that the observed differences are due to the manipulation of the credibility of the warrants. This approach focuses on the posttest to evaluate how participants' perception of credibility towards the same claims changes following exposure to arguments with credible premises. This change would reflect the transfer of credibility from the premises to the conclusion through the warrant, constituting the most critical and revealing aspect of our research.

In our study, we first assessed the initial (a priori) credibility of the critical claims without argumentative context during the pretest phase. Later, in the posttest phase, we measured their credibility again (a posteriori), but this time in the presence of an argumentative context. It is crucial to ensure that any observed difference in the posttest phase is owed solely to the introduction of

such context, and not merely to the passage of time. External factors, such as news or personal experiences, could influence participants' credibility perceptions between the pretest and posttest. To control these effects and provide validity to our findings, the experimental design included a control group that evaluated the critical claims in the same isolated conditions (i.e., without argumentative context) both in the pretest and posttest phases.

2. Material Generation

For the implementation of the experiment, it was necessary to generate a large number of experimental materials that met the established criteria. Below, the process of creating data premises, warrants, and critical claims is detailed.

Generation of Data Premises

For the experiment, it was necessary to develop materials that met specific criteria in terms of both plausibility and relevance. The data premises were based on information widely disseminated in media and social networks, selecting facts and opinions commonly accepted to reflect prevalent perceptions in society.

The constructed sentences contained a grammatical subject, which in some cases was preceded by a locative complement, and an orational predicate. The referents of the grammatical subject and, where applicable, of the locative complement, were elements *known* by the members of the epistemic community. This allowed for genuine credibility to be attributed to the content of these sentences. On the other hand, the orational predicates needed to function as potential causes of future events.

The following examples illustrate the type of statements used as data premises:

- Many young Spaniards frequently use drugs.
- In our society, many people choose to become first-time mothers at the age of 40.

These are two examples of news contents which frequently appear in the media, featuring a recognizable subject (sometimes preceded by a locative phrase) and a predicate that could function as the cause of something in the future. A total of 50 data premises were constructed.

Generation of Warrant Premises

In developing the warrants, we used responses from a preliminary sample of participants obtained through a questionnaire designed specifically for this purpose. This strategy ensured the relevance and plausibility of the materials generated. As detailed in App. C, warrants typically take a canonical form implicit in the conditional sentence structure 'if *p*, then *q*' (where '*p*' is the antecedent and '*q*' the consequent).

A crucial aspect in the generation of the warrants was ensuring that the experiment participants perceived the contents as causal relationships. For this purpose, we designed questionnaires that presented potential causes (the predicates of the data premises) in the form of antecedents of conditional sentences. For example:

- If someone frequently uses drugs...
- If one opts to become a first-time mother at 40...

Participants were asked to generate, in a limited time, as many consequent effects as possible⁵⁰. This acquisition of the consequents ensured the perception of causality of the warrant premises. For each cause proposed as an antecedent, an effect mentioned by many participants was considered strongly associated with its cause, and one infrequently mentioned, weakly associated. The former constituted warrants of strong association (S), and the latter, of weak association (W). Furthermore, this method enabled variations in the plausibility level of the warrants. By selecting effects based on their mention frequency, we determined their degree of association and plausibility. The more frequently mentioned effects reflect the consensus of the majority, making them more plausible. From this perspective, the strength of association and plausibility are interconnected through the frequency of mentions, suggesting that a higher frequency may indicate a more robust perception of a causal relationship. Thus, the collective consensus is equated with the perceived credibility of the warrants. This methodology is inspired by previous research, such as that of Quinn & Markovits (1998), who also based their studies on the frequency of mentions to measure the strength of association, though without references to plausibility. This approach has allowed us not only to discern the strength of association based on how frequently effects are mentioned, but also to directly relate this frequency to perceived plausibility. Consequently, an effect that is frequently mentioned is considered more plausible and more strongly associated with its cause.

Method

Participants: A sample of 80 second-year psychology students voluntarily participated in this material generation task. Their average age was 19.62 (SD = 6.74).

Procedure: The 50 potential causes were divided between two paper questionnaires (available in App. A-I), each containing 25 causes. Different versions of each were created to balance the order of presentation of the causes. The material generation task was conducted in a single session with all students in a classroom. The two versions were alternately distributed to ensure that each participant responded originally to their set of questions. Once the questionnaires were distributed,

⁵⁰ This method is similar but opposite to that employed by Quinn & Markovits (1998), where participants were shown a set of effects and had to write down as many causes as they could think of in 30 seconds. By doing so, they could identify highly associated and weakly associated causes for the presented effects (for more details, see sect. 15.2., chap. 4, pt. I).

the experimenter signalled the start, timed each item (one minute), and then signalled for everyone to move to the next item simultaneously.

Results

Responses were grouped into clusters of meaning for each item, identifying the most recurrent effects among participants. Multiple effects with low frequency were observed, from which the least credible as a predicate for the future critical claim was selected. This process enabled the identification of effects strongly (S)and weakly (W) associated with the same cause, resulting in S and W warrants. For example:

- If someone frequently uses drugs, then:
 - o They will end up with irreversible brain damage (S).
 - o They will develop new cognitive skills (W).
- If one opts to become a first-time mother at 40, then:
 - o The likelihood of having children with deficiencies increases (S).
 - o The amount of spoiled children increases (W).

The generated materials had to meet a series of requirements to be included in the experiment. A key criterion was that the most frequently mentioned effect of a warrant reached at least 50% of the frequency of the most mentioned effect overall. Since the most mentioned effect in the questionnaires was by 45% of the participants, this meant that the threshold for selecting a warrant required its most frequent effect to be mentioned by at least 22.5% of the participants. Some warrants did not generate S effects and were eliminated, along with their respective data premises.

Generation of Conclusions (Critical Claims)

With both premises constructed, we proceeded to formulate the conclusions or critical assertions derived from their combination. The premises were organized into first-figure syllogisms, where the predicate of the major premise was also the predicate of the conclusion. Because of this, manipulating the warrants into S (strong) and W (weak) effects resulted in a duplication of arguments: each data premise led to an argument with an S effect in both the warrant and the conclusion, and another with a W effect in both the warrant and the conclusion.

For example:

Strongly Associated Effect	Weakly Associated Effect
If someone frequently uses drugs, then they will end up with irreversible brain damage.	If someone frequently uses drugs, then they will develop new cognitive skills.
Many young Spaniards frequently use drugs.	Many young Spaniards frequently use drugs.
Many young Spaniards will end up with irreversible brain damage.	Many young Spaniards will develop new cognitive skills.

For simplicity, from now on, we will refer to the critical assertion as 'conclusion,' both before (isolated conclusion) and after its argumentation (simply, conclusion), although technically, it only becomes a 'conclusion' when it is part of an argument. Furthermore, the control group will only be presented with isolated conclusions.

The resulting conclusions needed to be initially disbelieved when presented in isolation. Those that, according to the researcher's intuition, appeared highly credible were eliminated along with the complete argument in both its S and W versions. This led to a reduction from 50 to 31 data premises, generating a total of 62 arguments (31 S and 31 W).

3. Pretest Phase: Evaluation of Isolated Conclusions and Data Premises

In the previous section, we saw that each data premise was associated with two conclusions (one S and one W), totaling 62 conclusions (31 S and 31 W). To prevent the same participant from evaluating both S and W versions of conclusions corresponding to arguments with the same data premise, the following procedure was implemented: two sets of complementary items (A1 and A2) were created, where one set contained 15 S items and 16 W items, and the other contained 15 W items and 16 S items. The participant sample was divided into two groups (G1 and G2), assigning each group one of the item sets (A1 or A2). Thus, the control group was also divided into two (C1 and C2) so that they received the corresponding item sets (A1 and A2) in line with the experimental groups (G1 and G2). Therefore, control group 1 (C1) and experimental group 1 (G1) evaluated the items from set A1, while control group 2 (C2) and experimental group 2 (G2) evaluated the items from set A2. This distribution was also maintained in the posttest phase. Hence, with different participants evaluating different (yet complementary) conclusions, when analysing the results, it will be necessary to treat it as two separate experiments. Notwithstanding, a joint analysis of the two groups was also conducted and can be consulted in App. A-IX.

The pretest phase consisted of two tasks. The first task was to evaluate conclusions and was performed by both groups, that is, the experimental and the control groups. The second task was to evaluate data premises and was carried out only by the experimental group.

Participants: Sixty-five first-year psychology students voluntarily participated in the experimental group, divided into G1 (n=33) and G2 (n=32). Another forty-two students, also volunteers, formed the control group, distributed into two subgroups of 21 participants each (C1 and C2). The average age of all participants was 18.7 (SD = 5.24). None of the participants had prior training in logic, and all were native Spanish speakers. They were offered a small bonus in the final grade of a course, as well as a cafeteria voucher, as an incentive for their participation providing they also attended the posttest.

Materials: In the first task, each group (G1 and C1, as well as G2 and C2) evaluated 31 isolated conclusions. Given that low initial scores of subjective probability were expected for these

statements, 20 filler sentences with high intuitive credibility were included to prevent bias in the distribution of credibility scores. Without clearly credible sentences, conclusions that had medium-low credibility when isolated could have been rated as more credible. By introducing these filler items, participants could notice some variability in the credibility of the sentences. These sentences maintained the same structure and content as those corresponding to the isolated conclusions. For example:

- Next year, the Sagrada Familia in Barcelona will still be under construction.

All items for this task are included in App. A-II.

Similarly, in the second task, each experimental group evaluated 31 data premises. Since high initial scores of subjective probability were expected for these phrases, 20 filler sentences with low intuitive credibility were incorporated to prevent biases in the distribution of credibility scores. For example:

- Smoking is a cultural movement aimed at saving human lives.

All sentences for this task are detailed in App. A-III.

The Embedded Figures Test (Witkin, Oltman, Raskin, and Karp, 1971) was also used as a pencil-and-paper distractor task. It is a non-verbal task in which participants had to find hidden figures masked within other figures.

Procedure: The experimental group carried out the two tasks of evaluating subjective probability using the DMDX software (Forster and Forster, 2003). The first task involved evaluating isolated conclusions, while the second focused on data premises. Each task used a separate file, each containing 51 items that included the critical sentences for evaluation along with their respective fillers. Participants used an 11-point scale, ranging from 0 (not probable at all) to 1 (completely probable), with increments of 0.1, to estimate the probability of the situations presented in each sentence. Between the evaluation tasks, a distractor task using the Embedded Figures Test was interspersed to separate the evaluations and minimize biases.

Before starting, detailed instructions were provided to the participants assigned to the experimental group (see App. A-IV). They were instructed that the scores assigned should reflect the degree of credibility they attributed to the situations described in the sentences, without imposing any restrictions on how to interpret them. This allowed them to consider any relevant aspect during the task. The instructions also included a guide on the structure of the experiment, informing participants that, upon reaching the halfway point of each evaluation task, they would be asked to perform a pencil-and-paper task before continuing with the rest of the evaluations.

Each sentence remained visible on the screen until the participant pressed the keyboard button corresponding to the degree of subjective probability that best represented its credibility, as follows:

Please estimate the probability of the following situation occurring:

Many young Spaniards will end up with irreversible brain damage.

The experiment was conducted in groups of four participants. Upon arriving at the laboratory, they entered their identification code, and the experimenter loaded the file for the isolated conclusion evaluation task onto the computer. After reading the instructions, any questions were clarified, and 6 practice trials were conducted. Then, the experimenter left the room, and participants began this first task. After this block, participants performed the distractor task. Once completed, the experimenter loaded the files to carry out the data premise evaluation task and gave the start command. Both the presentation of the sentences and the recording of the evaluation of subjective probability were done using the DMDX program.

Participants in the control groups evaluated the critical sentences (i.e., the isolated conclusions) and the filler sentences using paper questionnaires (available in App. A-V), with instructions to estimate the degree to which they believed the situations described in the sentences could be (or are) true, just like the experimental group. However, unlike the experimental groups, the control groups did not evaluate sentences related to the data premises, mainly because the study's interest focused on the credibility of the critical claims, rather than on the premises. The aim was to ensure that any external influence did not affect the dependent variable of interest, which is the credibility of the conclusions; these will be compared in the analysis. Additionally, there is a remote possibility that knowing the data premises could influence the scores that participants in these control groups assign to the critical claims during the posttest phase.

Control groups C1 and C2 evaluated the sentences using the same 11-point scale mentioned earlier. Two versions of the questionnaire were developed for each control group, with the sentence presentation order randomised in each, to ensure procedural consistency between both groups. These questionnaires were administered simultaneously to all the participants in a classroom.

To encourage honest and uninfluenced responses, all evaluation tasks conducted by the experimental and control groups were carried out anonymously. Additionally, to allow for data linkage between the pretest and the posttest, each participant was provided with a unique identification code.

In this phase, an additional selection of items was made based on the standard of argumentation, which requires that the premises be considered more credible than the conclusions a priori. For an item to be selected, the credibility assigned to the data premise had to exceed that of its corresponding isolated conclusion. The average credibility obtained for the data premises and isolated conclusions of each item was compared. If the subjective probability of a data premise was equal to or less than that of its isolated conclusion, the item was removed in both versions (S and W)⁵¹. At the end of this process, 12 items meeting this criterion were retained.

⁵¹ There was one exception to this rule: if in one of the two groups the credibility of the data premise was equal to that of the critical claim, and in the other group the credibility of the data premise exceeded that of the critical claim by at least 3 points, then the item was retained in both the S and W versions.

4. Posttest Phase: Evaluation of Warrant Premises and Conclusions

In this phase, the experimental group (i.e., G1 and G2) assessed the warrant premises in the first task and the critical claims as conclusions of complete arguments in a second task. Conversely, the control group (i.e., C1 and C2) re-evaluated the isolated conclusions, just as in the pretest phase. The interval between the pretest phase and the current posttest phase was two months, ensuring that participants did not remember their previous scores for these items (Ferrando, 2003b).

Participants: Fifty-five first-year psychology students voluntarily participated in the experimental group, divided into G1 (n=30) and G2 (n=25). Ten participants from the pretest phase did not attend this posttest phase. The average age of both groups was 18.5 years (SD = 7.32). Another thirty-eight students, also volunteers, made up the control group, divided into two subgroups: C1 (n=20) and C2 (n=18). Three participants from the pretest phase did not attend this posttest phase. The average age for both groups was 19.02 (SD = 9.28).

Materials: For the first task of evaluating warrants performed by the experimental group, paper questionnaires⁵² were used, differing for G1 and G2, featuring 12 conditional sentences for assessment. To clarify the content of the questionnaire, let's say that in G1, sentences 1 to 6 presented warrants of strong association (S), while sentences 7 to 12 presented warrants of weak association (W). In G2, the complementary versions to those in G1 were presented: W from 1 to 6 and S from 7 to 12. However, it is important to note that the actual order of items with weak and strong association was randomly intermixed (see samples of the G1 and G2 questionnaires in App. A-VI). Furthermore, from this random order, twelve versions of each questionnaire were created, where the twelve sentences rotated their position: in version one, item one was first, in version two, item twelve was first and item one was second, and so on.

In the second task of evaluating conclusions, also conducted only by the experimental group, 12 arguments were used, each consisting of a warrant premise (evaluated in the previous task), a data premise, and a conclusion. Of these 12 arguments, 6 contained an S warrant premise and the other 6 a W warrant premise (see materials in App. A-VIII). The data premise was identical in the S and W version of each item across G1 and G2. The materials for the task performed by the control group (C1 and C2) consisted of all the sentences evaluated in the pretest phase, which included the 12 conclusions (always in isolation) and their respective filler items, using the same paper

⁵² For objectives not directly related to the central idea of this research, two methods were used to evaluate the credibility of the warrants: direct scoring and contingency tables. The contingency table method, also known as the 'truth table task,' has commonly been used to indirectly measure the implicit subjective probability of conditional statements (e.g., Over & Evans, 2003; Over et al., 2007; Skovgaard-Olsen et al., 2023). It divides the conditional into its four logical possibilities, allowing for the calculation of conditional probability. However, in this study, we will focus primarily on direct scoring as it is a more natural and intuitive method, suitable for a wider variety of statements. Additional details about the use of contingency tables in this study are available in App. A-X for those interested in delving deeper into this aspect. The complete version of these questionnaires using both methods is presented in App. A-VI.

questionnaires as before. These conclusions corresponded with those seen by G1 and G2, respectively.

Procedure: For the experimental group (G1 and G2), the experiment was conducted in subgroups of one to four participants. They were explained that they would participate in an informal reasoning experiment where they would estimate subjective probabilities about specific situations. It was emphasised that their estimates should be based on their knowledge and beliefs and that there were no right or wrong answers. They were also informed about the structure of two separate tasks.

In the *first task*, participants in the experimental group estimated the subjective probability of the situations described in 12 conditional sentences, which would serve as warrant premises in the arguments of the second task. Upon entering the laboratory, the corresponding questionnaires for each group (G1 or G2) were provided to evaluate these warrant premises. After entering their identification code on the questionnaire, participants read the instructions in detail. They were required to directly assign a subjective probability to each described situation, using a scale from 0 to 100% to estimate the percentage of confidence in the future occurrence of these events. For example, assigning 80% to the statement "If someone frequently uses drugs, then they will end up with irreversible brain damage," would indicate that they *consider* the situation to have an 80% likelihood of occurring (specific instructions are found in the questionnaires referenced in App. A-VI). Any questions were resolved before beginning the task, and items were presented in a rotating order among the questionnaires to minimise potential biases. After completing the task, participants were offered a short break, after which the experimenter provided them with vouchers for the cafeteria.

In the *second task*, participants in the experimental group evaluated the subjective probability of the situations described in the 12 selected critical assertions. These were presented on the computer screen as conclusions of complete arguments. These arguments included premises that had been evaluated in previous tasks, allowing participants to consider the full credibility of the argument when estimating the subjective probability of each conclusion. For this, we input the 12 arguments for each experimental group (G1 and G2) into the DMDX program (Forster and Forster, 2003), which allowed for the randomization of the presentation of arguments to each participant. The arguments were displayed on the screen sentence by sentence, each unfolding as the space bar on the keyboard was pressed. The sentences of the argument always appeared in the same order: first the warrant premise, followed by the data premise, and finally the conclusion. After receiving detailed instructions about the task, which included the structure and purpose of the evaluation, all participant questions were addressed, and they then began with four training items. They were reminded that their estimation of the third sentence (i.e., the conclusion) should be made considering the context provided by the two preceding sentences (i.e., the premises). Each sentence remained on the screen until the participant pressed the key chosen for their probability estimation.

After completing the examples, the experimenter ensured that the participants had understood the task and then left the room.

Participants in the experimental group were then required to estimate the probability that the situations described in the conclusions would occur. For this, they pressed the keyboard button corresponding to the degree of subjective probability that best represented their credibility, using the same 11-point scale employed in the pretest. Specific instructions for this task are presented in App. A-VII. Thus, the presentation of the arguments was as follows:

Given that:

If someone frequently uses drugs, they will end up with irreversible brain damage.

Many young Spaniards frequently use drugs.

Please estimate the probability of the following situation occurring:

Many young Spaniards will end up with irreversible brain damage.

Meanwhile, the task performed by the control group (C1 and C2) was the same as in the pretest. They were all gathered in a classroom and given the paper questionnaires to evaluate the situations described in the sentences. Just like in the pretest phase, the participants responded on the same 11-point scale.

5. Recap of Objectives and Predictions

This study assessed the influence of enthymemes, arguments that use prior beliefs as premises, on the credibility of initially less credible conclusions. The experiment explored how the manipulation of the credibility of warrant premises and the variation of the argumentative context affect the credibility of critical claims. Two independent variables were adjusted: the strength of the association between causes and effects in the warrants, and the presentation of conclusions in isolated or complete argumentative contexts. The experimental design included a pretest and a posttest, where the pretest represented the absence of argumentative context and the posttest its presence, allowing conclusions to be presented as the result of arguments with either strong or weak association warrants. A control group was included to ensure that the observed effects were solely due to the manipulated variables; this group evaluated the conclusions in isolation, both in the pretest and the posttest. The study aimed to determine if an argumentative context with plausible warrants allows the transfer of credibility from the data premise to the conclusion, thereby enhancing its credibility.

Regarding the main objective, we anticipate observing an interaction between the two independent variables. In the posttest, in the presence of the argumentative context, an increase in the credibility of conclusions derived from S arguments, which have a warrant with a strongly

associated effect to its cause, is expected. This would be the only expected difference in terms of an increase in credibility. Scores for conclusions from arguments with W condition warrants are not expected to vary between the posttest and the pretest. They are also not expected to have scores differing from those observed in the control group. A W warrant may hinder the transmission of credibility from the data premise to the critical claim. However, this does not necessarily imply a decrease in their credibility compared to before the presentation of the premises. It is plausible that they experience a decrease in credibility only if the individual relies more on the reasons provided in the premises than on their own prior reasons.

Therefore, before the introduction of the argumentative context, no difference in credibility scores between the control and experimental groups for isolated conclusions labelled as S or W is expected, since no manipulation has yet been applied. Subsequently, when comparing credibility scores before and after the introduction of the argumentative context, an increase in the S condition and possibly a decrease in the W condition are expected (depending on whether participants heed or ignore the premises). However, no difference in the scores of the control group is expected between before and after. Furthermore, focusing on the time when the argumentative context is present, differences are only expected in the experimental group. Specifically, an increase in credibility is expected for items in the S condition, and possibly, although not necessarily, a decrease in the W condition. There should be no differences between the control group scores for the S and W items, as this group has not received any treatment. These predictions apply equally to both G1 and G2 groups of this study, as well as their corresponding C1 and C2 groups.

6. Results

Repeated measures ANOVAs were conducted with 2 (Argumentative Context: Absent (pretest) vs. Present (posttest)) x 2 (Strength of Association: Weak vs. Strong) x 2 (Group: Experimental vs. Control) factors, where the first two variables are within-subjects and the last is between-subjects. Because both G1 and G2 only had 12 items (6 S and 6 W), no item-level analysis was conducted due to a lack of statistical power. Therefore, all analyses presented below are by subjects.

First, the means and standard deviations of the subjective probability scores of the premises and conclusions in the pretest and posttest phases obtained by the experimental and control groups in all conditions are presented (see Table 6, below).

The scores at the top of the table correspond to the experimental groups (G1 and G2), where the W and S columns indicate the strength of the effect's association with its cause in the warrants of the arguments. In this part of the table, you can observe:

- a) The scores for the data premise.
- b) The scores for the warrant premise measured using the direct scoring method.

- c) The average scores of the conclusions prior (pretest) and post (posttest). In the experimental group, 'prior' corresponds to the condition of absence of argumentative context and 'post' to the presence of such context. For example, a score in the W column and in the 'prior' row refers to the assessment conducted by an experimental group on a set of items designated to receive a W argumentative treatment, but before this was applied (i.e., in the pretest phase; that is, in the absence of the argumentative context).

The bottom part of the table shows the scores for isolated conclusions given by the control groups (C1 and C2) in the pretest and posttest phases. For example, the 'prior' score in the W column of C1 refers to the evaluation conducted by that control group in the pretest phase on the same set of items evaluated by group G1 in that phase.

Table 6

Subjective Probability Scores of the Premises and Critical Assertions

Experimental	G1		G2	
	W	S	W	S
Warrant	.50 (.11)	.69 (.12)	.54 (.15)	.63 (.13)
Data Premise	.76 (.10)	.75 (.12)	.78 (.14)	.86 (.09)
Conclusion				
Prior	.56 (.14)	.49 (.12)	.49 (.13)	.51 (.13)
Post	.47 (.12)	.64 (.12)	.49 (.13)	.66 (.11)
Control	C1		C2	
	W	S	W	S
Conclusion				
Prior	.52 (.06)	.50 (.17)	.44 (.18)	.50 (.06)
Post	.50 (.10)	.48 (.15)	.46 (.16)	.49 (.10)

It was previously established as a prerequisite for the feasibility of the experiment that the subjective probability scores associated with S warrants were higher than those for W warrants. Upon examining the corresponding mean comparisons, this condition was consistently met (all $ps < .001$). Both the means shown in Table 6, as well as in Figures 13 and 14, display a general pattern of results highlighting the increase in subjective probability that occurs in the strong association condition in the experimental group in the posttest phase. A decrease in the scores can also be observed in the weak association condition of the experimental group G1. Finally, it also appears that the control group does not differ in its scores, neither between the conditions of argumentative context (which it does not receive), nor between conditions of strength of association (which it also does not receive). Next, we will present the results of the study, focusing on the transmission of credibility through the warrant, starting with group G1 and then G2.

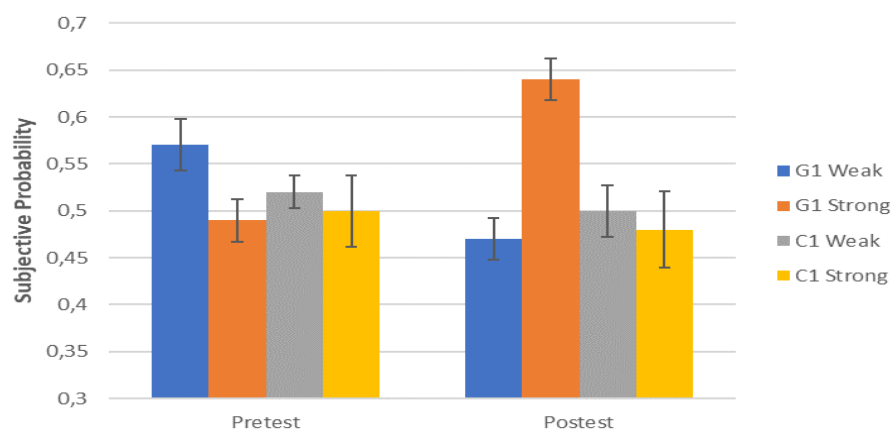
6.1. Results of Group G1 and C1 on the Transmission of Credibility

The general pattern of results is shown in Figure 13 (below). To test the main hypothesis of the experiment, a 2 (Argumentative Context: Absent (pretest) vs. Present (posttest)) x 2 (Strength of Association: Weak vs. Strong) x 2 (Group: Experimental vs. Control) ANOVA was conducted.

The analyses revealed an absence of main effects for the variables of argumentative context timing, strength of association of the warrant (both $F_s < 1$), and group ($F(1, 48) = 2.52, p = .119$).

Figure 13

Subjective Probabilities Given to the Critical Assertions by G1 and C1



Note: Mean subjective probabilities scores given by G1 and C1 as a function of argumentative treatment (prior – posterior) and the warrant association force (Strong or Weak). The error bars represent standard error.

Furthermore, no interaction was observed between the variables of argumentative context timing and group ($F(1, 48) = 2.43, p = .125$), nor between strength of association and group ($F(1, 48) = 2.70, p = .106$). However, the interaction between the variables of argumentative context timing and strength of association was significant ($F(1, 48) = 25.68, p < .001, \eta_p^2 = .349$). Post-hoc analyses were conducted using Bonferroni-adjusted paired t-tests. No significant differences were found in strength of association in the pretest (e.g., $p > .029$)⁵³, but there were significant differences in the posttest ($t(49) = -3.86, p < .001; d = .760$). In the posttest (i.e., in the presence of the argumentative context), the average subjective probability assigned to the conclusions in the W condition ($M_{Weak} = .48$) was significantly lower than that obtained in the S condition ($M_{Strong} = .58$). Significant differences were also observed between the pretest ($M_{Pretest} = .55$) and the posttest ($M_{Posttest} = .48$) in the W condition ($t(49) = 3.75, p < .001; d = .608$), as well as in the S condition ($t(49) = -3.73, p < .001; d = .620; M_{Pretest} = .49, M_{Posttest} = .58$). While scores decreased in the W condition from pretest to posttest, an increase was observed in the S condition.

⁵³ Given the adjustment by the Bonferroni method, this difference was not considered significant.

The analyses also showed a three-way interaction among the group variables, argumentative context, and strength of association ($F(1, 48) = 23.65, p < .001, \eta_p^2 = .330$). Post-hoc analyses were performed using t-tests adjusted by the Bonferroni method. In the pretest phase, no significant differences were found between the experimental and control groups in the weak association condition ($t(42.44) = 1.56, p = .126$)⁵⁴, nor in the strong association condition ($t(48) = -.17, p = .863$). In the posttest phase, there were no significant differences between the experimental and control groups in the weak association condition ($t(48) = -.837, p = .407$). However, there were significant differences between the means of both groups in the strong association condition ($t(48) = 3.95, p < .001; d = 1.18$), where the experimental group's mean was higher than that of the control group. Moreover, in the experimental group, there were differences in the means obtained between the pretest (absence of argumentative context) and the posttest (presence), both in the weak association condition ($t(29) = 4.01, p < .001; d = .766$) and in the strong association condition ($t(29) = -5.13, p < .001; d = 1.25$). In the weak association condition, the mean decreased from the pretest to the posttest. However, in the strong association condition, the mean increased. In the control group, no differences occurred between the pretest and posttest conditions, neither when the claims belonged to the weak association condition ($t(19) = .89, p = .385$) nor when they belonged to the strong association condition ($t(19) = .60, p = .557$).

In addition to these interactions, a separate ANOVA was conducted for the experimental group with the objective of observing the effects of the manipulated variables without the influence of the control group ratings. This analysis revealed a significant main effect of association strength ($F(1, 29) = 5.04, p = .032, \eta_p^2 = .148$) and, crucially, the interaction between argumentative context moment and warrant association strength ($F(1, 29) = 47.84, p < .001, \eta_p^2 = .623$).

Bonferroni adjusted post-hoc analyses showed that there were significant differences between the weak and strong association condition at the time both in the absence of the argumentative context, i.e., in the pretest ($t(29) = 2.59, p = .015; d = .588$) and in its presence in the posttest ($t(29) = -6.84, p < .001; d = 1.41$). A higher subjective probability was assigned to findings in the strong association condition compared to the weak association condition.

An equivalent analysis was performed in the control group and neither main effects nor interactions between variables were observed.

6.2. Discussion of the Results for Group G1

The results of Group G1 support the central hypothesis of the study. It was expected that the credibility of critical sentences in the experimental group would increase after the presentation of the argument and only if its warrant had a strong association between the effect and its cause. This prediction was clearly confirmed by the significant interactions observed.

⁵⁴ The Levene's test for equality of variances showed that the variances were different ($p = .002$) in the weak condition. Therefore, corrected degrees of freedom were used for the t-test.

A first testimony of supportive evidence was the significant interaction between the timing of the argument context and the strength of the association. Post-hoc analyses revealed that, prior to the presentation of the arguments, no differences were found in the credibility of critical sentences as a function of association strength. However, following exposure to the arguments, significant differences in the credibility of critical sentences were observed. Specifically, conclusions supported by arguments with strong association warrants were perceived as more credible than those supported by arguments with weak association warrants. In addition, the believability of conclusions supported by arguments with strong association warrants increased after exposure to the argument, while those supported by arguments with weak association warrants declined.

This interaction between the timing of the argumentative context and the strength of association included the evaluations of both the experimental and the control group. If in the control group there were no differences between the different levels of the two manipulated variables, then any observed differences would be due solely to the behaviour of the experimental group. As mentioned above, no differences were found in the scores of the control group. Consequently, the participation of the control group in this interaction could be attenuating the effects of the experimental group.

Thus an ANOVA was conducted for the experimental group. In this analysis, a main effect of the warrant's strength was observed, where the averages for the strong association condition were higher than those for the weak association condition. That is, in general, conclusions from arguments with warrants where the effect is strongly associated with its cause were perceived as more credible than when the warrant involved causally weakly related elements. Here too, the target interaction between the timing of the argumentative treatment and the strength of the warrant's association was observed. By excluding the control group from the analysis, the interaction showed the true magnitude of this relationship. The post-hoc analyses showed the same pattern as before but much more pronounced, and with a considerably higher effect size.

Furthermore, at this point, it is important to highlight a few things. One is that when the association in the warrant was weak, we had predicted that the credibility of the conclusions might remain unchanged or decrease in the posttest. This could depend on whether participants dismissed the premises as not credible or took them into consideration when assessing the credibility of the conclusion. In this study, participants in the experimental group were encouraged to consider the premises. This could explain the observed decrease in the credibility of the critical claims after the treatment, albeit with a medium effect size.

Another point to note is that in the control group, no significant changes were recorded between the pretest and the posttest. This group was designed to ensure that any changes observed in the posttest were attributable solely to the manipulated variables and not to external factors such as the sequence of events or the mere passage of time. Therefore, this result supports this assertion.

The three-way interaction found supports the central hypothesis of the study by indicating differences between the experimental and control groups in the pretest and posttest based on the

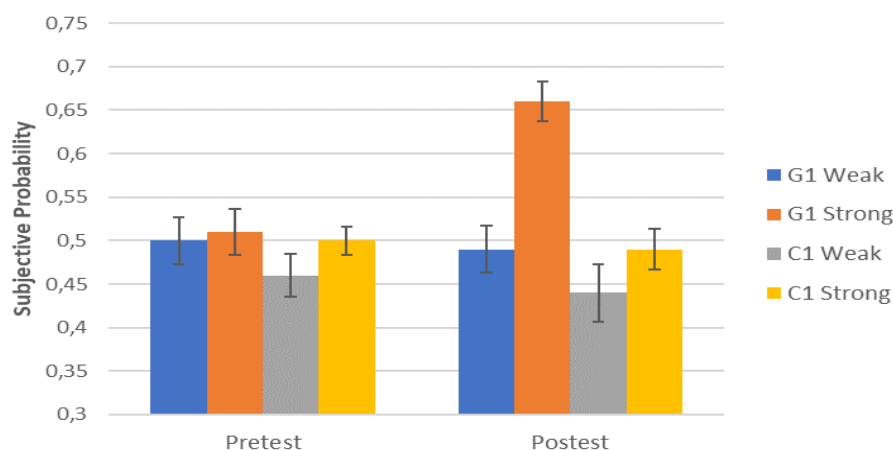
strength of association. In the pretest, the means of both groups do not differ significantly, as the experimental group has not yet seen the sentences as conclusions of arguments. However, in the posttest, the experimental group undergoes the expected changes, while the control group remains unchanged as it does not receive any argumentative treatment or warrant. The analyses show that the experimental group experiences a significant increase in the credibility of conclusions from arguments with strong association warrants, while those with weak association decrease.

6.3. Results of Group G2 on the Transmission of Credibility

The general pattern of results for G2 is shown in Figure 14 (below). As in the previous case, a 2 (Argumentative Context: Absent (pretest) vs. Present (posttest)) x 2 (Strength of Association: Weak vs. Strong) x 2 (Group: Experimental vs. Control) ANOVA was conducted.

Figure 14

Subjective Probabilities Given to the Critical Assertions by G2 and C2



Note: Mean subjective probabilities scores given by G2 as a function of argumentative treatment (prior – posterior) and the warrant association force (Strong or Weak). The error bars represent standard error.

The analyses revealed that the main effect of the argumentative context was not significant ($F(1, 41) = 2.73, p = .106$), while the main effect of the strength of association of the warrant was significant ($F(1, 41) = 11.29, p = .002, \eta_p^2 = .216$). The mean in the strong association condition ($M = .55$) was higher than in the weak association condition ($M = .47$). A main effect of the group was also found ($F(1, 41) = 9.69, p = .003, \eta_p^2 = .191$), indicating that the mean of the experimental group ($M = .54$) was higher than that of the control group ($M = .47$).

Furthermore, a significant interaction was observed between the timing of the application of the argumentative context and the group ($F(1, 41) = 6.61, p = .014, \eta_p^2 = .139$). Bonferroni-adjusted post-hoc comparisons revealed that there were no significant differences between the two groups in the pretest ($t(41) = .85, p = .397$), but there were indeed differences in the posttest ($t(41) = 3.87, p < .001; d = 1.45$), where the scores of the experimental group ($M = .58$), which had received the

argumentative context, were higher than those of the control group ($M = .46$). In addition, no significant differences were found between the pretest and the posttest in the control group ($t(17) = .80, p = .430$), but they were found in the experimental group ($t(24) = -2.84, p = .009; d = .80$), where the average increased from .50 in the absence of the argumentative context to .58 in its presence.

A significant interaction was also observed between the argumentative context and the strength of association of the warrant ($F(1, 41) = 11.03, p = .002, \eta_p^2 = .212$). In the pretest, no significant differences were found between the averages of the isolated conclusions for the strong and weak association conditions ($t(42) = -1.21, p = .232$). However, in the posttest, that is, in the presence of the argumentative context, significant differences were observed between these conditions ($t(42) = -4.76, p < .001; d = .92$). Conclusions with strong association had a higher average ($M = .59$) compared to those with weak association ($M = .47$). Additionally, no significant differences were detected between the pretest and posttest averages in the weak association condition ($t(42) = .411, p = .683$), but significant differences were observed in the strong association condition ($t(42) = -3.49, p = .001; d = .69$). In this association condition, the average obtained in the presence of the argumentative context in the posttest ($M = .59$) was higher than that obtained in its absence in the pretest ($M = .51$).

Finally, a three-way interaction was observed among the variables of argumentative context, strength of association, and group ($F(1, 41) = 10.26, p = .003, \eta_p^2 = .200$). Bonferroni-adjusted post-hoc t -tests indicated that, in the pretest phase, there were no significant differences between the averages of the experimental and control groups in relation to strength of association ($t(41) = 1.09, p = .282$ for weak association; $t(41) = .12, p = .900$ for strong association). In the posttest phase, no differences were observed between the averages of the experimental and control groups in the weak association condition ($t(41) = 1.34, p = .186$). However, in this posttest phase, a notable difference was observed between the averages of the experimental and control groups in the strong association condition ($t(41) = 5.11, p < .001; d = 1.61$). The mean of the experimental group was higher than that of the control group.

Additionally, these analyses revealed a lack of differences in the averages of the experimental group between the conditions of absence and presence of the argumentative context in the weak association condition ($t(24) = .08, p = .936$). However, significant differences were found between the conditions of presence and absence in the strong association condition ($t(24) = -5.22, p < .001; d = 1.24$). In contrast, in the control group, no differences were observed between the pretest and the posttest, either in the weak association condition ($t(17) = .68, p = .505$) or in the strong association condition ($t(17) = .58, p = .570$).

As with G1, we will conduct a 2×2 ANOVA to analyse the interaction between the variables of argumentative context and strength of association of the warrant, solely for the experimental group. This analysis showed main effects of both the argumentative context ($F(1, 24) = 8.10, p = .009, \eta_p^2 =$

.253) and the strength of association ($F(1, 24) = 13.46, p = .001, \eta_p^2 = .359$). Crucially, the expected interaction was observed ($F(1, 24) = 22.12, p < .001, \eta_p^2 = .480$). Bonferroni-adjusted post-hoc analyses indicated that there were no differences between the weak and strong association conditions in the absence of the argumentative context ($t(24) = -.38, p = .705$). However, differences between these conditions were observed in its presence ($t(24) = -7.14, p < .001; d = 1.41$), with a higher subjective probability assigned to the conclusions in the strong association condition compared to the weak association condition.

An equivalent analysis performed for the control group revealed no differences in any condition.

6.4. Discussion of the Results for Group G2

The results obtained in G2 replicate those we obtained in G1, except for the fact that in G2, the credibility of the sentences in the condition of warrants with weak association does not decrease in the presence of the argumentative context but remains unchanged.

Therefore, the results once again support the central hypothesis of the study. As expected, only the experimental group showed a significant increase in the credibility granted to the critical sentences after applying the argumentative context with strong association warrants. In the other conditions, no significant changes were observed from the pretest to the posttest, confirming the anticipated interaction.

We found a significant main effect of the strength of association of the warrant, with a higher overall mean for the strong association condition. This indicates that, in the posttest, the overall mean of subjective probability for the critical sentences was higher in the strong association condition than in the weak association condition, considering both the experimental and control groups. It is important to highlight that this effect is attributed exclusively to the increased credibility of the sentences derived from an argument with a high association warrant in the experimental group's posttest. This notably high score significantly influenced the value of the overall mean, being the only case where significant differences were observed. Therefore, the presence of this main effect is essentially due to the drag effect of the mean derived from this single case, which confirms the central hypothesis of the study.

Something similar can be said about the main effect found in relation to the group. While the control group received no argumentative treatment, the experimental group did. This latter group experienced two types of treatment: one presenting the critical sentences as conclusions of arguments with causal warrants with a weak association between cause and effect, and another where the effect was strongly associated with its cause. The presence of this main effect is primarily due to the treatment that included warrants with strong association, resulting in a higher overall mean in the experimental group compared to the control group.

An interaction was observed between the timing of the application of the argumentative context and the group, confirming the absence of differences between the groups before applying the argumentative treatment. After the application of the treatment, which, let us remember, was only applied to the experimental group, the overall mean (collapsing both conditions of the warrant's strength of association) of the experimental group was higher than that of the control group.

Furthermore, another interaction was found between the timing of the application of the argumentative context and the strength of association. This result indicated that there were no differences in the credibility of the critical sentences based on the strength of association before the application of the argumentative context (collapsing the conditions of the group variable). However, after the application of the treatment, it was observed that the average credibility granted to the sentences in the strong association condition of the warrant was greater than that granted to the sentences with weak association warrants. In other words, in the posttest, conclusions from arguments with strong association warrants were perceived as more credible than those derived from arguments with weak association warrants.

These results are complemented by observing the found three-way interaction, which indicates that the experimental group exhibited different behaviour from the control group in the posttest, and not in the pretest. In the posttest, while the control group evaluated the same sentences under the same conditions as in the pretest, the experimental group considered them as conclusions of arguments with warrants of varying strength of association. The control group did not show differences in their evaluation of the sentences, neither after compared to before, nor as a function of the strength of association condition (which, in fact, only applied to the experimental group). However, as expected, differences emerged in the experimental group, where the conclusions of arguments with more credible warrants (i.e., strong association) were evaluated as more credible than in the absence of the argumentative context, and more credible than the critical sentences seen as conclusions of arguments with less credible warrants (weak association).

Finally, the analysis conducted for the experimental group allowed for evaluating the magnitude of the interaction effects between the variables of argumentative context and strength of association, uninfluenced by the scores of the control group. We observed that, despite no significant differences based on the strength of association in the absence of the argumentative context, substantial differences occurred in its presence. Specifically, the credibility of the conclusions from arguments with strong association warrants experienced a considerable increase in the presence of the argumentative context. In contrast, the credibility of the conclusions from arguments with weak association warrants remained unchanged between the absence and presence of the argument. On the other hand, when conducting an equivalent analysis for the control group, no significant differences were found.

7. Study Conclusions

The discussion of the results from groups G1 and G2 reveals similarities, and some differences, in how participants evaluated the credibility of the conclusions of arguments based on the strength of association of the warrant. Both groups supported the central hypothesis of the study, showing a significant increase in the credibility of conclusions supported by arguments with strong association warrants after exposure to the argumentative treatment. The interaction between the timing of the argumentative context and the strength of association was consistent across both groups, suggesting a consistent response in relation to the experimental manipulation.

Independent analyses for each group confirmed the significance of the interaction between the timing of the argumentative context and the strength of association of the warrant in the perception of the credibility of the conclusions. Additionally, in both groups, a significant main effect of the strength of association of the warrant was observed, with greater credibility attributed to conclusions supported by arguments with strong association warrants compared to those with weak association. Furthermore, the interaction between the timing of the argumentative context and the group (experimental vs. control) revealed that the experimental group experienced significant changes in the credibility of the conclusions after the argumentative treatment, while the control group remained unchanged. This indicates that changes in the perception of the credibility of the conclusions were specific to the experimental treatment and were not attributable to external factors.

The presence of three-way interactions between the timing of the argumentative treatment, the strength of association of the warrant, and the group supports the idea that the interpretation of the credibility of conclusions varies with the experimental manipulation. Additionally, this interpretation significantly depends on the strength of association of the warrant. Thus, the results from both experimental groups suggest that exposure to the argumentative treatment influences the perception of credibility of the conclusions, and that this influence is mediated by the strength of association of the warrant: only when it is strong does the credibility of the conclusions increase.

However, differences emerged between the groups concerning the weak association condition of the warrant. While in group G1 a significant decrease in the credibility of the conclusions supported by arguments with weak association warrants was observed after the argumentative treatment, in group G2 the credibility of these conclusions remained unchanged. This discrepancy may indicate variability in the participants' response to the experimental manipulation between the two groups, requiring further detailed evaluation to determine its exact cause. As a possible tentative interpretation, it could have been the case that participants in group G1 relied on the premises to a greater extent when estimating the credibility of the conclusions. In contrast, those in group G2 might have based their responses more on their a priori credibility when addressing the conclusions under this condition.

8. General Discussion

The belief system model proposed in this thesis is based on argumentative paradigms. According to this model, a representation becomes believed when it achieves a sufficient degree of justification for the individual. All representations stored in the belief system are subjectively justified and can convey that justification to new representations.

In this work, it is considered that this transmission is carried out through inferential links consisting of topical relation rules, which can be of varying strength. In this way, the system preserves the credibility of the representations through inferential steps of differing intensities. The amount of justification is measured in terms of subjective probabilities, and the transmission operations respect probabilistic validity and consistency. This implies that if a representation is maximally justified and the relation rule is maximally strong, then the new representation incorporated into the system will also become maximally justified for the individual. Otherwise, a contradiction would occur, violating the principle of consistency that underpins the system. Thus, the parallel established between the elements of the belief system we postulate and the elements of the arguments we employ becomes clear. The justified representation is identified with the data premise, as it refers to real-world situations. The candidate representation corresponds to the critical claim or conclusion, while the topical relation rule acting as an inferential link corresponds to the warrant premise. When the critical claim represents content foreign to the belief system and the two premises represent internal elements of the system, the conditions set by the standard of argumentation are met. Arguments that meet this standard are considered enthymemes. If an argument is not an enthymeme, it will not be effective in producing an increase in the credibility granted to the critical claim or conclusion. In these cases, the credibility of the conclusion may remain unchanged if the individual ignores the offered premises or decrease if they consider them and they result in a credibility lower than the initial one.

8.1. On the Argumentative Nature of the Deductive Aspect of the Belief System

The main goal of this experiment was to gather evidence about the argumentative nature of the deductive dynamics of beliefs assumed by the theory. This was achieved in two ways: by presenting the conclusion to participants twice, first in isolation (in the pretest) and then as the actual conclusion of an argument (in the posttest); and by manipulating the credibility of the warrants through the strength of association between the antecedent and the consequent, while the credibility of the data premise was kept constant at high levels. To our knowledge, this is the first study on belief change that employs a pretest-posttest structure. It was hypothesised that, through these manipulations, the transmission of justification should vary, occurring only after the argumentative treatment when the warrant was highly believed. Therefore, the subjective probability of the conclusions should be significantly higher when presented as conclusions of enthymemes than when presented in isolation or as conclusions of non-enthymematic arguments.

The experimental results indicated that the transmission of credibility occurred as predicted. Generally, before the argumentative treatment, the critical sentences were believed at similar levels by the participants, except for the sentences associated with the weak association level in G1, which were believed to a greater extent⁵⁵. However, significant differences were observed after the argumentative treatment. Specifically, the conclusions linked to the data premise by a strong association warrant achieved greater credibility compared to others. On the other hand, in G1, the conclusions preceded by an argument with a weak association warrant were less believed than others, while their credibility remained unchanged in G2. Finally, no differences were found in the credibility assigned to the conclusions by the control group before and after the argumentative treatment.

The results suggest that the credibility of the warrant, which functions as a connector, plays a fundamental role in the transmission of justification that takes place in the belief system.

8.2. Comments on Specific Results: The Case of G2 in the Weak Association Condition

An increase in the credibility of the conclusions in the weak association condition was not expected, and indeed, it was not observed in either group. However, observing the subjective probabilities assigned to the premises in group G2 (see Table 6), one might expect an increase in the probabilities assigned to the conclusions, given that this condition meets the standard of argumentation⁵⁶. But there was none. This may be due to the fact that the a priori credibility of the conclusions in isolation was very close to the credibility of the premises. Perhaps, to increase credibility, there must be a significantly greater difference between the credibility of the premises and that of the conclusions in isolation, with the premises being notably higher. But, if this is the case, then it raises the question of how much more credible the premises must be than the a priori credibility of the conclusions in isolation to see an increase.

8.3. Advantages and Disadvantages of a Pretest-Posttest Study with Two Identical Experiments

Manipulating the plausibility of the warrants by varying the strength of association of the consequents for the same antecedent led to a study composed of two different experiments with the same design and procedure. This presents both negative and positive aspects. Among the negatives are the complexity of the design and the reduction in statistical power. However, the effect sizes reported have been consistently high. Moreover, the analyses reveal that the observed power remained at high levels, ranging from .999 to .997 for all effects in G1, and from .995 to .879 for the effects in G2.

⁵⁵ This difference is likely due to the content of those sentences, which must have seemed more credible, but only to participants in the experimental group and not to those in the control group. However, this difference in the credibility of the sentences belonging to the weak association condition is irrelevant since before the treatment they were believed more than the rest, not less. In contrast, after the treatment, they were believed less than the rest.

⁵⁶ In other words, although it was not desired, a minimal increase in this condition was expected, since the arguments met, albeit minimally, the standard of argumentation.

On the positive side, the main advantage lies in the observation of the replication of results, which strengthens the conclusions. Additionally, it has allowed for the observation of diversity in the behaviour of the subjective probability of the conclusions derived from arguments with weak association warrants. The main conclusion is that, in these cases, the subjective probability granted to the conclusion never increases or surpasses the credibility of any of the premises. The results in both groups indicate that the premises have been considered, as requested in the experiment instructions.

On the other hand, the experimental test-posttest structure also has negative and positive aspects. A significant negative aspect is experimental attrition, which poses a risk when there are limitations in the sample size. In this case, 13 participants were lost from the pretest to the posttest. However, the fact that it was possible to directly measure the change in credibility produced by the argumentative treatment for the same sentences with the same participants for the first time is a significant positive aspect.

8.4. Manipulation of Credibility in Warrants and Its Implications

In this study, we manipulated the credibility of warrants in conditional sentences through a task in which participants, different from those in the main experiment, generated effects for specific causal antecedents. The effects mentioned more frequently were considered strongly associated with their antecedent, while the less frequent ones were considered weakly associated. This strategy aimed to create highly credible warrants. The method was based on rhetorical plausibility, which seeks to start from premises considered probable by all or most members of an audience. In our case, that audience was the psychology student community at the University of Tarragona. If the majority mention an effect for a cause, then they believe that effect follows from the cause. The prevailing belief among the public is that the occurrence of the effect is likely if the cause is present. This is what was reflected in the averages obtained for the credibility of the strong association warrants.

In their study on the impact of the strength of association between cause and effect on the acceptance of conclusions, Quinn & Markovits (1998) discovered that strongly associated causes could act as refuters⁵⁷ for conditionals that included weakly associated causes to the same effect. Participants tended to reject to a greater extent the conclusions derived from the invalid AC and DA arguments when the conditional presented a weakly associated cause in its antecedent. The ease of retrieving from LTM alternative causes with a strong association showed them that the initial cause was not necessary, leading them to dismiss the conclusions of such arguments. However, in cases of strong associations, it was more difficult to identify alternative causes with even stronger associations, which increased the acceptance of such conclusions.

In our study, the strength of association between effects and causes could influence the credibility of conditionals in a similar way, affecting the acceptance of conclusions in valid

⁵⁷ They acted as the *undercutting defeaters* described by Pollock (1987); see sec. 13.2., chap 4, pt. I.

arguments. It is likely that strongly associated effects serve as refuters in conditionals with weakly associated effects to the same cause, especially if they are incompatible. Following the logic of Quinn & Markovits, participants would more readily find strongly associated and incompatible effects when the association is weak, indicating that the cause is not sufficient for the effect, which would reduce the acceptance of conclusions based on that premise. Conversely, with strong associations, it would be difficult to find an even more associated and incompatible effect, thus increasing the acceptance of the conclusions.

However, not all retrieved effects will be incompatible. For example, consider that 'a woman is a first-time mother at forty' (cause); that 'the child has deficiencies' (strong effect) does not refute that 'the child is an only child' (weak effect). The latter, weak effect, does not automatically lead to refuters or decreased credibility. Apart from these cases, it is also possible that some believe that a general event C usually does not produce effects, but if any effect occurs, it would be E. This scenario could result in many participants mentioning E in relation to the potential cause C, even though they consider the credibility of the conditional 'if C, then E' to be low. In this sense, factors such as the *perceived* covariation between cause and effect, the presence of refuters, and their relative frequency significantly influence.

Perceived covariation—the frequency with which an effect is observed given a cause, or the frequency of exceptions to this rule—directly impacts the credibility of conditionals (Geiger & Oberauer, 2007). Another important factor is the number and strength of association of refuters (de Neys et al., 2003ab), among which '*disablers*' or '*rebutting defeaters*' stand out (for example, Cummins, 1995; Pollock, 1987), as described in sec. 14, chap. 4, pt. I. These elements, which can prevent the manifestation of an effect despite the presence of its cause, do not necessarily consist of strongly associated and incompatible effects. Rather, they are often external events related to the causal situation described in the conditional.

Thus, the relationship between the strength of association of causes and effects, the credibility of conditionals, and their influence on the acceptance of conclusions is certainly complex. Causal conditionals that align with common opinion (*endoxa*) tend to show strong association; that is, when a cause evokes a specific effect for the majority, the resulting conditional is considered doxastic. Moreover, the direct or vicarious perception of the frequency of occurrence of an effect given a cause significantly contributes to this association in the subject's mind. However, as explained, a strong association does not automatically imply high credibility, nor does a weak association necessarily entail low credibility. These variables, although closely linked, depend on the interaction with other factors to influence the perception of credibility.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

EXPERIMENT 2

Theoretical Foundations

In the theory proposed in this thesis, the belief system is conceptualized as a cognitive tool for understanding the world, enabling the prediction of incoming events to achieve efficient environmental adaptation. Within this system, concepts, events, and situations are organized, classified by domains, and linked under different relational criteria.

Belief systems are formed through experience by the interaction of individuals' cognitive systems with their surrounding world. The basic cognitive mechanisms in belief system construction described in the thesis are threefold: the intuitive-inferential, the perceptual, and the socio-communicative.

Incoming information that ends up as a belief can become either a foundational belief or a derived belief. They differ, among other things, in their source of justification. The justification for foundational beliefs does not come from inference about other beliefs, but from implicit confidence in how the cognitive system operates. These mechanisms are considered highly effective and reliable for obtaining accurate information. Thus, foundational beliefs can combine to form new beliefs called derived beliefs. Therefore, the justification for derived beliefs stems from inference about other previously justified beliefs. Whenever formed by inference, these derived beliefs will have an explanation, both for their content and, above all, for their justification. Something becomes believable whenever it can be explained. Three general explanation procedures have been proposed: abduction, induction, and deduction. The main focus of the present experiment is on the deductive aspect of belief formation and dynamics. Before delving into the deductive aspect, it is essential to understand the roles of induction and deduction within the belief system.

1. Induction and Deduction

Over time, individuals accumulate experiences involving a myriad of objects, events, situations, and relationships; some of such experiences are stored in episodic memory. Many of these experiences can be grouped by similarity into different classes. People perceive that most times, A goes with B, although there are exceptions. Based on the frequencies of these binary relationships and their exceptions, processes of association and induction naturally unfold. Through these cognitive processes, individuals gradually form general concepts and relationships that describe how things typically work in the world. This accumulated knowledge is stored in semantic memory. These derived beliefs can take the form of prototypical and schematic mental representations, encompassing objects, events, situations, as well as general rules of relationships, all with varying degrees of abstraction. For example, individuals may have a general rule like 'if the sky is dark, grey and cloudy, it will rain', or at a higher level of abstraction, 'if a cause has occurred, then its effect will follow'. It is important to note that these representations and rules take exceptions into account and remain susceptible to updates based on confirmatory or refutative evidence from subsequent experiences. Therefore, although they have reached high levels of stability, they are fallible (i.e., defeasible). Essentially, this process aligns with Bayesian belief updating.

These schematic general rules establish coherence relations among the elements composing complex mental representations and among different representations. A subset of these coherence relations is called topical relations (detailed in App. C). This type of relations allows for the transmission of justification between representations. Ultimately, our belief system consists of networks with flexible representations interconnected by adaptable relational rules, forming domains of knowledge governed by implicit theoretical principles. These domains enable us to derive predictions. When incoming content activates an implicit theory⁵⁸ and aligns consistently and coherently with it, it becomes predictable, explainable, and therefore easily accepted and believed. The system seeks to verify the consistency and coherence of each piece of incoming information with previously established beliefs. If the new information can be coherently connected with an existing belief through a topical relation, part of the existing justification is transmitted to the new information. However, these processes are no longer inductive but rather deductive in nature.

Deduction utilizes the final outcomes of induction; inductive processes result in justified content 'from the world to the mind', and deduction transmits that pre-existing justification to the incoming candidate content. If the amount of transferred justification is *high enough*, the new content is integrated into the belief network in long-term memory. Therefore, the justification of incoming content depends on the set of previously activated beliefs in working memory, termed the "local belief system" (LBS). This justification process involves a subjective estimation of the

⁵⁸ I use the term 'Implicit Theory' in a sense similar to that developed by Tenenbaum, Griffiths, & Kemp (2006).

probability of the incoming content given the LBS. Justifying derived beliefs through deduction operates conceptually similar to conditional probability.

2. Degrees of Belief

Many authors contend that foundational beliefs are *prima facie* subjectively justified, leading to maximum credibility. However, over time, individuals may encounter information contradicting their foundational beliefs (i.e., defeaters). This contradicting new information may also constitute a foundational belief. Depending on the frequency of occurrence of these defeaters, the credibility of the initial belief may decrease or even vanish. Accordingly, a decrease in the credibility of a foundational belief caused by a defeater arises from processes of inductive inference. Consequently, the initial belief ceases to be foundational and becomes derived. Conversely, there may be derived beliefs with maximal credibility (typically, though not exclusively, many formal beliefs). However, all representations with credibilities below the maximum are derived beliefs. Thus, the system is structured with foundational and derived beliefs interconnected by relationships of justification and refutation, which determine the credibilities of their associated representations. When incoming information activates the LBS, it triggers its representations with different levels of justification, along with these two types of relationships. The degree of justification it receives from the LBS depends on the elements it connects with. If it activates more refutative than justificative elements, it will be rejected. The elements of the belief system vary in their level of justification, needing to exceed a minimum threshold to be included. Maximum justification reflects total subjective and positive certainty about the necessary occurrence of an event, while lower degrees indicate lesser certainty. A certain degree of positive certainty is sufficient for the mental model of the situation to become part of the system. Negative certainty implies the belief in non-occurrence, reaching its maximum in the consideration of impossibility. Total uncertainty indicates the absence of certainty, where occurrence is considered necessarily as impossible, manifesting in doubts. And what is doubted does not form part of the belief system.

As developed in previous chapters (3, 4 of Part I and 6 of Part II), degrees of justification, and consequently, degrees of belief, are typically measured in subjective probabilities. A probability of 0 corresponds to total negative certainty (i.e., impossibility). A probability of 1 corresponds to total positive certainty (i.e., necessity). Finally, a probability of 0.5 corresponds to total uncertainty, where the principle of indifference applies (introduced in sec. 8, chap. 2, pt. I). Therefore, all representations within the belief system are assigned a subjective probability greater than 0.5, with a maximum of 1 (i.e., $0.5 < P(\text{beliefs}) \leq 1$). Consequently, any content evaluated with a probability equal to or less than 0.5 will not be part of the belief system. If I attribute a probability of 0.6 to incoming content as a result of the activated LBS, I would assign a probability of 0.4 to its negation or contradiction (i.e., its complementary probability). The probability scale allows for the expression

of different states of belief (when greater than 0.5), uncertainty (when equal to 0.5), and incredulity (when less than 0.5) regarding any given content.

Numerical quantification is useful in analysing credibility, as it allows for comparisons and mathematical operations such as addition, averaging, and multiplication. This enables operations with the qualitative properties of objects, properties that are present to varying degrees. In this context, this property corresponds to the credibility of representations, which, given in degrees, can be quantified. It is important to note that this quantification does not imply that people assign numerical values to their thoughts in daily life, nor do they consciously perform mathematical calculations when reflecting on or predicting events. Alternative non-numeric measurement methods have been proposed, such as iconic representations along a continuum or a scale with 8 modal certainty qualifiers (for example, Khemlani et al., 2015). However, these proposals do not differ substantially from using numerical values. Previous research on everyday reasoning has required participants to provide direct probability estimates for potential scenarios (for example, Evans et al., 2015; Over et al., 2007). Participants have not reported problems with this approach, even though they normally do not think in numerical terms. Therefore, in the present study, participants will be asked to provide a direct numerical rating for the probability of occurrence of various potential scenarios.

3. Deduction in Belief Formation and Dynamics

As previously mentioned, the credibility of a representation arises when an individual attributes it a sufficient degree of justification, expressed in subjective probabilities. Building upon this concept, justification is transmitted to incoming information through coherent connections within the LBS. The subjective probability assigned to the new representation is determined by its relationship with the epistemic framework of LBS, thus conditioning its subjective probability to the prior beliefs activated within LBS. Essentially, new information inherits the probabilities associated with relevant existing beliefs. The aim of this experiment is to investigate this process of probability inheritance in new information with respect to relevant existing beliefs.

Real-world experiences are stored in episodic memory and then subjected to inductive processes of decontextualization, abstraction, and generalization. These processes link various instances of experiences to construct general and schematic representational and relational rules, also considering their exceptions. Consequently, representations and rules have varying degrees of credibility. Subsequently, deductive processes can be employed on the general representations and rules to justify new information, thus generating derived beliefs. LBS representations transmit their justification to incoming information through relational rules. Our main hypothesis is that, in this transmission, the magnitudes of justification of the representation and the rule are mechanically combined according to a systematic pattern. We consider LBS elements to act as premises, and incoming information as the conclusion of an argument. Drawing inspiration from Toulmin's

argumentation model (1958), we refer to justified LBS representations as 'data premises' and justified general relational rules as 'warrant premises'. Incoming information is termed a 'claim' or 'critical assertion'. Given its deductive nature, the inferential transmission process is expected to meet three essential requirements:

- a) Analyticity: There should be nothing in the conclusion that is not already present in the premises, both in terms of content and degree of justification.
- b) Validity: It ensures the preservation of justification (i.e., credibility) from premises to conclusion, which, within the framework of probability logic, is interpreted as the preservation of probability.
- c) Consistency: Inferences must adhere to the principle of non-contradiction, implying that premises must be consistent with each other, and the conclusion must be consistent with them.

In the following sections, I will provide a brief explanation of the concept of probabilistic validity and delve into the notion of argumentative consistency. The proposed operation for transmitting justification in this thesis is primarily based on consistency, along with the discursive concept of coherence.

4. Probabilistic Validity

Deductive logical validity involves preserving truth from the premises to the conclusion. In a valid argument, if the premises are true, then the conclusion is necessarily true. The truth of the conclusion logically follows from the truth of the premises. However, the reverse does not hold; the falsity of the premises does not necessarily make the conclusion false. It is possible to have a valid argument where the premises are false, but the conclusion is true.

Some authors, such as Adams (1998) and Evans and Over (2013), argue that probabilistic deductive validity entails *preserving subjective probability from premises to conclusion*. According to this perspective, if premises are assigned a maximum degree of probability, then the conclusion should be deemed totally probable. Under this notion, *a deductively probabilistically valid (p-valid) conclusion cannot coherently be more uncertain than the total uncertainty of the premises*. In this context, 'uncertainty' is construed as 'disbelief' or 'incredibility'. Thus, it is posited that *the level of disbelief in the conclusion cannot exceed the total degree of disbelief in the premises*, that is, the sum of disbelief in the premises. In the context of quantifying believability in terms of probability, this principle can be succinctly expressed as follows: (where $\bar{P}(x)$ is the complementary probability of x , i.e. $1 - P(x)$):

$$p\text{-validity: } \bar{P}(\text{conclusion}) \leq \sum \bar{P}(\text{premises})$$

In its direct counterpart, p -validity sets a minimum threshold for the credibility of the conclusion, expressed as $\sum P(\text{premises}) - 1$. For instance, let's consider a scenario where I believe in the content of a 'data-premise' to a degree of 0.7. Additionally, I hold a belief in a rule stated in the 'warrant premise' also at a level of 0.7. Following the deductive process of argumentative justification, my degree of belief in the incoming information (i.e., the 'claim' or 'critical assertion') cannot fall below 0.4 (i.e., $(0.7 + 0.7) - 1 = 0.4$). These are the guidelines provided by p -validity on how to combine the credibilities of the premises.

4.1. Probabilistic Validity in Argumentation

The principle of analyticity states that disbelief in the conclusion cannot exceed the combined disbelief in the premises, as the conclusion cannot contain more than what is already present in the premises. Thus, if there is no disbelief in the premises, there will be none in the conclusion. Additional disbelief cannot be introduced from the premises to the conclusion during the inferential process (Evans & Over, 2013). However, it's important to note that complete disbelief in both premises does not necessarily determine the credibility of the conclusion. While logical validity allows for arguments with false premises and a true conclusion, probabilistic validity permits arguments with premises that are not believed and a conclusion that is believed (according to the p -validity formula). What is asserted in the conclusion could be believed *a priori* for reasons independent of the premises. Nevertheless, this diverges from typical argumentative practice, where arguers seek to persuade their audience of a claim's truth.

If the total disbelief in the premises exceeds 1, the argument will always be p -valid, as the resulting incredibility in the conclusion cannot surpass this sum. Similarly, if the prior believability in the conclusion exceeds the minimum threshold of believability derived from the premises (i.e., $\sum P(\text{premises}) - 1$), the argument will also be p -valid. Therefore, for p -validity to be a significant criterion in argumentation and belief modification, the total sum of incredibility of the premises should be less than 1 and greater than the prior belief in the conclusion. Ideally, if the prior belief in the conclusion were 0, there would be no alternative reason within LBS that would allow for the possibility of the truth in the assertion apart from the stated premises.

4.2. Boundaries and Scenarios in p -Validity Applied to Argumentation

In realistic argumentative scenarios, the prior credibility of the critical claim tends to be greater than 0 and, when deemed not credible, it is typically equal to or less than 0.5. The absence of a content within the belief system does not necessarily imply that it is considered impossible. In such contexts, the lower bound of p -validity becomes relevant if the total sum of incredibilities in the premises is less than the prior incredibility in the critical claim. Given that the prior credibility of a critical claim can assume any value on the subjective probability scale, let us examine some extreme cases:

1. The prior believability of the critical claim may be 0. In this idealized argumentative scenario, any level of believability in the premises exceeding the minimum threshold of 0 should influence the subsequent believability of the critical claim. In such cases, the

believability of the conclusion should be explained and predicted by the believability of the premises.

2. The prior believability of the critical claim may be 1. In this extreme 'non-argumentative' situation, disbelief in the premises may not necessarily impact subsequent belief in the conclusion. This will depend on whether the listeners rely more on their prior believability or on the believability of the premises presented when evaluating the conclusion.
3. The prior believability of the claim could be greater than 0 and less than 1. The influence of the believability of the premises on the posterior believability of the conclusion could depend on the combined incredibility of the premises relative to the prior incredibility of the conclusion. If the sum of the incredibility of the premises is less than the prior incredibility of the critical claim, this will affect the posterior believability of the conclusion, but not necessarily in the opposite direction.

All these considerations are implicit in the p -validity formula.

4.3. Implications and Variations of p -Validity

The total incredibility of the premises can range from 0 to 2. When it exceeds 1, any outcome in the conclusion believability will be valid, as it can vary between 0 and 1. Furthermore, if the believability of each premise surpasses the prior believability of the critical claim, the believability of the conclusion should increase according to the argumentative standard. One crucial aspect of the lower limit for the credibility of the conclusion, established by p -validity, is its dependence on the number of premises (Demey, Kooi & Sack, 2019). According to the p -validity formula, as the number of premises with a credibility below 1 increase, the probability of the conclusion decreases. This dynamic reflects how incredibility among the premises accumulates in the conclusion.

5. Consistency

In classical logic, a set of propositions is considered consistent if they can all be simultaneously true, and thus, no inherent contradiction within the set can be proved. Similarly, a set of beliefs is regarded as consistent if an individual can assign a high subjective probability to all of them concurrently, meaning these beliefs do not conflict. Contradictions typically arise when ideas are related in some manner. Therefore, ideas that are subjectively or objectively independent of each other tend to remain consistent. For instance, believing 'tomorrow will be a sunny day' and knowing 'wolves are mammals' are perfectly consistent in context.

Potential contradictions arise when ideas are interrelated. If two ideas are contradictory, at least one of them must be incorrect, making it untenable to hold them simultaneously and in the same interpretation. This aligns with the principle of non-contradiction in its doxastic form (Gottlieb, 2008), presented here as fundamental to the belief system. Consistency is often considered the hallmark of rationality by many authors (e.g., Kruglanski & Orehek, 2009; Nitsch & Kalenscher, 2021;

Legrenzi, Girotto & Johnson-Laird, 2003; Ragni, Khemlani & Johnson-Laird, 2014). A contradictory system reduces to absurdity, blurring the line between truth and falsehood and hindering environmental comprehension.

Accordingly, Chapter 1 of Part II discusses how the belief system aims to primarily eliminate contradictions within the LBS. In contrast, within a consistent system, certain formulas logically derive while others do not. Consequently, specific premises allow deducing or predicting outcomes, establishing criteria for distinguishing what is potentially true from what is not, enabling learning, prediction, and adaptation to the environment.

6. Contradiction with Degrees of Belief

As previously mentioned, the principle of non-contradiction prohibits simultaneous belief in something and its negation. Belief in a representation implies firm acceptance within the system, typically with a probability exceeding 0.5. For example, a belief with a probability of 0.6 inherently rejects its contradictory counterpart, which falls below the 0.5 threshold and is thus absent from the belief system. Subjective probabilities adhere to probability axioms, including finite addition and complementarity. Consequently, the negation's probability is the complementary value of 0.4, representing rejected ideas *based on existing beliefs*.

Content enters the belief system by resisting rebuttals or defeaters, yet this resistance doesn't ensure absolute credibility. Assigning a probability of 0.6 to a content acknowledges the presence of rebuttals or defeaters, but it remains a belief. Consequently, a probability of 0.4 corresponds to its contradiction. This consistent distribution adheres to axioms, ensuring probabilities sum to 1. Failure to adhere to this sum constitutes a contradiction and inconsistency. A subjective probability of 0.5 is assigned when an event is subjectively equally possible and impossible, reflecting pure uncertainty.

7. Belief Consistency in Argument Components and Equitable Believability Transfer

The argument, comprising the data and the warrant, must exhibit internal consistency: Consistency between the data and the warrant is ensured by their independence as events: the probability assigned to one does not rely on the verification of the other. Therefore, complete belief in both simultaneously is possible without contradiction. For instance, one can believe that 'means of transport emit greenhouse gases' independently of whether one believes that 'greenhouse gases contribute to the extinction of polar bears.'

Data-premises and warrants in enthymemes must be consistent with the LBS: Believability is rooted in pre-existing beliefs, guided by relational rules already integrated into the belief system.

These elements harmonize with other components within the local belief system, facilitating the transmission of justification from previously justified representations to incoming ones. Enthymeme premises exhibit probabilistic consistency and coherence with other elements in the active LBS. Consistency ensures the absence of contradiction, while coherence implies semantic relevance. This dual characteristic facilitates believability transmission within the system, impacting the subjective probability assigned to new representations. An argument is not an enthymeme if at least one of its premises is incoherent (i.e., irrelevant) or inconsistent (i.e., contradictory) with some element activated in the LBS.

The claim must be consistent with both the data and the warrant: The claim lacks initial justification to enter the belief system if it is inconsistent or irrelevant to it. In an idealized argumentative scenario, when a claim is entirely inconsistent with the LBS, processes to resolve the inconsistency are initiated when presented as a deductive conclusion derived from believed premises at a later moment. In this case, the credibility of the claim is systematically derived from the believability of the already accepted premises, contingent upon that of the LBS. The degrees of believability associated with the premises determine that of the conclusion, with *both premises equally contributing to the transmission of believability*. Without a justified data representation, there would be no source of credibility from which to transfer justification to the incoming representation. Similarly, without a topical relational rule, the justified data representation would not connect with the incoming representation, hindering the transmission of credibility, as the incoming information would not be linked to the LBS. This holds true regardless of the arbitrary believability of the other premise. This is referred to as the '*Equitable Transfer Hypothesis*'.

8. Believability Combination in Deductive Reasoning: The Role of Multiplication

Within the belief system, mental representations are interconnected with varying degrees of intensity, corresponding to their believability. During believability transmission processes, these varying degrees of belief in mental representations are combined, resulting in a new representation with its own degree of believability. The question arises: how can we best describe the operation governing this combination of beliefs in the deductive context?

The resulting degree of believability in the new representation can take different forms. It may surpass the believability of any individual element being combined, then having an *increasing magnitude*. Alternatively, the combined result could be higher than one element but lower than the other, representing an *internal magnitude*. Lastly, the outcome could be less than the believability associated with any of the elements being combined, referred to as *decreasing magnitude* (Díez & Moulines, 1999).

Arguments may consist of multiple premises, with believability having an upper limit of 1. If it were of the increasing type, the upper limit could easily be exceeded, leading to inconsistency. Hence, this option is ruled out. If it were of the internal or decreasing type, it would never surpass

the upper limit. However, the model assumes that the absence of either the warrant or the data completely prevents the transmission of believability. Let's consider a scenario where one premise has a believability of 1 and the other of 0. If an internal magnitude were considered, the credibility of the conclusion could vary between 0 and 1. However, this contradicts the presented model. According to it, both premises are equally essential for credibility transmission, so the absence of either of them prevents such transmission. Therefore, this scenario would not be applicable.

As previously mentioned, probabilistic validity implies that the conclusion accumulates a proportion of the disbelief present in the premises. Consequently, as the number of premises with varying degrees of incredibility increases, the believability of the conclusion decreases. When combined with the imperative not to exceed the upper limit, the decreasing magnitude is the most suitable fit for the model. Given that believability varies between 0 and 1, decreasing as the number of premises with some degree of incredibility increases, and considering that both types of premises are equally crucial in credibility transmission, arithmetic multiplication of premise probabilities emerges as the most suitable operation for this scenario. Through this operation, the believability of the conclusion becomes entirely dependent on and determined by the believability of the premises. This can be expressed as follows:

$$P(\text{claim}) = P(\text{Data}) \cdot P(\text{Warrant})$$

We refer to this theoretical property as the '*Hypothesis of Multiplicative Combination*'. Unlike Bayes, a priori believability of the conclusion is not incorporated into the formula. If the standard of argumentation is met, multiplication should suffice in describing believability transmission. However, a priori beliefs may impact the credibility of the claim when the argumentative standard is not met or when the result of multiplication is lower than this prior believability.

9. Hypothetical Consequences of the Multiplication Operation

9.1. The Weakest Part Epistemic Rule

In this model, we assume that justification is transferred from a previously justified representation (data) to the candidate representation (critical assertion) through a similarly justified relational rule (warrant). Both the data and the warrant can be either fully or partially justified. When both are fully justified (i.e., with a probability of 1), the assertion will also be fully justified (i.e., with a probability of 1). When the warrant is fully justified but the data is not (for example, with a probability of 0.4), all the justification from the data is transmitted unrestrictedly to the critical assertion (reaching, at least, a probability of 0.4). However, when the data is fully justified but the warrant is not (for example, with a justification of 0.2), the warrant imposes strong restrictions on the amount of justification that can be transferred through it, limited by its own justification (i.e., 0.2). Finally, when neither premise is fully justified, both impose limitations on the justification that the critical assertion can receive. The conclusion accumulates the disbelief of the premises, with the

less justified premise having a more decisive impact on the transferred credibility. If either premise is considered completely improbable (probability of 0), no transmission occurs. This consequence of the multiplicative combination is referred to as the '*Hypothesis of the Weakest Part Epistemic Rule*'.

It's important to note that we are discussing the transfer of credibility from the premises to the conclusion, not the final credibility that the conclusion may have. The Weakest Part Rule can have various consequences on the credibility of the conclusion, depending on the relative magnitude of the weakest premise compared to the a priori credibility of the critical claim. When this credibility exceeds 0, the weakest premise can be greater than, equal to, or less than the initial credibility. We will explore the different ways in which this magnitude relationship could influence the subsequent credibility of the critical claim.

9.1.1. Idealized Argument Conditions: Prior Believability of the Claim at 0

One consequence of multiplication is that if the a priori credibility of the conclusion is 0, its posterior credibility will either match or be lower than the credibility of the least credible premise. Specifically, if one premise has a credibility of 1, the posterior credibility will be equal to that of the other premise. Conversely, when both premises have a credibility less than 1, the posterior credibility of the conclusion will be lower than that of the least credible premise (e.g., $0.9 \times 0.8 = 0.72$). Consequently, while validity establishes a lower limit for the believability of the conclusion post-combination, the principle of consistency via multiplication sets an upper limit. This should be true under this ideal scenario where the prior believability of the conclusion is 0.

In these circumstances, the Weakest Part Rule stipulates that the credibility of the conclusion cannot surpass that of the least credible premise. A debate exists between proponents of probabilistic reasoning, who endorse this rule (e.g., Hahn & Oaksford, 2006b; Hahn et al., 2013), and proponents of plausible reasoning (e.g., Pollock, 2001; Rescher, 1976; Walton, 2004, 2008). The latter contend that the credibility of the conclusion cannot fall below that of the least credible premise, proposing that credibility functions not as a diminishing magnitude but as an intrinsic one. They have referred to this rule as the Weakest Link Rule, the Theophrastus Rule, or the MAXMIN Rule. However, none of these positions has undergone experimental scrutiny. To our knowledge, the only studies that have addressed this topic are those by George (1997)⁵⁹ and by Hahn et al. (2013), whose findings largely supported the Weakest Part Rule described here.

9.1.2. Realistic Argument Conditions: Non-Extreme Prior Believability of the Claim

In realistic scenarios, very few things are considered entirely impossible or necessary. When assessing the believability of a proposition, individuals draw upon both supporting and refuting information from their local belief system. As a result, individuals assign a degree of believability to the proposition, ranging between 0 and 1. When an individual does not believe the critical assertion

⁵⁹ Strictly speaking, George (1997) did not directly address this issue. However, the results obtained are directly relevant to the topic.

(i.e., the claim), they may assign it a subjective probability between 0 and 0.5. There are two possibilities.

1. In *enthymematic arguments*, premises typically receive subjective probabilities between 0.5 and 1. We posit that if the result of the multiplication equals or exceeds the initial credibility of the critical assertion, it approximates the posterior credibility. This prediction, termed '*Multiplication Exceeding Initial Credibility*', will be evaluated in the current experiment. Therefore, the observed result aligns with expectations under idealized conditions, as adherence to the '*Multiplication Exceeding*' prediction also implies adherence to the Weakest Part rule. The essential condition for rule fulfilment is that the derived believability from the argumentative treatment (i.e., the multiplication result) must exceed the prior believability of the claim.
2. In non-enthymematic arguments, at least one premise will typically be not believed, having a probability between 0 and 0.5. Consequently, it is possible for the prior believability of the critical assertion to exceed the multiplication result and even that of the least believed premise. In such cases, the posterior believability may remain unchanged or decrease, depending on whether the individual disregards or takes into account the offered premises. Previous studies (e.g., Cummins, 1995; George, 1995; 1997) suggest that individuals tend to dismiss what appears incredible to them. In either case, the posterior believability should range between the multiplication result as its lower limit and the prior believability as its upper limit. This prediction, termed '*Multiplication Falling Below Initial Credibility*', will be tested in this experiment. This can be expressed as follows (where Π stands for multiplication):

if ($\Pi P(\text{premises}) \leq P(\text{Claim Prior})$) *then* ($\Pi P(\text{premises}) \leq P(\text{Claim Post}) \leq P(\text{Claim Prior})$)

9.2. Revisiting the Argumentation Standard: Required Premise Probability

The first experiment introduced the concept of argumentation standard and its significance in bolstering the credibility of a claim. According to this standard, premises should possess a higher credibility than the prior believability of the claim. However, in some instances, this did not result in a significant increase in the posterior credibility. The credibility of the premises was only slightly higher than the priori credibility of the claim, which may explain the lack of effectiveness. This aligns with our proposal that believability is a decreasing magnitude but raises the question of how much more credible the premises should be.

To address this question, we need to establish a systematic relationship between prior believability and the product of the premises. If a claim has an a priori subjective probability of 0.4, and the premises have probabilities of 0.9 and 0.8, the resulting probability increases to 0.72. The argumentation standard successfully leads to increased credibility. However, if both premises have probabilities of 0.6, the probability for the claim only reaches 0.36. Despite meeting the standard, credibility does not increase. This is because the believability of the premises remains close to the prior believability of the claim, illustrating the decreasing nature of credibility. We will seek to

establish a general criterion to ensure a minimal increase in the credibility of the assertion compared to its prior believability. To achieve this, it will be necessary to identify premises whose combined incredibility does not exceed the prior incredibility of the assertion. This prediction, termed the '*Enhanced Argumentation Standard*,' will be assessed in the current experiment. Formally, it is expressed as follows:

$$\sum \bar{P}(\text{premises}) \leq \bar{P}(\text{Claim Prior})$$

In the current example, the claim's prior incredibility is 0.6. Therefore, the combined incredibility of the premises must not exceed 0.6. For example, if each premise has an incredibility of 0.3 (i.e., $0.3 + 0.3 = 0.6$), their believability would be 0.7 each. Consequently, $0.7 \times 0.7 = 0.49$, which exceeds 0.4. This principle applies to scenarios like 0.8 and 0.6, or 0.9 and 0.5, and so on. Even better results are obtained when the combined incredibility of the premises is gradually lower than the prior incredibility of the claim. For example, in a scenario where the premises have believabilities of 0.7 and 0.8 (resulting in a joint incredibility of 0.5), the posterior believability of the conclusion is 0.56. When this principle is not upheld, the argumentation standard fails to enhance the initial credibility. For instance, if premises have probabilities of 0.6 and 0.5, their combined incredibility of 0.9 surpasses the initial incredibility of 0.6. In such instances, it is unlikely for the standard to produce an increase in the initial credibility, even if technically met (since 0.6 and 0.5 exceed 0.4). The resulting probability of their multiplication would be 0.3, lower than the original value of 0.4. While not all violations of the revised standard lead to this outcome, it is prevalent in most cases. Box 1 below relates the concept of the weakest part rule to the revisited standard of argumentation.

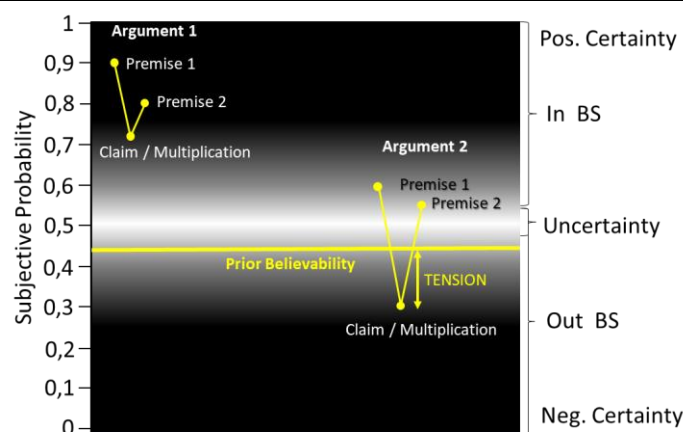
9.3. Argument Strength and a Criterion to Become a Belief

The revised argumentation standard sets forth the minimum criteria for an argument to impact the credibility of a claim. When a pair of premises can raise the subjective probability of a claim, they constitute a strong enthymeme or argument. The strength of the argument is thus determined by the initial believability of the claim. However, it is essential to recognize that a strong argument doesn't automatically render the claim believable. In some instances, it may only shift from negative certainty (0 probability) to uncertainty (0.5 probability), without actually convincing the individual about the claim. In this context, for the conclusion's credibility to be accepted, the combined incredibility of the premises must not exceed 0.5, as outlined in the following criterion:

$$(\sum \bar{P}(\text{premises}) < 0.5) \rightarrow (\Pi P(\text{premises}) > 0.5)$$

In essence, for the product of the premises' believabilities to surpass 0.5, the combined unbelievability of the premises must be less than 0.5. Consequently, a claim resulting from such an argument will be believed, irrespective of its prior believability. This prediction, termed the '*Criterion to Becoming Belief*,' will be empirically tested in the upcoming experiment.

Box 1: Conceptual Relations in the Revised Argumentation Standard with the Weakest Part Rule under Realistic Conditions



The figure illustrates the model's predictions for two different arguments concerning their relationship with the prior believability of the claim.

The left side of the figure displays a subjective probability scale from 0 to 1, with a central white line representing the threshold of uncertainty. Above this threshold lies the belief system (BS) space, where positive certainty reaches its maximum. Below this threshold, external contents to the system are found, with negative certainty at its lower extreme. The horizontal yellow line represents a prior probability assigned to a critical assertion (0.45).

The credibility of the premises in Argument 1 is 0.9 and 0.8, meeting the argumentation standard and resulting in an increase in the posterior credibility of the critical assertion to 0.72.

In Argument 2, although the argumentation standard is met (with values of 0.6 and 0.55, both exceeding 0.45), it fails to increase the subjective probability of the critical assertion. The resulting product of the premises' credibility is 0.33, lower than 0.45.

In this scenario, the individual may maintain their evaluation based on their unchanged prior credibility, or they may consider the presented premises, reducing their credibility in the critical assertion. Additionally, both variables could influence their evaluation, creating some tension between these forces.

The present study

1. Aims of this Study

The primary objective of this experiment is to empirically investigate and validate key hypotheses and predictions related to the proposed model of argumentation and credibility transfer. These hypotheses and predictions are pivotal for understanding deductive reasoning credibility dynamics, argument strength, and credibility transmission.

The Belief System is conceptualized as a network or argumentation system, where prior beliefs take the form of event representations or schematic rules of topical relations. These rules facilitate the relevant connection between different event representations within the belief system, ultimately serving as premises for argumentation. Conclusions drawn from these premises can

encompass both internal contents (already partially believed elements) and external contents (previously unaccepted elements). Subsequently, through the mechanisms of justification, these conclusions will either gain or lose credibility based on the believability of the premises.

This experiment aims to investigate the credibility enhancement of previously non-believed information and elucidate the underlying mechanisms facilitating this process. We hypothesize that these mechanisms involve the transmission of justification from already accepted representations of events (acting as data premises) to candidate representations seeking entry into the system, which are not yet believed (functioning as conclusions or claims). This transmission occurs through the rules of topical relations, which are also believed (acting as warrant premises). In this transmission, both the credibility of the data premise and the warrant premise are equally important and mutually constrain each other. Ultimately, what reaches the conclusion is a product of combining the credibility of both premises. We propose that the rule governing this mutual constraint of credibilities is probabilistic consistency, rooted in the principle of non-contradiction. Our hypothesis is that this consistent combination of credibilities is implemented through an arithmetic multiplication operation. If proven, this implies several consequences impacting the credibility limits that the system can assign to conclusion contents and the prerequisites an argument must meet to enhance the credibility of a representation or incorporate previously disbelieved information into the belief system.

In any case, the prior credibility of a claim plays a crucial role. If the product of the premises' credibility is lower than the prior credibility of the claim, it is not justified to abandon the latter to solely base the evaluation of the conclusion on the premises. In these cases, the posterior credibility may equal the prior credibility, suggesting that the premises do not influence the credibility of the conclusion. However, it is also possible for the premises to strongly influence the evaluation of the conclusion, resulting in a posterior credibility determined by the multiplication of the premises' credibilities. These trends may intermix, yielding evaluations that range between the multiplication of the premises' credibilities as a lower limit and the prior credibility as an upper limit. In such cases, the prior credibility could modify the posterior credibility, deviating it from the magnitudes expected by the multiplication, potentially surpassing the credibility of the least believed premise. Consequently, the posterior credibility may not be solely determined by the premises, and the prior credibility could account for part of the observed results. This demonstrates how the increase in posterior credibility of the claim relies on the product of the premises' credibility exceeding the prior credibility of the claim, akin to how the transition from disbelief to belief depends on the product of the premises' credibility being greater than 0.5.

These assumptions stem from the theoretical model of the belief system proposed in this thesis. The primary goal of this experiment is to empirically validate its psychological reality. To achieve this, we have organized these issues into hypotheses presented throughout the development of the theoretical framework in the preceding sections. Here, we offer an overview of the core hypotheses and predictions that will be tested in the study.

1. *Equitable Transfer Hypothesis*: It assumes that both premises play an equally important role in this transfer, that is, no premise has a privileged role over the other.
2. *Hypothesis of Multiplicative Combination*: It suggests that credibility is combined and transmitted through a multiplication operation in deductive reasoning.

Hypothetical Consequences of the Multiplication Operation: Expanding on the concept of multiplication, we aim to uncover the hypothetical impact on argument strength, particularly concerning its relationship with prior believability of the claim. This exploration will help us understand the implications of this operation on the overall belief system.

3. *Hypothesis of the Weakest Part Epistemic Rule*: This hypothesis asserts that the credibility of an argument is often determined by its least believed premise. Our goal is to examine the validity of this rule and understand how it operates in different argumentative scenarios.
 - a. *The Multiplication Exceeding Prediction*: In enthymematic arguments, this prediction suggests that the Weakest Part Rule will consistently apply when the product of the premises' believability exceeds the prior credibility of the critical assertion.
 - b. *The Multiplication Falling Below Prediction*: In non-enthymematic arguments, this prediction suggests that the Weakest Part Rule will not be as influential when the product of the premises' believability falls below the prior credibility of the critical assertion.
4. *The Enhanced Argumentation Standard Hypothesis*: It goes further than the basic argumentation standard by proposing that the multiplication of premise credibility should surpass the prior credibility of the conclusion. This prediction necessitates exploring suitable combinations of premise credibility for a given prior believability of the critical assertion. We propose that the appropriate combination is one in which the joint incredibility of the premises does not exceed the prior incredibility of the critical assertion.
5. *Criterion to Becoming Belief Hypothesis*: The aim is to assess argument strength in terms of the change in credibility required for the critical assertion to be accepted into the belief system.

Through systematic experimentation, we seek to gain insights into the dynamics of credibility transfer in deductive reasoning, the influence of argument structure on believability, and the validity of key epistemic rules. These objectives will contribute to a deeper understanding of the interplay between premises and conclusions in argumentation, shedding light on the complexities of belief modification in everyday reasoning processes.

2. General Description of the Present Experiment

To address these questions, an experiment was conducted where participants assessed the credibility of critical statements. These statements were presented either in isolation, to measure their *a priori* credibility, or as conclusions of arguments, to measure their *a posteriori* credibility. Each argument consisted of two premises. The credibility of the premises was manipulated. The primary focus of the study was to examine how the combination of premise credibilities influenced the credibility of the critical statements, considering the potential moderating effect of a priori credibility.

Critical statements concerning future events in the participants' surrounding world were constructed to be perceived as improbable, with the expectation of having *low* a priori credibility. An example of such a critical statement is:

'Advancements in medicine will lead to an increase in the number of people with some illness.'

Given that a priori credibility assessment of the critical sentences presented in isolation was needed, the experiment was divided into two main phases. The first phase, referred to as the 'Initial Credibility Establishment Phase,' aimed to empirically determine the values of a priori credibility. In the second phase, hereinafter the 'Experimental Phase', the influence of premises' credibility on the credibility of conclusions was assessed. It's important to note that the critical statements (or conclusions) remained identical across both phases, while the participants varied.

During the Initial Credibility Establishment Phase, all participants were presented with the critical statements. They were asked to assess the credibility of these statements using an 11-point subjective probability scale ranging from 0 (Not Probable at all) to 1 (Totally Probable), with increments of 0.1. The credibility of the two premises of the argument was manipulated based on the researcher's intuitions. During the Experimental Phase, the manipulation's effectiveness was tested through two main tasks: first, participants evaluated the premises' credibility, and second, the effect of the manipulation on the credibility of the critical statements was examined. In both tasks, participants were required to assess the credibility of the situations described in the sentences using the same scale described in the Initial Credibility Establishment Phase.

Arguments consist of:

1. A 'warrant' premise: a general rule of relation, perceived as causal and expressed in conditional form (e.g., 'if cause, then effect').
2. A 'data' premise: a statement describing a potential specific fact related to the everyday world surrounding the participants.
3. A conclusion / claim: constituting critical statement about a hypothetical future event, debatable and of initially expected low credibility.

All the arguments in this Experimental Phase were syllogisms of the first figure, without quantifiers or negations. Here is the traditional form of a syllogism of the first figure, along with the nomenclature used to identify the premises and conclusion in this experiment:

Warrant Premise:	Middle Term	Predicate
Data Premise:	Subject	Middle Term
Conclusion / Claim:	Subject	Predicate

Three terms can be identified: the subject of the conclusion, a middle term present in both premises but absent in the conclusion, and the predicate of the conclusion. To ensure a fair comparison of how premise credibility affects the conclusion, varying premise credibilities while maintaining the same conclusion across different variations was crucial. This approach prevents differences in credibility from being attributed to the statements themselves. Consistency in both subject and predicate between the conclusion and premises was maintained, even through various combinations of premise credibility, eliminating potential extraneous variables from wording differences.

Thus, the only way to vary premise credibility was through the modification of the middle term. When the middle term was semantically coherent with the subject or predicate of the premise, it resulted in high credibility (i.e., H). Otherwise, it led to low credibility premises (i.e., L). Below are examples of high credibility premises:

Warrant Premise: If the number of elderly individuals increases, then there will be an increase in the number of people with some illness.

Data Premise: Advancements in medicine are causing an increase in the number of elderly individuals.

The following sentences exemplify premises of low credibility:

Warrant Premise: If pharmacies start giving away contraceptives, then there will be an increase in the number of people with some illness.

Data Premise: Advancements in medicine have enabled pharmacies to start giving away contraceptives.

These examples show scenarios where both premises have either high or low credibility, yielding the same conclusion. Additionally, some combinations feature one premise with high credibility and the other with low credibility. We denote these four combinations using the following abbreviations:

Warrant	Data-premise	Abbreviation
High	High	HH
High	Low	HL
Low	High	LH
Low	Low	LL

Note that the first letter always refers to the warrant premise, while the second refers to the data premise. Thus, the credibility of the warrant premise (H and L) and the credibility of the data premise (H and L) were the two independent variables in the experiment. Additionally, the order of premise presentation was controlled, with half in warrant-first order and half in data-first order.

This study examines how qualitative reasoning quantifies non-numeric magnitudes. Participants' credibility scores reflect implicit confidence in presented situations and rules. The measured scores interpret confidence levels numerically. Explicit numerical translation increases the risk of numerical reasoning, which we aim to avoid by separating credibility assessment into two tasks. The first task involved evaluating the credibility of two separate blocks of sentences. One block contained warrant premises conditional sentences, while the other block included data premises affirmative sentences. The use of conditional sentences for warrant premises aimed to maintain consistency with argumentation tradition and because they are well-known structures in the psychology of reasoning. In the second task, participants evaluated the credibility of the conclusion based on the information provided by the premises. In this task, participants were presented with complete arguments as shown below and were asked to assign a credibility rating exclusively to the conclusion on the subjective probability scale. For example:

Given that:

Premise 1: Warrant or Data

e.g. If greenhouse gases continue to be emitted, then it will contribute to the extinction of polar bears.

Premise 2: Data or Warrant

e.g. Means of transport emits greenhouse gases.

Please estimate the probability of the following situation:

Conclusion

e.g. Means of transport will contribute to the extinction of polar bears.

3. Initial Credibility Establishment Phase

The determination of a priori credibilities for critical statements plays a pivotal role in our research for several fundamental reasons. Firstly, having knowledge of the a priori credibility of these statements is crucial for verifying whether they genuinely represent assertions that participants do not consider credible. This forms the basis for assessing the impact of the argumentative strategies we employ, particularly enthymemes, in altering the initial perception of item credibility. Furthermore, obtaining a priori credibility is necessary for the application of the 'enhanced argumentation standard' in argument evaluation and for calibrating argument strength. This information enables us to analyse with precision and rigor the effectiveness of argumentative strategies in credibility changes and to determine if a substantial increase in item credibility is achieved.

A priori credibility also plays a pivotal role in assessing our hypotheses and predictions. Knowing it allows us to test our expectations regarding the outcomes of multiplying premise credibilities. It also enables us to explore its role as an upper limit for posteriori credibility in situations where multiplication yields a value lower than a priori credibility. Ultimately, obtaining a priori credibilities is essential for comprehending their role in determining posteriori credibility across the various combinations of premise credibilities explored in our study.

Participants: Twenty-five third-year Psychology students from the Universitat Rovira i Virgili in Tarragona voluntarily took part in this study in exchange for receiving an additional bonus towards their final grade in a course. Their main age was 20.81 ($Sd = 4.95$).

Materials: A total of 24 critical sentences were employed. All sentences were about future events in participants' everyday surroundings.

It's important to note that the key objective was to have all critical items in this group perceived as having low credibility (i.e., subjective probability ratings below 0.5). However, the absence of high-credibility items could introduce biases into the distribution of credibility scores, potentially resulting in inflated ratings for these critical items that would have received medium-low scores if high-credibility items were present. Therefore, 18 filler items were included. The filler items had the same structure as the critical sentences, and their content was similar, but their credibility was intuitively high. All these materials (critical and filler items) are included in App. B-II (items for the Initial Credibility Establishment Phase).

Procedure: Participants were asked to estimate the subjective probability of 42 sentences (24 critical and 18 filler items) on an 11-point scale, where 0 indicated "Not Likely at All" and 1 "Completely Likely," with increments of 0.1. The instructions are presented in App. B-I (these instructions were used for both this task and the first task of the Experimental Phase).

Participants entered the laboratory booths in groups of up to four people and sat in front of the computers. After reading the instructions, they entered their identification details into the

program and completed four practice items. Subsequently, questions, if any, were addressed, and the experimenter left the room as participants commenced the task. Sentences were presented individually on the screen and remained until participants assigned a subjective probability score by pressing one of the buttons on the keyboard's top row. Stimulus presentation and response recording were performed using the DMDX program (Forster & Forster, 2003).

Results: The a priori credibility of the critical statements received an average rating below 0.5, as expected: $\bar{X} = .439$; $SD = .172$. However, not all items received ratings below 0.5. The data range was 0.62, with a minimum value of 0.16 and a maximum value of 0.78. The percentage of items above 0.5 was 37.5%. These results indicated the appropriateness of phrase selection. Despite the percentage of ratings exceeding 0.5, the average remained below this threshold. In this regard, we assume that these cases will not hinder the experiment's objectives.

4. Experimental Phase

Before going into the details of the experiment, we will present its main predictions, outlining them according to each hypothesis listed in the main study objectives section.

1. *Predictions for the Equitable Transfer Hypothesis:*

We assume that both premises hold equal relevance in determining the credibility of the conclusion, that is, both premises will significantly influence the credibility of the conclusion. When the data premise has high credibility, the conclusions should have higher values compared to situations where the data premise has low credibility. Similarly, when the warrant premise has high credibility, the conclusions should have higher values compared to situations where the warrant premise has low credibility.

Furthermore, we expect an interaction between the two independent variables. When one premise has completely high credibility, all variability in the credibility of the conclusion will be attributed to the credibility of the other premise. However, if one premise has completely low credibility, no credibility will be transmitted. Therefore, we expect to observe clear effects of the credibility of one premise when the other has high credibility. However, these effects will significantly diminish when the other premise has low credibility.

2. *Predictions for the Hypothesis of Multiplicative Combination:*

This model suggests that the credibility of the two premises combine through an operation similar to multiplying their credibilities. Assume the premises reach extreme credibility values, either completely credible (1) or totally lacking credibility (0). In the HH condition, where both premises are fully credible (1×1), the conclusion would achieve maximum credibility, equivalent to a value of 1. In contrast, in the HL, LH, and LL conditions, where at least one premise has zero credibility,

the credibility of the conclusions would be zero. Thus, the credibility of the conclusion in the HH condition would surpass that in other conditions, with no significant credibility differences among them. However, it is unlikely that individuals would assign extreme scores. In such cases, we would expect the credibility of the conclusion in the HH condition to be the highest, and in the LL condition to be the lowest. Furthermore, if there are no discernible differences in the mean credibility ratings between the HL and LH premises, we would not expect variations in the credibility of the conclusions in these conditions, as the order of factors does not alter the product. These outcomes would support the multiplication hypothesis.

Additionally, if the credibility of the conclusion stems from the product of the credibility of the premises, then significant correlations are expected between the credibility observed in the conclusions and the credibility predicted by the multiplication model, across the four credibility conditions of the argument (i.e., HH, HL, LH, and LL). The strength of this correlations may vary depending on the a priori credibility. As discussed in the objectives and hypotheses section, prior credibility can influence the credibility derived from the premises, particularly in cases of low credibility, thereby impacting the correlations mentioned.

3. *Predictions for the Hypothesis of the Weakest Part Epistemic Rule:*

This hypothesis postulates that the credibility of the less believed premise predominantly determines the credibility of the conclusion: if one of the premises has low credibility, low credibility will be attributed to the conclusion. In fact, due to the nature of multiplication, the credibility of the conclusion should not exceed that of the least believed premise. However, again, the a priori credibility of the conclusion may modulate this strict interpretation, depending on whether the a posteriori credibility is greater or less than the a priori credibility. If it is greater, we expect the strict rule to apply; if it is less, a more flexible interpretation is possible.

- a. *The Multiplication Exceeding Prediction:* In the HH condition, the credibility of the conclusion is limited by the credibility of the less credible premise, whether it be the data or the warrant, depending on the item. Although both premises are generally considered credible (with values above 0.5), the one with lower credibility defines the maximum level of credibility that the conclusion can attain. This dynamic directly influences the distribution of the conclusion's credibility. Therefore, since either premise can be the less credible in different scenarios, the credibility of the conclusion correlates with that of both premises.
- b. *The Multiplication Falling Below Prediction:* when one premise has low credibility, the multiplication of the credibilities of the premises may result in a value lower than the a priori credibility of the conclusion. Consequently, it is expected that this will also occur with the a posteriori believability of the conclusion. In such cases, participants could base their assessment of the credibility of the conclusion on both the credibilities of the premises and the conclusion's priori credibility. If the multiplication hypothesis and the Weakest Part Rule are met, we expect that in the HL and LH conditions, the credibility of the conclusion

will significantly correlate with the credibility of the low-credibility premises but not with the high-credibility ones. However, it is possible that a priori credibility will influence the observed credibility of the conclusions, causing a deviation from the strict interpretation of the Weakest Part Rule. That is, the effect of a priori credibility could be reflected in posterior credibility values of the conclusion that are higher than the credibility of the least believed premise.

Finally, in the LL condition, if participants do not completely disregard the information from the premises, the credibility of both premises may significantly correlate with that of the conclusion. However, since neither premise is considered credible in this case, there may be no premise that allows for clear variations in the credibility of the conclusion based on the credibility of the other premise. Consequently, it is possible that the credibility of neither premise will significantly correlate with the observed credibility in the conclusion. Thus, we do not have very clear predictions in this regard in this condition. On the other hand, here too, we expect that the posterior credibility of the conclusion will surpass the credibility of the least believed premise due to a priori credibility.

4. Predictions for the Range Limits of Posterior Credibility

The theoretical introduction posited that multiplication, a priori credibility, and the weakest part (i.e., the least believed premise) would delineate the limits of conclusion posterior credibility. In the HH condition, conclusion credibility is expected not to exceed that of the least believed premise or fall below the credibility predicted by multiplication. In other conditions, conclusion credibility is expected not to surpass a priori credibility or fall below the credibility predicted by multiplication.

5. Predictions on the Influence of Prior Credibility and Premise Multiplication on the Variability of Posterior Credibility

In the HH condition, an increase in the credibility of the conclusions is expected compared to the a priori credibility. This increase is solely attributed to the credibility of the premises. Consequently, it is inferred that the a priori credibility plays no significant role in explaining the observed variability in the credibility of the conclusions. On the other hand, it is expected that in the L conditions (i.e., HL, LH and LL), the posterior credibility would be equal to or lower than the a priori credibility. If it is lower, it has been assumed that it could surpass the credibility of the least believed premise due to the effect of the a priori credibility. Therefore, in the L cases, the a priori credibility should account for a significant portion of the variability.

Overall, the distribution of posterior credibility levels may be influenced by the a priori credibility, the credibility of each premise, and the multiplication as an integrated combination of that credibility. It is expected that the a priori credibility will be the weakest predictor, and the multiplication will be stronger than each individual premise, but equally strong as both together, as it integrates both. All these predictions will be tested through regression analysis.

6. *Predictions for the Enhanced Argumentation Standard Hypothesis:*

The strength of the argument is assessed by an increase in the conclusion's credibility between the prior and the posterior credibility. For this increase to occur, the premises must have a higher credibility than the conclusion in its initial state. However, this alone is not sufficient. It is also necessary to establish a criterion for the product of the premises' credibilities to exceed the a priori credibility of the conclusion. This criterion states that the joint incredibility of the premises should not surpass the a priori incredibility of the conclusion. Additionally, the lower the joint incredibility of the premises compared to the a priori incredibility of the conclusion, the greater the increase in the conclusion's posterior credibility, indicating a stronger argument. Therefore, we expect negative correlations between the degree of joint incredibility of the premises relative to the a priori incredibility of the conclusion and the extent of increase in the posterior credibility of the conclusion relative to the a priori credibility.

7. *Predictions for the Criterion to Becoming Belief Hypothesis:*

The final prediction focuses on the requirements for an argument to lead to the acceptance of a content within the belief system. As mentioned earlier, if the joint incredibility of the premises is less than 0.5, we anticipate that the result of the multiplication will be greater than 0.5. Since we expect the system to combine the credibility of the premises in a manner similar to multiplication, we always anticipate that when the premises meet this criterion, the credibility of the conclusion will be greater than 0.5, being therefore accepted.

4.1. Method

Participants: Sixty-four third-year psychology students from the Universitat Rovira i Virgili in Tarragona participated voluntarily in this experimental phase of the study to earn an additional bonus in their final course grade. The mean age was 20.04 (SD = 6.41). Only two participants had received formal logic training; however, their responses did not significantly deviate from the overall pattern. Hence, their data were included in the analysis. Participants in the Experimental Phase were not the same as those in the Initial Credibility Establishment Phase, though both groups belonged to the same population (psychology students from the URV).

Design: We conducted a fully factorial repeated measures design with a 2 (believability of the warrant premise: high vs. low) x 2 (believability of the data-premise: high vs. low) x 2 (order of the premises: warrant-first vs. data-first) configuration. Items (critical assertions) and participants were systematically rotated across all believability and order conditions, resembling a Latin square design. This ensured that no participant encountered the same critical assertion in more than one believability condition or order. Apart from that, some analyses involved the comparison with prior believability values. In this case, there was a single 2 (argumentative treatment: prior vs. posterior) intra-items factor. The dependent variable was the participants' estimated subjective probability of the critical assertions on an eleven-point scale ranging from 0 to 1 (i.e. identical to the one used in the Initial Credibility Establishment Phase).

Materials: The same 24 critical sentences described in the previous phase were employed, this time for use as conclusions in the experimental phase's arguments. Each conclusion was paired with four arguments, one in each of the four experimental conditions. This generated a total of 96 arguments (available in App. B-II; arguments used). Table 7 presents an example of the four possible combinations of premise credibilities for one conclusion.

Table 7

Example of the Materials in their Four Believability Conditions

Data-premise	Warrant Premise	
	High	Low
High	If greenhouse gases continue to be emitted, then it will contribute to the extinction of polar bears.	If the import and export of products is facilitated, then it will contribute to the extinction of polar bears.
	Means of transport emits greenhouse gases.	Means of transport facilitate the import and export of products.
Low	If poaching of large mammals is favoured, then it will contribute to the extinction of polar bears.	If mobile phone sales increase, then it will contribute to the extinction of polar bears.
	Means of transport favours the poaching of large mammals.	Means of transport increase mobile phone sales.
Claim	Means of transport will contribute to the extinction of polar bears.	

Note: Here, only the four believability conditions resulting from the combination of the warrant and data premises' believability are shown; order manipulation is not included. Furthermore, it is possible to observe how the mere modification of the middle term results in differences in premise credibility. In some cases, the middle term is coherent in both premises, other times in one but not the other, and, finally, at times in neither of the two.

Each argument consists of two premises: a conditional warrant premise and an affirmative data premise, equally divided into high (H) and low (L) credibility. Some arguments were taken from Experiment 1, but new conclusions and premises were also added based on the researcher's intuitions regarding their credibility.

In the first task of the experiment, in which participants assessed the credibility of premises, 20 filler items (10 conditional and 10 affirmative sentences, each with 5 high credibility and 5 low credibility sentences) were included, along with 6 practice trials.

Procedure: the experiment comprised two main tasks, with a distractor task between the two tasks. In both main tasks, participants were required to provide subjective probability estimates for a series of written sentences. The DMDX software (Forster and Forster, 2003) was used to display the written materials on the computer screen and record participants' responses. As in the Initial Credibility Establishment Phase, participants rated the subjective probability of sentence content

using an 11-point scale, which ranged from 0 (not likely at all) to 1 (completely likely), with intervals of 0.1, by pressing the corresponding keys on the keyboard.

The *first task* involved assessing the subjective probability of both data and warrant premises. These premises were displayed one at a time on the screen, remaining visible until participants rated them using the subjective probability scale. Since presenting a data premise immediately after a warrant premise could lead participants to generate a possible conclusion, we chose to present the two types of premises separately, i.e. in two blocks: one for warrant premises and another for data premises. These blocks were presented consecutively, with the second block starting immediately after the completion of the first one, creating a seamless transition for participants. This setup aimed to make the premises appear as isolated sentences rather than as argument premises. The order of block presentation was counterbalanced across participants, and sentence presentation within each block was randomized. To further mask the transition between blocks, we included the filler items in each block, aligning with the grammatical structure of the opposite block. For instance, affirmative sentences were included in the warrant premise block, while conditional sentences were included in the data premise block. In this first task, each participant assessed a total of 68 sentences: 24 warrant premises, 24 data premises plus the 20 fillers already described in the materials section. Half of the warrant premises, as well as half of the data premises had high subjective probability while the other half had low subjective probability.

In the *intermediate task*, participants engaged in the Masked Figures Test (Witkin, Oltman, Raskin, & Karp, 1971), a paper-and-pencil distractor. The test was positioned face down beside each computer. The test involved a non-verbal challenge, in which participants were tasked with identifying concealed figures within other figures. Participants completed this task over a five-minute period.

In the *main task*, each participant saw a total of 24 arguments, 6 from each credibility condition. Within each condition, 3 arguments had the warrant-first, and 3 had the data-first order. The order of argument presentation was randomized for each participant. Participants advanced through the elements of the argument using the spacebar. First, they saw "Given the probability that:". Next, the first premise (warrant or data) appeared, followed by the second premise. Finally, "Please estimate the probability of the following situation:" was shown, along with the conclusion and the scale. Each element remained on the screen until the participant assessed the subjective probability of the conclusion. Participants conducted the experiment individually, entering the laboratory booths in groups of up to four people at a time. Once seated in front of the computer screen, they were informed that they would be participating in an informal reasoning experiment. They were asked to estimate subjective probabilities regarding certain situations presented on the screen. The experimenter emphasized that there were no right or wrong answers and that their responses should be based solely on their own beliefs about the given situations. It was crucial that participants felt completely free to evaluate the statements. To ensure this, it was explicitly mentioned that there were no restrictions on how they could interpret the premises, and their

responses should reflect their genuine beliefs. The study was conducted anonymously to further encourage this type of response. The experimenter ensured participant anonymity by providing an experiment form with an identification code.

After giving participants time to read the instructions for the first task, the experimenter reiterated them verbally and addressed any questions. Participants completed 4 practice trials to familiarize themselves with the task under the experimenter's supervision. Once the practice trials were done, the experimenter confirmed participants' understanding by asking them about their estimates. After finishing the first task, participants engaged in a distractor task. After five minutes, the experimenter returned to introduce the instructions for the second task, which involved 6 practice trials. Before leaving the room, the experimenter emphasized considering the 'context sentences' (i.e. the premises) when estimating the third sentence.

4.2. Results

In what follows, we will first present the results of the initial task of the experimental phase. Subsequently, we will present the results of the main task. The results will be organized according to each research hypothesis. To enhance readability, a succinct summary of the prediction associated with each hypothesis will be provided.

4.2.1. Results of the First Task

The descriptive statistics for the credibility ratings of the premises obtained in Task 1 are represented in Table 8.

Table 8
Believability Ratings of the Premises

	Bel. Condition	Premises		Average
		Warrant-Premise	Data-Premise	
<i>a</i>	HH	.68 (.12)	.70 (.12)	.69 (.09)
	HL	.75 (.10)	.26 (.14)	.51 (.09)
	LH	.20 (.13)	.78 (.12)	.49 (.08)
	LL	.14 (.08)	.30 (.13)	.21 (.09)
<i>b</i>	Total Averages			
	High	.72 (.08)	.74 (.10)	.73 (.07)
	Low	.17 (.08)	.28 (.09)	.22 (.06)

Note: This table presents the mean subjective probability rating for the premises in all conditions. At the top of Table (a), we have the first column (Bel. Condition) representing the four argument credibility conditions. Moving to the right, we find columns displaying the average credibility (and SD) of the warrant and data premises for each credibility condition of the argument. The final column (Average) indicates the combined average credibility of both premises for each condition.

At the bottom of Table (b), the total averages for the High and Low premises are presented. In the warrant premise column, you can see the average of warrants in the High credibility condition (i.e., the average of warrants in HH and HL) and in the Low credibility condition (i.e., in LH and LL). Similarly, total averages for the credibility of data premises in the High and Low credibility conditions are provided. Finally, in the last column, you will find the overall average of all premises with both High and Low credibility.

The results show that highly credible premises obtained subjective probability scores above 0.5 (0.72 for the warrant and 0.74 for the data-premise), while low credibility premises received scores below 0.5 (0.17 for the warrant and 0.28 for the data-premise). When grouping premises by credibility conditions, HH had the highest mean (0.69), while LL had the lowest (0.21), with HL (0.51) and LH (0.49) falling in between. As can be observed, the data align with the expected outcomes, thereby validating the manipulated independent variables.

4.2.2. Results of the Second Task

Results Concerning the Equitable Transfer Hypothesis

A *main effect* is expected for both warrant and data premise credibility, as this model assumes their equal relevance in influencing the credibility of the conclusion. When either the data or the warrant premise has high credibility, it is expected to result in a higher credibility of the conclusion compared to situations where either premise has low credibility. An *interaction* between the two independent variables is also expected. When one premise has high credibility, the conclusion's credibility varies with the other premise. Conversely, if one premise has low credibility, minimal credibility is transmitted. Clear effects of one premise's credibility are expected when the other has high credibility but diminish notably when it has low credibility.

A repeated measures ANOVA with 2 (warrant believability: high vs. low) x 2 (data-premise believability: high vs. low) x 2 (premise order: warrant-first vs. data-first) was conducted. The analyses were conducted both by subjects and by items. Table 9 displays the descriptive statistics of credibility ratings for the conclusions in this Task.

No main effect was found for the *premise order* factor, $F_1(1, 63) = .01, p = .922, \eta_p^2 = .00$; $F_2(1, 23) = .00, p = .934, \eta_p^2 = .00$. Furthermore, the premise order factor did not interact with any of the other factors (all $ps > .1$). This suggests that this factor plays no role in determining believability. Consequently, it is omitted from Figure 15, which presents the overall pattern of results (see below).

Table 9

Conclusions' Subjective Probability Scores as a Function of Order and Believability Conditions

Order Conditions	Argument Believability Conditions				Average
	HH	HL	LH	LL	
Warrant-First	.56 (.15)	.40 (.14)	.27 (.12)	.22 (.09)	.37 (.07)
Data-First	.57 (.15)	.38 (.14)	.28 (.10)	.21 (.10)	.36 (.08)

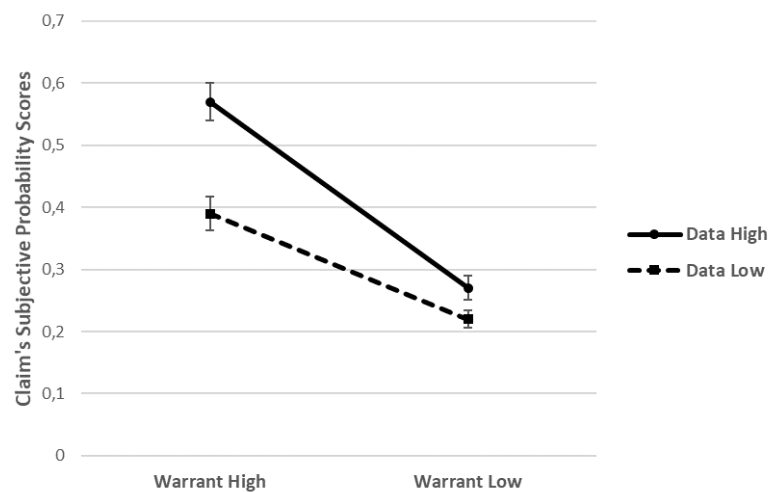
Note: This table shows the mean scores (and standard deviations) of claims' subjective probabilities in the four conditions of believability resulting from the combination of both warrant and data-premises believability (HH, HL, LH and LL), together with the order conditions (Warrant-First and Data-First). The last column (Average) shows total mean scores for warrant-first and data-first orders.

A significant main effect emerged for the *warrant's believability* factor, $F_1(1, 63) = 31.44, p < .001, \eta_p^2 = .83$; $F_2(1, 23) = 66.98, p < .001, \eta_p^2 = .74$. When the warrant is highly believable, the claim's believability ($M = .48$) is higher compared to when the warrant is of low believability ($M = .25$).

Similarly, a significant main effect was observed for the *data-premise believability* factor, $F_1 (1, 63) = 82.35, p < .001, \eta_p^2 = .57$; $F_2 (1, 23) = 75.97, p < .001, \eta_p^2 = .77$. When the data-premise is highly believable, the claim's believability ($M = .42$) is higher compared to when the data-premise is of low believability ($M = .3$). A significant interaction between the *warrant's believability* and the *data-premise's believability* was observed, $F_1 (1, 63) = 24.29, p < .001, \eta_p^2 = .27$; $F_2 (1, 23) = 11.47, p = .003, \eta_p^2 = .33$. It's worth noting that all obtained values of the effect size exceed .14, indicating large effects (Cohen, 1992).

Figure 15

Graphical Illustration of the Claim's Averaged Scores as a Function of Believability Condition



Note: The graph illustrates the two main effects together with an interaction effect. The means represented in this graph are detailed numerically in Table 11, below.

To understand the nature of this interaction, post hoc analyses were conducted to compare the effects of premise credibility at specific levels, all of them corrected using Bonferroni corrections. Paired t-tests were performed to compare the effects of data-premise credibility across warrant credibility levels. The results of the paired t-tests revealed significant differences between HH and HL, $t_1 (63) = 9.16, p < .001$; $t_2 (23) = 7.22, p < .001$ ($d = 1.30$), as well as between LH and LL, $t_1 (63) = 3.57, p = .001$; $t_2 (23) = 2.90, p = .008$ ($d = .60$). In either case, Cohen's d indicated a large effect in the comparison between HH and HL, while only a moderate effect was observed in the comparison between LH and LL. Furthermore, paired t-tests were also conducted to examine the effects of warrant premise credibility across data-premise credibility levels. Significant differences were observed between HH and LH, $t_1 (63) = 15.63, p < .001$; $t_2 (23) = 7.82, p < .001$ ($d = 2.46$), as well as between HL and LL, $t_1 (63) = 10.16, p < .001$; $t_2 (23) = 6, p < .001$ ($d = 1.60$).

Results Concerning the Hypothesis of Multiplicative Combination

The model suggests that premise credibility is transmitted to the conclusion through a multiplication-like operation. Extreme scenarios, where the conclusion is deemed a priori impossible and both premises are considered necessary, lead to a posteriori necessity of the conclusion. However, such extremes are unlikely in everyday life. Instead, predictions are proposed considering more realistic scenarios with low a priori probability for the conclusion and varying probability for the premises. Bearing this in mind:

1. It is expected that in HH, conclusion credibility will be the highest; in LL, the lowest; and in HL and LH, intermediate.

Posthoc T-tests in the previous section partially covered the first of these predictions, specifically comparing HH with HL and LH, and LL with HL and LH. For the missing comparison between HH and LL, a new T-test was conducted, yielding $t_1(63) = 17.38, p < .001$; $t_2(23) = 11.79, p < .001$ ($d = 3.09$). Observations revealed the highest conclusion credibility in HH, the lowest in LL, and intermediate values in HL and LH.

2. If there are no differences in the average credibility of the premises between HL and LH, no differences in the credibility of their respective conclusions are expected.

Regarding the second prediction, we first checked for differences in the mean credibility of the premises between HL and LH. Analyses revealed differences among subjects but not among items for both H premises ($t_1(63) = -2, p = .05$; $t_2(23) = -.8, p = .434$) and L premises ($t_1(63) = 3, p = .004$; $t_2(23) = -1.44, p = .162$). Next, we compared the mean credibility of the conclusions between HL and LH. Significant differences emerged between the credibility levels achieved in the conclusions of HL and LH ($t_1(63) = 7.09, p < .001$; $t_2(23) = 3.50, p = .002$ ($d = 1.03$).

3. A positive and significant correlation is expected between the credibilities reported by participants (observed scores) and those predicted by the multiplication model. Additionally, if the a posteriori believability of the conclusion surpasses its a priori believability (i.e. the HH case), it should rely solely on the combination of premises' believability. Consequently, a higher correlation is expected in the HH condition compared to others. Finally, as multiplication integrates both premises' believability, it is expected to correlate more consistently with the conclusion's believability than each premise separately.

Regarding this composite prediction, a series of correlations were computed between the subjective probability scores observed in the conclusions with: the credibility scores of the warrant; the credibility scores of the data; and the credibility scores predicted by the multiplication model.

The procedure involved the following steps. Initially, the mean believability score for each claim within each believability condition was computed across all participants, yielding the *observed* believability scores for the claims. Subsequently, the mean believability scores for the warrants and

data-premises of each argument within each believability condition were determined by considering the participants' scores. The *predicted* believability of the claims was then calculated by multiplying the believability values of the warrants by the corresponding values of the data-premises for each argument in every believability condition. Finally, correlations were computed between the 24 observed scores and the 24 warrants' scores, the 24 data premises' scores and the 24 predicted scores for each of the four argument believability conditions.

These (by item) correlations are detailed in Table 10 (first row):

Table 10

Correlations between Observed vs. Model-Predicted, Warrant and Data Premises' Believabilities across Argument Conditions

	HH	HL	LH	LL
Multiplication	.90**	.69**	.71**	.38*
Warrant	.65**	.24	.65**	.15
Data-premise	.67**	.63**	.02	.43*

Note: () $p < .05$; (**) $p < .001$*

Focusing on these last predictions, Table 10 shows significant correlations between observed and multiplication model-predicted scores for claims across all conditions. Particularly, in the HH condition, where the argumentation standard is met⁶⁰, this correlation reaches 0.90, the highest observed. Additionally, a visual inspection reveals a clear superiority of the correlation in HH compared to the other conditions.

Finally, it is observed that the correlation between the ratings predicted by the multiplication model and the observed ratings was significant in all conditions, whereas the same did not occur with the premises taken separately.

Results Concerning the Hypothesis of the Weakest Part Epistemic Rule

This hypothesis suggests that the credibility of the less-believed premise mainly influences the credibility of the conclusion. Strictly speaking, the conclusion's credibility should not exceed that of the least-believed premise due to multiplication. However, the a priori credibility of the conclusion may modify this strict rule, depending on whether the result is greater or less than the a priori credibility. If greater, we expect the strict rule to apply; if less, a more flexible interpretation can be taken.

This hypothesis will be subdivided into two more specific predictions: the *Multiplication Exceeding Prediction* and the *Multiplication Falling Below Prediction*. These predictions will be

⁶⁰ The a posteriori believability of the conclusion in HH surpasses its own a priori believability. But this will be developed further on.

explored in subsequent sections, involving comparisons between the values of a priori credibility (acquired in the Initial Credibility Establishment Phase) and the values of a posteriori credibility in the four conditions (obtained in the Experimental Phase). Although the means of the four posterior credibility conditions of the conclusions and the a priori credibility have been previously presented in previous sections, we recapitulate them in the following Table 11 to aid readability.

Table 11

Posterior and Prior Claims' Subjective Probability Scores

	Argument Believability Conditions				Prior Bel.
	HH	HL	LH	LL	
Means (Sd)	.57 (.14)	.39 (.13)	.27 (.09)	.22 (.07)	.44 (.17)

Note: Under the label 'Argument Believability Condition,' the averages (and their standard deviations) of subjective probability obtained under different argument credibility conditions during the Experimental Phase are depicted. The label 'Prior Bel.' presents the average and standard deviation of subjective probability obtained in the 'Initial Credibility Establishment Phase,' serving as the a priori credibility.

Results Concerning the Multiplication Exceeding Prediction: We assume that the weakest part epistemic rule strictly applies when the multiplication, and thus the posterior credibility of the conclusion, surpasses the prior credibility, specifically in the HH condition. This implies that, in each HH item, the credibility of the conclusion will not exceed that of the least believed premise. Consequently, the credibility of both premises shapes the credibility of the conclusions, with a correlation expected between them.

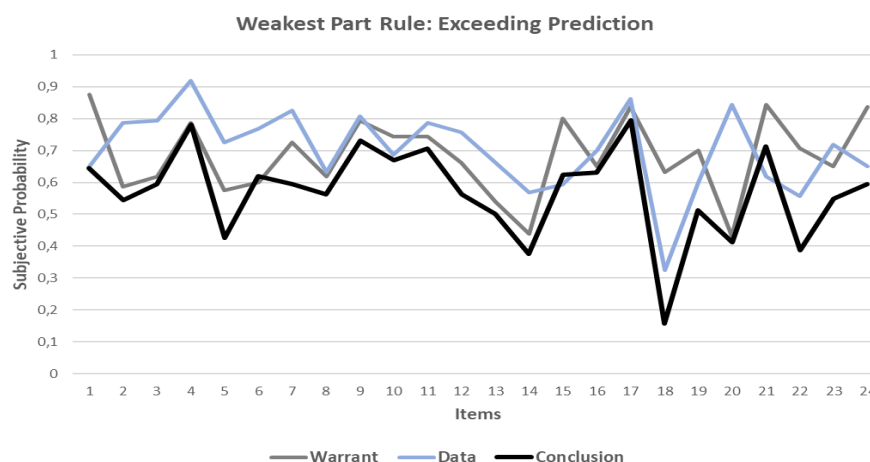
First, a paired-samples t-test was conducted to test if the posterior credibility of the conclusions exceeds the prior credibility in HH. The analysis revealed significant differences between the two factors, $t(23) = -5.23, p < .001 (d = .82)$. Second, the number of items in HH where the credibility of the conclusions was lower than that of the least believed premise was examined. This examination revealed that 87% of the conclusions (i.e., 21 out of 24) received subjective probability ratings lower than those of the least believed premise. The remaining three scores were only slightly higher than the least believed premise, specifically with differences of 0.01, 0.03, and 0.09, respectively. Additionally, a paired-samples t-test was conducted to determine if the credibility of the conclusions was significantly lower than that of the least believed premises ($M = .62$). The result of the t-test analysis indicated that this difference was statistically significant⁶¹, $t(23) = 4.07, p < .001$.

Figure 16 graphically illustrates the subjective probability scores of both premises where we can find the least believed premise in each case, demonstrating how the credibility of the conclusions remains below that of the least believed premise.

⁶¹ Since all that matters here is whether the means are equal or whether one is statistically higher or lower than the other, Cohen's d will not be calculated.

Figure 16

Strict Conformity of Conclusions with the Weakest Part Rule in HH



Note: This graph illustrates that the subjective probability of the conclusions (black line) has consistently been lower than that of the least believed premise in each item, whether it is the warrant premise (grey line) or the data premise (blue line). The alternating nature of both premises as the least believed highlights their equal influence on the credibility of the conclusion.

Finally, significant correlations were found between the credibility of the conclusion and the credibility of both premises, as indicated in Table 10.

Regarding the Issue of the Range Limits of Posterior Credibility in HH

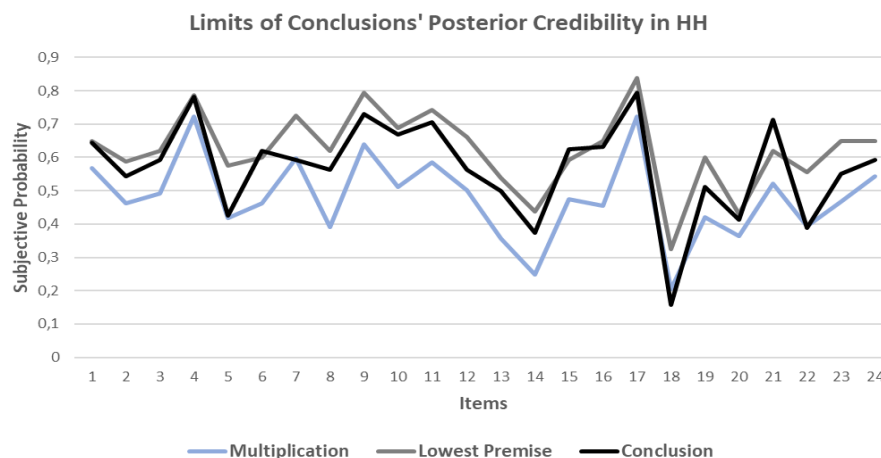
The analysis from the previous section revealed that, in the HH condition, the credibility of the conclusions did not exceed that of the least believed premise, and it is also expected not to fall below the credibility predicted by the multiplication hypothesis.

Indeed, it was observed that 87% of the conclusions (i.e., 21 out of 24) obtained a posteriori credibility higher than that estimated by multiplying the credibility of the premises. The remaining three conclusions had a credibility equal to that estimated by multiplication (with specific differences of 0.004, 0.005, and 0.04, respectively). Additionally, the mean credibility observed in the conclusions (see Table 11) was compared with the mean credibility predicted by the multiplication hypothesis ($M = .48$). The analysis revealed a significant difference between the two ($t(23) = -6.95, p < .001$). Therefore, the a posteriori credibility of the conclusions in the HH condition has the least believed premise as its upper limit and the credibility predicted by the multiplication hypothesis as its lower limit.

The graph presented in Figure 17 illustrates this result.

Figure 17

Upper and Lower Limits of Conclusions' Subjective Probability in HH



Note: The graph illustrates that the credibility of the conclusions (black line) typically falls within the bounds set by the least believed premise at each point (grey line) as its upper limit, and the multiplication of the premises' credibility (blue line) as its lower limit. The grey line represents the minimum value between the two HH premises at each point.

Results Concerning the Multiplication Falling Below Prediction: In the L conditions (HL, LH, and LL), where the product of the premises' credibility may be lower than the a priori credibility, it is possible for the premises to be disregarded when evaluating the conclusions as they do not constitute convincing arguments. In such cases, the posterior credibility would equal the prior credibility. However, if participants assess the conclusions considering the premises, the posterior credibility would reflect the product of the premises' credibility, being influenced by the least believed premise. If some participants dismiss the premises while others consider them, the effects are averaged in the results. This could lead to the multiplication effect being reflected, thereby influencing the conclusions with the least believed premise, possibly even surpassing it due to the influence of the a priori credibility.

First, comparisons were conducted using the Bonferroni-adjusted Student's t-test (i.e., $p < .017$) to determine whether the posterior credibility of conclusions under the L conditions equalled or was lower than the prior credibility. Student's t-tests revealed significant differences between the mean of prior credibility and posterior credibility in LH, $t(23) = 4.44$, $p < .001$ ($d = 1.21$); and in LL, $t(23) = 5.94$, $p < .001$ ($d = 1.67$). However, this was not the case in the HL condition, $t(23) = 1.53$, $p = .138$.

When examining the correlations between the posterior credibility of the conclusions and the premises in the HL and LH conditions (see Table 10), significant correlations were found with the

least believed premise (L), and not with the most believed (H). In the LL condition, the posterior credibility correlated significantly with the data premise, which was the more believed of the two⁶².

To ascertain if the posterior credibility of the conclusions in the L conditions exceeded that of the least believed premise, the number of conclusions that were equal to or less than premise L was tallied. Only 16% met this criterion (i.e., 4 out of 24). Subsequently, Student's *t*-tests were conducted to compare the means of the subjective probability of the conclusions with the means of their respective least believed premises. The averages of the conclusions were presented in Table 11. Regarding the averages of premise L, they were $M = 0.26$ in HL, $M = 0.20$ in LH, and in LL, the least believed of the two was $M = 0.14$ (all these means were presented in Table 8). All comparisons exhibited significant differences, with $t(23) = -5.33$, $p < .001$ in HL; $t(23) = -3.42$, $p = .002$ in LH; and $t(23) = -4.1$, $p < .001$ in LL, respectively.

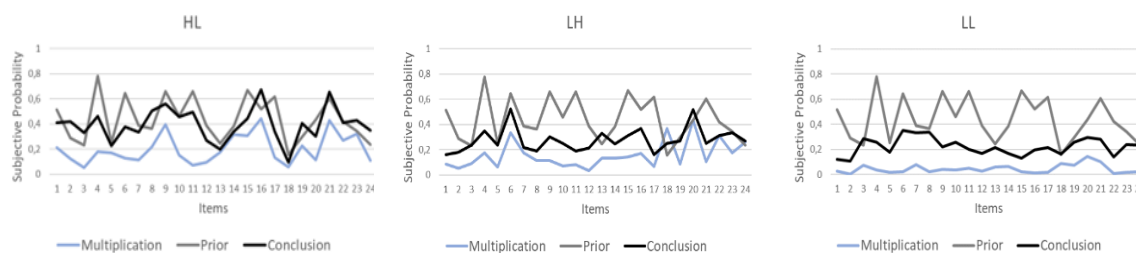
Regarding the Issue of the Range Limits of Posterior Credibility in L conditions

The arguments containing a premise L lack sufficient strength to increase the posterior credibility of the conclusions, implying that the latter must be equal to or lower than the a priori credibility. As previously discussed, the averages of LH and LL were significantly lower than the a priori credibility, while that of HL was equal to it.

Furthermore, it was expected that the posterior credibility of the conclusions would not be lower than the credibility predicted by the multiplication hypothesis. As demonstrated earlier, the averages of the L conditions exceeded the least believed premise, which, in turn, has a credibility equal to or higher than the product of multiplying it by the other premise. Therefore, the posterior credibility of the conclusions in L conditions was always greater than that established by the multiplication. These results are presented graphically in Figure 18.

Figure 18

Upper and Lower Limits of Conclusions' Posterior Subjective Probability in L Conditions



Note: The graph illustrates that the credibility of the conclusions (black line) typically falls within the bounds set by the a priori credibility (prior: grey line) as its upper limit, and the multiplication of the premises' credibility (multiplication: blue line) as its lower limit.

⁶² Correlations were also conducted at the subject level ($n = 64$). These correlations are not reported here due to mostly redundant information. However, it is important to note that in the LL condition, observed scores correlated with the resulting subjective probability of multiplication ($r = .40$; $p < .001$), the data premise ($r = .39$; $p < .001$), and also the warrant ($r = .39$; $p < .01$). In other words, similar to what occurred in HH, here it correlated with the subjective probability ratings of both premises.

Collectively, these data show that the credibility scores of conclusions in the L conditions have not surpassed a priori credibility, nor fallen below the credibility expected by the multiplication model.

Results Concerning the Influence of Prior Credibility and Premise Multiplication on the Variability of Posterior Credibility

In the HH condition, an increase in the credibility of the conclusions is expected solely due to the credibility of the premises, implying minimal influence of the a priori credibility. In the L conditions, it is expected that the posterior credibility will be equal to or lower than the a priori credibility; if lower, it could surpass the credibility of the least believed premise due to the effect of the a priori credibility. Overall, the distribution of posterior credibility can be influenced by the a priori credibility, that of each premise, and the multiplication as an integrated combination of these factors. The a priori credibility is expected to be a weaker predictor compared to the multiplication, which is predicted to be stronger than each premise individually and equally strong as both combined, as it integrates them. Multiple linear regressions were conducted to investigate these aspects. The subjective probability of the conclusions, also referred to as posterior credibility, was set as the dependent variable. Predictors included prior credibility, warrant credibility, data premise credibility, and multiplication.

The multiplication of premise credibilities is closely linked to the individual premises' credibilities, as it depends entirely on them. Collinearity tests revealed tolerances below .1 between the multiplication and premise credibilities (both in HH; with L in HL, and LH conditions; and with the data premise in LL). To address this collinearity, we conducted two separate regression analyses for each believability condition, following López-González (1998)'s recommendation. One analysis included both premises' credibilities and the a priori credibility of the conclusion⁶³, while the other incorporated the multiplication result and a priori credibility. This approach ensured that all tolerances remained within an acceptable range (between .60 and .96)⁶⁴. It allowed for a more precise examination of how these predictors influence the credibility scores of the conclusions in each condition. The results of these analyses are presented in Table 12.

⁶³ We posited that the relationship between the predictors 'data premise' and 'warrant premise' is *multiplicative*, rather than *additive*. This implies that if one of the predictors is low, the dependent variable will also be low, regardless of the other predictor. However, the results obtained after applying a logarithmic transformation to the predictors did not differ significantly from the untransformed values in the multiple regression analysis. Therefore, we chose to present the analyses using the original values.

⁶⁴ By another hand, the Durbin-Watson statistic was close to 2 in all cases. Strictly speaking, the assumption of independence is fulfilled if the Durbin-Watson statistic is between 1.5 and 2.5. The HL, LH and LL conditions clearly meet this requirement, while HH (with 2.7) is very close to it.

Table 12

Multiple Regression Analysis for the Models Constructed for the Premises and for the Multiplication, Including the A Priori Credibility Factor

Condition	Premises Model			
	Predictor	β weight	p -Value	Adjusted R^2
HH	A Priori Bel.	.25	< .050	.83
	Warrant Bel.	.49	< .001	
	Data Bel.	.52	< .001	
HL	A Priori Bel.	.42	= .010	.55
	Warrant Bel.	.16	ns	
	Data Bel.	.50	< .010	
LH	A Priori Bel.	.39	< .050	.48
	Warrant Bel.	.72	< .001	
	Data Bel.	-.06	ns	
LL	A Priori Bel.	.04	ns	.06
	Warrant Bel.	.07	ns	
	Data Bel.	.40	ns	

Condition	Multiplication Model			
	Predictor	β weight	p -Value	Adjusted R^2
HH	A Priori Bel.	.25	< .050	.83
	Multiplication	.74	< .001	
HL	A Priori Bel.	.41	< .010	.60
	Multiplication	.55	= .001	
LH	A Priori Bel.	.27	ns	.53
	Multiplication	.73	< .001	
LL	A Priori Bel.	.14	ns	.09
	Multiplication	.40	= .058	

We conducted multiple regression analyses using the “enter” method, which includes all predictors initially. While this approach does not optimize the model, it offers a detailed understanding of each predictor’s contribution to the dependent variable.

The table indicates that prior credibility plays a significant role in explaining the variance under HH conditions, with a β weight of .25 in both the premises and multiplication models. In the HL condition, this weight increases to .42 and .41, respectively in the two models. In the LH condition, the β weight is .39 for the premises model but does not reach significance in the multiplication model. However, in LL, prior credibility does not make a significant contribution in either model.

Therefore, it significantly influences the HH and HL conditions and, in the premises model, LH as well, but does not appear to influence the LL condition in any model.

It is observed that In HH, both premises contribute significantly and equally to explaining the variance, while in HL and LH, only the L premise is significant, and in LL, no premise contributes. On the other hand, multiplication is significant in all conditions except in LL. Moreover, the R^2 values are similar across both models, albeit slightly higher in the multiplication model. Among all predictors, prior credibility presents the lowest β weights, and multiplication the highest, with the premises occupying a significant mid-high position.

In terms of explanatory capacity (R^2), both models in the HH condition achieve high levels of variance explanation (83%). This capacity slightly decreases in conditions with an L premise (HL and LH), showing values between 60% and 50%, and markedly declines in LL⁶⁵.

Results Concerning the Enhanced Argumentation Standard Hypothesis

Under the Enhanced Argumentation Standard Hypothesis, argumentation seeks to bolster the credibility of a claim via persuasive reasoning, requiring the premises to be more credible than the claim itself. However, this is not enough. A criterion posits that for argumentation to effectively enhance credibility, the cumulative incredibility of the premises must be equal to or less than the initial incredibility of the claim. Thus, the lower the combined incredibility of the premises relative to the claim's initial incredibility, the more they can enhance the claim's credibility as a conclusion. This establishes a crucial link between the premises' credibility and the conclusion's prior credibility for an effective increase in argumentative credibility.

To rigorously test this hypothesis, we adopted a structured methodology as outlined below, showing the meticulousness with which we approached the verification of our assumptions:

- a) The a priori subjective probability of critical assertions, termed 'Prior', from the Initial Credibility Establishment Phase. For example, the claim "Transportation will contribute to the extinction of polar bears" had a Prior of 0.38.
- b) The subjective probability for the warrant premise, 'W', from the Experimental Phase. For instance, "If greenhouse gases continue to be emitted, it will contribute to the extinction of polar bears" had a W of 0.66.
- c) The subjective probability for the data premise, 'D', also from the Experimental Phase. For example, "Transportation emits greenhouse gases" had a D of 0.75.
- d) The a posteriori subjective probability of the critical assertions, now deemed a conclusion 'C', in the experimental phase's second task, was 0.56.

⁶⁵ Subject-level regression analyses were also conducted (not included here due to space constraints). In these analyses, both the premise model and the multiplication model yielded significant changes in F.

We then calculated the combined incredibility of the premises ($\bar{W} + \bar{D}$) and subtracted the conclusion's a priori incredibility ($\overline{Prior}$):

$$\bar{W} = 1 - 0.66 = 0.34; \bar{D} = 1 - 0.75 = 0.25; \overline{Prior} = 1 - 0.38 = 0.62$$

$$\text{Thus, } \bar{W} + \bar{D} - \overline{Prior} = 0.59 - 0.62 = -0.03$$

This process was replicated across all arguments in the four believability conditions, producing a dataset to assess the premises' incredibility against the conclusions' initial incredibility, introducing the "Relative Disbelief Index" (RDI), where, in our example, RDI = -0.03. The goal is for RDI values to be equal to or less than 0, indicating enhanced credibility. Additionally, we calculated the change from a priori to a posteriori believability for each argument across all conditions, creating the "Relative Believability Index" (RBI), where RBI = 0.56 - 0.38 = 0.18 in our example.

We anticipated a negative correlation between RDI and RBI, signifying that as RDI decreases (becoming more negative), RBI should increase. This was tested through one-tailed correlation analyses. Table 13 presents these correlations along with their 90% confidence intervals and the interpreted effect size.

Table 13

Correlations between RDI and RBI together with their effect size

Bel.	<i>r</i>	90 % CI		Effect Size
		LL	UL	
HH	-.85***	.72	.92	Large
HL	-.72***	.50	.85	Large
LH	-.75***	.55	.87	Large
LL	-.75***	.55	.87	Large

Note: CI = confidence interval; Bel. = believability condition; *r* = Pearson correlation; LL = lower limit; UL = upper limit; Effect Size = *r* effect size given by Cohen (1992).

****p* < .001

The table shows strong negative correlations between RDI and RBI across all believability conditions, suggesting that when the combined incredibility of the premises is equal to or less than the initial incredibility of the claim, the conclusion's believability tends to increase. The confidence intervals' lower limits significantly surpass 0, indicating a large effect at a 90% confidence level.

Results Concerning the Criterion to Becoming Belief

This section aimed to identify the prerequisites for a premise combination that results in a conclusion's credibility exceeding 0.5, thereby persuading the audience. Following the enhanced

argumentation standard discussed previously, we hypothesized that a combined premises' incredibility below 0.5 would lead to a conclusion's credibility above this threshold.

Our analysis confirmed this hypothesis across the dataset: in every case where the premises' combined incredibility was below 0.5, the conclusion's credibility exceeded 0.5. Interestingly, there were instances where the premises' combined incredibility was above 0.5, yet the conclusion's credibility still exceeded this mark, indicating a conservative threshold for premises' incredibility. Further scrutiny suggests the actual limit for premises' incredibility to be around 0.75, beyond which no conclusion reached a credibility of over 0.5. Therefore, no conclusions with credibilities above 0.5 were observed when the combined incredibility of the premises exceeded this more lenient threshold of 0.75.

5. Discussion

This study presents a model for understanding the human belief system, offering insights into their dynamics and how beliefs are formed and adjusted. It aligns with existing research on probabilistic deductive reasoning (e.g., Evans et al., 2015; Pfeifer & Kleiter, 2009, 2010; Politzer, 2016), emphasizing the model's relevance and its psychological validity. The core objective was to validate the theoretical constructs proposed, highlighting the belief system as dynamic, influenced by environmental inputs and internal operational mechanisms. Beliefs evolve through the acceptance or rejection of information, incorporating both inductive and deductive reasoning processes. Inductive reasoning is informed by the frequency and similarity of experiences, while deductive reasoning in the belief system relies on formal connectors and content-relevant schemas to adjust belief intensity, combining degrees of belief.

Experiment 2 focused on the combination of believability degrees, exploring the hypothesis that credibility transmission resembles a multiplicative operation. Using the affirmative indefinite hypothetical syllogism of the first figure as the argumentative form, we systematically manipulated the credibility of the warrant and the data premise by modifying the middle term. This resulted in four premise credibility conditions (HH, HL, LH, LL), allowing for controlled analysis. The findings explore how premises collectively influence credibility, testing the multiplicative hypothesis and the epistemic weakest part rule. This involved examining how posterior credibility compares with a priori credibility and its boundaries, further exploring the impact of premise credibility and the efficacy of multiplication in integrating these elements cohesively.

Significant enhancements to the argumentation standard were proposed and tested, particularly regarding the minimal premise credibility combination required for persuasive argumentation. These efforts shed light on the processes behind credibility evaluation and transmission. The following sections will detail Experiment 2's findings, examining the prerequisites for premise and claim credibility, the replicability of credibility increases from Experiment 1, and a review of each hypothesis against the findings. We will then address alternative interpretations for unexpected results and link specific findings to broader psychological and argumentation theories.

The discussion concludes with methodological reflections and a summary of contributions to the overarching thesis.

5.1. Prerequisites on the Credibility of Premises and Claims

The belief system outlined in this thesis is inspired by argumentative systems. Argumentation is a communicative activity where a speaker, a listener, and a message converge. This study focused on neutralizing the speaker's role, concentrating on the listener and the message, presented via computer and devoid of a specific communicator or singular theme. Participants were expected to perceive critical claims as having low a priori credibility and premises as either high or low credibility, depending on the argument condition. This led to the formation of enthymematic (HH) and non-enthymematic arguments (HL, LH, LL) if participants assessed the premises as expected. The results largely confirmed the prerequisites, allowing for the establishment of the four argument credibility conditions.

In enthymematic arguments, an increase in credibility was expected through a multiplicative operation of the premises' credibility. For non-enthymematic arguments, a multiplicative influence was also anticipated, albeit resulting in lower credibilities and responses primarily influenced by the least believed premise, along with some influence from the a priori credibility of the claim.

5.2. Effectiveness of Argumentative Treatment

The experiment aimed to confirm two key aspects: firstly, whether enthymematic argumentation could enhance the credibility of claims beyond their a priori levels, consistent with previous findings. A significant effect was observed: claims were considered more credible following HH premises than when presented in isolation. Secondly, the study examined whether this increase was due to the high credibility of the premises rather than being presented as a conclusion from a formal argumentative structure. The results revealed that only under the HH condition did the credibility of the claims exceed the initial levels, thus underscoring the positive influence of enthymematic argumentation.

This is evidence in favour of the credibility of conclusions being intrinsically linked to the credibility of the premises on which they depend.

5.3. Regarding the Equitable Transfer Hypothesis

The proposed theory highlights two crucial components in the transmission of credibility: event representations and topical relation rules, corresponding to the data premise and the warrant, respectively. According to this theory, both components are essential for transferring credibility to the assertion. The results showed significant effects of the credibility of both premises on the credibility of the conclusion, with large and similar effect sizes.

The interaction between the data premise and the warrant suggests that clear variations in the conclusion's credibility based on one premise's credibility only manifest when the other is highly credible. This was demonstrated in comparisons such as HH versus HL or HH versus LH. A high level

of credibility in the relation rule allows the credibility of the event representation, whether high or low, to be transferred to the conclusion. Reciprocally, a highly credible event representation facilitates the observation of differences influenced by the credibility of the relation rule. If the credibility of this rule is low, it will restrict the amount of credibility that can be transferred from the event representation to the conclusion.

When one of the elements is low (L), the differences are significantly reduced, though not entirely eliminated, as reflected in comparisons such as LH versus LL or HL versus LL. A relation rule of low credibility will allow only a limited transfer of credibility from the event representation. Similarly, if the event representation is of low credibility, that will be the credibility transferred to the conclusion, regardless of the high credibility of the rule. These findings reinforce the validity of the proposed theory and the multiplication model.

5.4. Regarding the Hypothesis of Multiplication in the Combination and Transmission of Credibility

The model proposes that the credibility of premises is transmitted to conclusions via a multiplication-like operation. Recognizing the rarity of extreme values such as 0 or 1 in realistic contexts, the model anticipates that credibility will be highest in the HH condition and lowest in LL, with HL and LH offering intermediate credibility levels. The ANOVA confirmed significant treatment level differences, reinforcing the multiplication operation's role in credibility transmission. However, the anticipated equivalence between HL and LH was not observed, a discrepancy further examined in the unexpected results section.

In examining the relationship between premises and conclusion credibility, we proposed that the conclusion's credibility adjusts to that of the premises through a multiplication-like process. This hypothesis received strong empirical support, with observed conclusion credibility and the product of premises' credibility showing high correlation rates, most notably a 0.90 correlation in the HH condition. However, this correlation diminishes—though remains significant—when the credibility of one or both premises is lowered, falling to around 0.70 with one low-credibility premise and to 0.40 with two. This highlights the crucial role of premise credibility in shaping how the conclusion's credibility is perceived.

The HH condition significantly boosted claim credibility beyond a priori levels, suggesting participants matched conclusion credibility to that of the highly credible premises. Conversely, in conditions with lower premise credibility (L), a decrease in conclusion credibility was observed, indicating a weaker argument. Despite this, the presence of significant correlations with the multiplication model indicates that participants considered premise credibility, possibly due to explicit instructions or the act of presenting the premises, thus focusing attention on the articulated belief system. Yet, in such scenarios, posterior credibility may also rely on a priori credibility, suggesting a complex interplay of factors beyond simple multiplication. This complexity, particularly evident in non-HH cases, offers a tentative explanation for the observed decrease in correlations in L conditions.

5.5. Regarding the Hypothesis of the Weakest Part Epistemic Rule

The rule posits that the conclusion's credibility is primarily influenced by the less credible premise when at least one premise is fully justified. This principle, an extension of the multiplication hypothesis, aligns with the traditional weakest part syllogistic rule which is introduced in app. C (i.e., If there is a negation or an existential in the premises of the syllogism, then it will also be present in the conclusion). However, it transitions from a bivalent logical framework, focusing on formal elements like negations and quantifiers, to a probabilistic system emphasizing credibility levels. It asserts that a conclusion cannot be more credible than its least credible premise, thus the conclusion's credibility is mainly shaped by the lowest credibility among the premises. This interaction also considers how the premises' credibility compares with the claim's a priori credibility.

In deductive reasoning, logical validity ensures that true premises lead to a true conclusion, whereas false premises may result in a conclusion that is either true or false. Conversely, in rhetorical arguments grounded in beliefs, the full belief in the premises necessitates a fully believed conclusion. Yet, when premises are disbelieved, the conclusion's perceived credibility can span from disbelief to full belief, influenced by the individual's prior belief in the claim. In this type of arguments, extreme credibility values for premises—and consequently, for conclusions—are improbable. The premises' credibility typically leads to intermediate credibility levels for the conclusion, which may either exceed or fall short of the claim's a priori credibility based on the premises' relative credibility.

5.5.1. Regarding the Multiplication Exceeding Prediction: the HH case

When premises' credibility surpasses a priori levels, the model indicates that the conclusion's credibility is solely determined by the combined premises' credibility, limited to that of the least credible premise, as evidenced in the HH condition. This aligns with the result that posterior credibility in HH significantly exceeds a priori levels. Given that either premise may be the least credible depending on the specific item, the conclusion's credibility is determined by both, confirmed by significant correlations with both premises. Furthermore, the conclusion's credibility does not surpass that of the least credible premise, supporting the epistemic rule that the weakest part dictates the conclusion's credibility limit.

Regarding the Range Limits of Posterior Credibility in HH

The results show that posterior credibility, when exceeding a priori levels, is capped by the credibility of the least credible premise, serving as the upper boundary. This finding aligns with the discussion on Theophrastus' rule, also known as the Weakest Link. Additionally, previous sections posited that posterior credibility cannot fall below the product of the premises' credibility without resulting in inconsistency. The results confirm that posterior credibility indeed remains within defined limits: the credibility of the least credible premise as the upper limit and the multiplication of the premises' credibility as the lower limit.

5.5.2. Regarding the Multiplication Falling Below Prediction

Arguments featuring a low-credibility (L) premise are typically perceived as less convincing, potentially leading to the retention of a priori credibility without considering the premises. However, it was anticipated that evaluating these premises is feasible; in such cases, the conclusions' credibility would be adjusted by multiplying the premises' credibility, significantly relying on the least credible premise to determine the conclusions' final credibility. These tendencies might be reflected in the results, particularly if some participants disregard the premises while others consider them, impacting the average credibility of the conclusions and possibly raising it above that of the least credible premise.

In L conditions, both the content and credibility of the premises were considered. The reduction of conclusions' credibility below the a priori in LH and LL clearly indicates the premises' impact. Conversely, conclusions' credibility in HL being equivalent to the a priori might imply that the premises are being overlooked. However, correlations between the conclusions' credibility and the premises across all conditions, including HL, show that the premises were evaluated.

According to the epistemic rule of the weakest part, it was expected that the conclusion's credibility would mirror that of the L premise, not the H, which was confirmed in HL and LH. Unexpectedly, in LL, the conclusions' credibility correlated with the data premise, the more credible of the two. By other hand, only 16% of conclusions in L conditions were equal to or lower than the L premise, and the analysis revealed significant differences between the credibility of conclusions and the least believed premise in all L conditions. This points to a potential influence of a priori credibility, suggesting that the interaction between it and the premises introduces variability not fully anticipated by the rule of the weakest part.

Finally, it appears that, even in weak arguments, participants actively seek credible elements within the premises, possibly driven by experiment's instructions or by a natural tendency to adhere to the pragmatic principles of relevance, cooperation, and charity in communication.

Regarding the Range Limits of Posterior Credibility in L Conditions

In L conditions, the premises' multiplication sets a posteriori credibility's lower limit, while a priori credibility establishes the new upper limit. To maintain consistency, the posterior credibility of the conclusion cannot fall below the product of the premises' credibilities through multiplication. Conversely, when this combination falls below the a priori credibility, the argument does not provide reasons that justify either decreasing or increasing the conclusion's credibility relative to its a priori level. Hence, conclusions' credibility in L conditions is expected to fall between these defined limits—multiplication as the lower and a priori credibility as the upper. The results confirm this, underlining that conclusions' credibility, when stemming from less credible premises than a priori, adheres to these boundaries, suggesting a priori credibility may influence but not exclusively determine posterior credibility.

5.6. Regarding the Influence of Prior Credibility and Premise Multiplication on the Variability of Posterior Credibility

In the HH condition, an increase in the credibility of the conclusions was expected to arise solely from the credibility of the premises, thereby minimizing the role of a priori credibility in the observed variability. It would be illogical to assume that a priori credibility could cause its own increase. Conversely, in the L conditions (HL, LH, LL), posterior credibility could not only equal or be lower than a priori credibility but might also potentially exceed the least credible premise's credibility. This is because weak arguments might lead some participants to retain their initial credibility, disregarding the premises and thus raising the average credibility above what the premises suggest, assigning a significant role to a priori credibility in these conditions' variability. The distribution of posterior credibility is determined by a priori credibility, each premise's credibility, and their multiplication as an integrated combination. The latter acts as a robust predictor that adjusts the conclusions' credibility into a coherent unit. The multiplication process consistently impacts posterior credibility, even if one of the premises contributes minimally. This mechanism ensures that its predictive and explanatory power is as effective as the integration of both premises, capturing both significant and minor contributions. This approach highlights the importance of considering both multiplication and the influence of a priori credibility to fully understand the variability in posterior credibility under different conditions.

Multiple regression analyses were conducted with two key objectives. First, we explored the potential influence of prior credibility on the posterior credibility of conclusions at different levels of argument credibility. Second, we sought to confirm that the multiplication of premise credibility indeed integrates the effects of credibility from each premise into a consistent unit. Furthermore, these analyses allowed for an examination of the relationships between the effects of argument credibility and prior credibility, as well as the identification of the strongest and most robust predictors in explaining the observed variance in conclusion credibility. As explained earlier, to mitigate multicollinearity between the premises and their product, separate regression analyses were conducted. This facilitated the assessment of a model incorporating a priori credibility alongside the premises, as well as another model involving a priori credibility and the multiplication of premise credibilities.

5.6.1. Regarding Prior Credibility

The results clearly demonstrate the impact of prior credibility on the posterior credibility of conclusions, but they diverged from our predictions. We expected no effect in the HH condition and positive effects in the L conditions (HL, LH, LL). However, regression analyses showed that prior credibility significantly contributed to variance in HH, HL, and LH, but not in LL, indicating its impact in scenarios with at least one high-credibility premise. Conversely, in the multiplication model, prior credibility significantly influenced variance only in HH and HL, not in LH or LL, suggesting its effect in this model is limited to conditions with a high-credibility warrant.

It's crucial to note that prior credibility only accounted for a small portion of the variability in HH across both models. Although it didn't influence the increase in credibility, it could affect the distribution of posterior credibility by limiting its increase. For instance, when considering two statements with equally credible premises (0.8), the presence of a prior credibility of 0.4 in one and 0.2 in the other could facilitate the former reaching a credibility of 0.64 more easily than the latter.

Prior credibility significantly impacted posterior credibility in HL in both models, and in LH, but only within the premises model. The absence of its influence in the multiplication model in LH suggests that the contribution of prior credibility might be redundant and already incorporated within the multiplication calculation. While prior credibility seemed to act as an attractive factor in HH and L conditions, its lack of effect in LL and in LH's multiplication model weakens this assumption. These findings could indicate that prior credibility does not necessarily justify changes in correlation in L conditions nor an increase in the conclusions' credibility above that of the least credible premise.

5.6.2. Regarding the Validation of Multiplication as an Integrating Unit of Premise Credibility

The proportion of variance explained by the premise and multiplication models will be compared across conditions, specifically focusing on the impact of each model element.

In the HH condition, both models equally accounted for 83% of the variance. Within the premise model, each premise equally contributed to the explained variance, with very similar β weights. Conversely, the multiplication achieved a higher beta weight than each premise individually, indicating that the multiplication model effectively integrates the explanatory potential of both premises into a single explanatory unit.

In the HL and LH conditions, the premise and multiplication models explained similar proportions of variance, slightly higher in the multiplication models. In the premise models, only the least credible premise significantly contributed to explaining the variance observed in the conclusions, being the warrant in one case and the data premise in the other. The multiplication explained a significant proportion of variance in both models, showcasing its robustness. This pattern reinforces the weakest part rule and demonstrates the coherent integration of premises' credibility by the multiplication model, aligning the results from both conditions consistently. In LL, both models explained a minimal and nonsignificant portion of the variance, indicating their equivalent but limited effectiveness in this condition.

The multiplication model consistently integrates contributions from both premises into a coherent unit, even when one premise's contribution is minimal, demonstrating equivalent explanatory power across both models.

5.6.3. Strength and Robustness of Different Predictors and Their Relationships

The multiplication model exhibits similar explanatory and predictive strength to the premise-based model, yet it slightly surpasses it in explaining variance in conclusions, though the difference is very small. It emerges as a more stable and robust predictor than individual premises, significantly

explaining variance in three out of four conditions, unlike each premise, which does so in only two conditions.

Within the premise model, both premises contribute equally to variance explanation. Including prior credibility factor in the model shifts some focus towards the warrant premise, especially notable in the LH condition with a beta weight of .72, surpassing the data premise's influence in HL. However, excluding prior credibility from the model shows balanced contributions from both premises in all but the LL condition.

Prior credibility, overall, has less explanatory potential. In the HH condition, it contributes less than premises and multiplication. Its inclusion marginally (though significantly) increases explanatory power, more so in HL where its impact is more pronounced, suggesting a stronger role when the argument's believability is low. Still, it's below the impact of the L premise. In the LH condition, the influence of prior credibility slightly diminishes in the premise model, though it remains substantial, and it disappears entirely in the multiplication model. Finally, in LL prior credibility has no impact in any model.

Thus, prior credibility plays a nuanced role, being more influential when premises have differing credibilities. This suggests that in less believable arguments, prior credibility's influence on conclusion credibility becomes more apparent, though the LL condition's results challenge this straightforward interpretation.

5.7. Regarding the Enhanced Argumentation Standard Hypothesis

This experiment aimed at refining argumentation standards for increased effectiveness. Prior work (chap. 2, pt. I; chap. 6, pt. II) highlighted that subjective probabilities are not absolute; they're influenced by prior knowledge and how it contrasts with new information, generating a prior credibility for that new information. The lower this initial credibility, the greater the potential for its increase via argumentation, which requires presenting conclusions derived from premises credible to the individual.

For an argument to bolster a conclusion's credibility, the premises must initially be seen as more credible. However, as demonstrated in Experiment 1, this condition, while necessary, is not always sufficient due to the diminishing nature of credibility, especially when combined through multiplication. Then, a key question emerges: How much more credible must the premises be relative to the claim's a priori credibility to enhance it a posteriori? This inquiry leads to the concept of maximum allowable incredibility for premises, posited to be the level of the claim's a priori incredibility. If premises' combined incredibility surpasses this, they fail to augment the claim's credibility⁶⁶.

Two indices were introduced to quantify this relationship: the Relative Disbelief Index (RDI) and the Relative Believability Index (RBI), relating to the a priori incredibility and the potential increase

⁶⁶ In this context, 'combined' refers to the sum of incredibilities, not their multiplication.

in credibility, respectively. It was hypothesized that a lower RDI should correlate with a higher RBI, with expected negative correlations between these indices across different conditions. The observed significant negative correlations, with large effect sizes, support the hypothesized mechanisms for credibility enhancement via argumentation.

This evidence supports the refined argumentation standard based on the premise's credibility relative to the claim's initial incredibility, underscoring the multiplication hypothesis's role in argument effectiveness.

5.8. Regarding the Criterion to Becoming Belief

The study's final goal was to determine the conditions under which combined premises lead to a conclusion with a credibility exceeding 0.5, effectively convincing the audience. This issue mirrors the structure of the previous problem, applying a similar solution: if the combined incredibility of premises is below 0.5, their credibility product in the conclusion is likely above 0.5. This pattern is supported by findings that, when the conclusion's credibility falls below 0.5, the premises' combined incredibility exceeds this threshold.

However, exceeding a 0.5 incredibility for premises doesn't always prevent a conclusion from being credible. Analysis revealed that no conclusions reached a credibility above 0.5 when their combined incredibility surpassed 0.75—a potential maximum of 2. This suggests that while a combined incredibility below 0.5 is adequate for a credible conclusion, it's not strictly necessary; premises with combined incredibility up to 0.75 can still yield a conclusion credibility above 0.5. Therefore, while maintaining combined incredibility below 0.75 is ideal, 0.5 serves as a reliable threshold for achieving desired credibility levels.

5.9. Unexpected Results and Alternative Interpretations

Differences between HL and LH Conditions: The first unexpected finding was the observed difference in averages between the HL and LH conditions, specifically that $HL > LH$. This difference could be attributed to inherent differences in the premises' credibility. Post-hoc analyses (refer to App. B-III) identified differences in premises' subjective probabilities: higher for premise H in LH and higher for premise L in HL, though noted only at the subject level. Moreover, only the differences in subjective probabilities of premise L seem to go in the right direction to explain the observed results. This unexpected outcome could be due to variations in the premises' credibility, but it's not clear. Other explanations, like inherent differences in the premises' nature, should not be dismissed. Further research is needed to clarify that question.

The LL case: The second unexpected result concerns the LL condition, where, despite significant, the correlation between observed and predicted credibility scores was the lowest. The regression analysis suggests that this reduced correlation is not attributable to the effect of a priori believability.

The belief model posits that low credibility in one premise restricts the variability of credibility transmitted to the conclusion, resulting in decreased conclusion variability. Conversely, high

credibility in one premise attributes variations in conclusion credibility to the other premise. Specifically, a premise with high credibility (close to 1) means conclusion credibility will closely follow the variability of the other premise. For example, if we have arguments with one high-value premise and the other varying between 0.3 and 0.9, the resulting conclusion's credibility closely mirrors these variations. Conversely, a premise with low credibility (close to 0) leads to reduced or even negligible conclusion variability.

Consistent with this, the variability of conclusion credibility is highest in the HH condition and lowest in LL, due to differences in premise credibility. This difference in variability affects correlational analyses capacity to detect systematic relationships between premises and conclusions, with HH offering the most conducive environment for such analysis and LL the least. The standard deviation and range of variation in the HH condition were the highest (.14 and .63, respectively), indicating greater data variability, whereas LL had the lowest (.06 and .24, respectively), as also reflected in predicted scores.

Table 14 displays variability measures for observed conclusions and scores predicted by the multiplication model, including standard deviations and range. Higher values indicate greater variability.

Table 14
Variability Measures for Observed and Predicted Credibility of Conclusions

Stand. Deviation	HH	HL	LH	LL
Conclusions	0,1449	0,1334	0,0952	0,0694
Multiplication	0,1262	0,1156	0,1079	0,0349
Range	HH	HL	LH	LL
Conclusions	0,63	0,59	0,37	0,24
Multiplication	0,51	0,39	0,41	0,13

Note: Standard deviations and range of variation for conclusions and multiplication are shown here. It is possible to observe how, whenever the average credibility decreases, the variability also decreases.

Furthermore, a subject-level analysis in the LL condition revealed significant correlations with both premises, suggesting that low correlations may be due to the restricted data variability inherent to low premise credibility. This limited variability impacts the coefficient of determination, which is the square of the correlation, thereby reducing it.

5.10. Comparison with Previous Studies in the New Paradigm of Deductive Reasoning

Developed independently from ongoing debates in psychology regarding reasoning, discourse, or argumentation, this research intersects with key topics in the field. Unlike research focused on belief biases, Bayes' theorem's relevance, fallacies, information validation, argumentative skill enhancement, or argument quality assessment, our research doesn't directly address these areas. It also diverges from psychological investigations into pathological beliefs, conspiracy theories, or fake news.

However, our findings align with the broader psychology of reasoning literature, particularly the shift from bivalent logic to viewing subjective probability as a measure of belief. This paradigm shift recognizes individuals' tendency to evaluate knowledge via degrees of belief, implicitly assigning these to complex mental representations and everyday discourse content of sentences. Our contribution lies in detailing how the credibility of premises impacts the perceived credibility of conclusions in a probabilistic deductive process, thereby enriching discussions on rational confidence levels in derived conclusions and situating our work within a contemporary understanding of reasoning processes.

On Credibility Intervals and Limits in Probabilistic Reasoning

In the domain of probabilistic reasoning, research primarily explores belief computation through Bayes' theorem and its conditional updating. Studies such as Heit & Rotello (2010) and Oaksford & Chater (2007, 2009) examine inductive reasoning, while others analyze deductive reasoning's probabilistic consistency, based on De Finetti's theories, for subjective probability assignments in uncertain deductive reasoning. While these studies are considered Bayesian in computing subjective probabilities, they do not centrally employ Bayes' theorem in constructing their reasoning models.

Research in deductive models emphasizes the coherent transfer of credibility from premises to conclusion. This transfer occurs through a consequence relationship rather than through an uncertain relationship between premises and conclusion. This relationship ensures p -validity and consistency (referred to as coherence in this context). Noteworthy studies in this field include those of Pfeifer & Kleiter (2006, 2009, 2011), Politzer (2016), Baratgin & Politzer (2016), Evans & Over (2013), and Evans et al. (2015). These studies mainly explore conditional reasoning and inferential schemes like Modus Ponens and Modus Tollens; essentially, two-term arguments with a conditional in the major premise.

Building on De Finetti (1937), these studies propose that the evaluation of deductive inferences is conducted in terms of the consistency of probability preservation. Based on the type of conditional inference and the probabilities assigned to the premises, an interval is computed to indicate the expected probability range for the conclusion. This interval sets the bounds for the degrees of belief that a rational individual should maintain to avoid inconsistencies. In the case of Modus Ponens, they propose multiplying the probabilities of the premises for the lower limit and then adding the complement of the smaller categorical premise for the upper limit:

$$P(\text{Conclusion}) \in [\Pi P(\text{premises}); (\Pi P(\text{premises}) + (1 - P(\text{minor premise}))]$$

The proposed lower limit in this interval aligns with our multiplication model for first-figure arguments, showing conceptual parallels despite different origins. Modus Ponens, akin to the first-figure syllogism, represents the most natural inferential rule. To some extent, Modus Ponens can be considered a special case of the first figure, where the first premise is a conditional proposition, and

the second premise is the affirmation of the truth of the conditional's antecedent. I would like to recall both structures here⁶⁷:

Modus Ponens	Syllogism
$p \rightarrow q$	$M \rightarrow P$
p	$S \text{ 'is' } M$
q	$S \text{ 'is' } P$

However, it's important to note that they are not equivalent in that syllogisms incorporate three terms, while Modus Ponens use two, affecting dynamics significantly. For instance, the probability of the minor premise (or data premise) " $P (S \text{ is } M)$ " in a syllogism does not equate to that of " $P(p)$ " in Modus Ponens due to their structural disparities: " p " of Modus Ponens just could be assimilated to " M " in the syllogism, but not to " $S \text{ is } M$ ". The probability given to "there will be an increase in greenhouse gases" will not be identical to that of "transportation will produce an increase in greenhouse gases" or, let's say, "dogs will produce an increase...". These sentences involve a subject S from which M is predicated, and the subjective probability a person attributes to the statement " $S \text{ is } M$ " will depend on how predicative M seems of S .

Previous studies establish intervals for the expected a posteriori probability of a conclusion. In Modus Ponens, the lower limit is calculated by multiplying the probabilities of the premises, while the upper limit includes the non-believed part of the minor premise (the subjective probability that it is false) ⁶⁸ These limits formally follow from the *total probability fundamental theorem*. If the a posteriori probability falls below this lower limit, the argument is p -invalid; exceeding the upper limit indicates an inconsistency known as 'conjunction fallacy'. These limits, derived from the theorem, do not depend on a priori credibility. Therefore, the a priori credibility could exceed the calculated upper limit of the interval. This would necessitate a reduction in credibility a posteriori to be consistent with that of the premises. However, this outcome is not assured: if the a priori credibility is at its maximum, it is not required to decrease to conform to a flawed argument.

Despite of the similarities, the studies discussed and the current thesis operate within distinct frameworks, leading to different limitations and outcomes. In the model depicted in this thesis, the ideal goal is to provide exact scores directly derived from multiplication. However, the integration of the weakest link epistemic rule, validity, and its interaction with a priori credibility necessitates the acceptance of relatively narrow intervals. While the lower limit mirrors other studies by multiplying both premises' probabilities, the upper limit adjusts based on this product's comparison to the a priori probability of the conclusion. Typically, in an argument (HH scenario), this product exceeds the a priori probability, eliminating the need for an interval. Yet, to accommodate potential

⁶⁷ The syllogism presented here corresponds to that employed in the experiments of this thesis: the structure of the hypothetical syllogism of the first figure without quantifiers or negations.

⁶⁸ Refer to Politzer (2016) for an intuitive interpretation of these intervals based on an analogy involving a four-part water tank. In this analogy, water represents the credibility or probability of the conclusion, and different sections of the tank represent the probabilities of the premises and their complements. An innovative geometric approach to calculating the interval limits is elaborated, conceptualizing them as the areas of a rectangle.

errors, the posteriori probability should not exceed the upper limit set by the least credible premise's probability, as dictated by multiplication's nature. Notably, this resulting interval is exceptionally narrow.

When the probability resulting from multiplication falls below the a priori probability, the model sets the a priori credibility as the upper limit of the interval. This reflects the principle that credibility is transferred from premises to conclusion. If the conclusion initially seems more probable than its supporting premises, the lower credibility of these premises should not markedly influence the conclusion's credibility, indicating the premises fail to strengthen the initial belief. Consider a scenario where I am 0.7 certain (or have a low acceptance of up to 0.3) that 'the Pyrenees will have no glaciers in 30 years.' If someone presents an argument intending to boost my belief in this prediction but uses premises, I find less credible, resulting in a combined credibility of just 0.6 (or 0.2), my subsequent belief should not exceed my original 0.7 certainty, nor be lower than 0.6 (or 0.2). This low-quality argument, due to its less convincing premises, fails to challenge my initial reasons and thus does not lower my initial belief of 0.7 certainty (or 0.3). Hence, my belief in the claim fluctuates between the product of the premises' credibility and my initial belief, depending on how much weight I give to the provided premises—whether I consider or disregard them. This phenomenon was supported by the findings. Moreover, the limits defined in this thesis lean more towards an empirical than a strictly formal perspective, resulting in notably tight ranges.

On the Theophrastus' Rule, Consequence Condition, Weakest Link, and MIN Rule in Argumentation and Plausibilistic Reasoning

The Rule

The consideration of plausible limits leads to exploring the so called "Theophrastus' rule" (among other denominations) in the realm of argumentative validity assessment. This rule, often misinterpreted, has its roots in the *sylogistic rule of the weakest part*, presented in app. C. In this thesis, it has also been linked to the epistemic rule of the weakest part.

Theophrastus, as per recent research (Ierodiakonou, 2020; Rescher, 2018; Vega, 1998), never formulated what is now known as the "Theophrastus' rule" in the context of informal logic but rather the *sylogistic rule of the weakest part* (also known as the *de peiorem sequitur conclusion partem rule*) in assertoric and modal logic. Its later transformation into the realm of plausibilistic reasoning and argumentation is attributed to Rescher (1976), who proposed this rule as a consequence condition for plausible reasoning. The rule or condition states:

"...when a set of mutually consistent propositions $A_1, \dots, A_n$, implies some other proposition B by valid deductive argument, then the plausibility of B cannot be less than the plausibility value of the least plausible proposition among the set $A_1, \dots, A_n$." Rescher (1976, p. 15)

In other words, the plausibility of the conclusion must be, at a minimum, as plausible as the least plausible premise (it can be more plausible than this⁶⁹). According to Walton (1992), this rule can be summarized as follows:

$$\text{If } A_1, \dots, A_n \text{ implies } B, \text{ then } \text{plaus}(B) \geq \text{MIN } \text{plaus}(A_1, \dots, A_n)$$

This rule operates within a typical argumentative context where an arguer aims to convince the listener of a seemingly implausible idea by framing it as the conclusion supported by plausible premises. However, the plausibility of these premises can vary in degree. When one premise is perceived as highly plausible and another as less plausible, the overall plausibility of the argument is strongly influenced by the least plausible premise (as mirroring the Theophrastus' *de peiorem rule* applied to modal logic). In this conventional argumentative context, the least plausible premise will still be considered plausible, meaning its content will be perceived as more true than false.

From this perspective, deductive reasoning about plausible premises resembles deductive reasoning about true premises. If an argument is deductively valid and both premises are considered plausible, then the conclusion must necessarily be considered plausible. At least at the same level of plausibility as the least plausible premise. This is perceived as a condition of consequence. From this viewpoint, it should not happen that from justified premises an unjustified conclusion is derived in valid deductive reasoning. Therefore, it is postulated that the argumentative strength of a deductive argument is, at a minimum, the lesser of the degrees of justification of its premises. For this reason, this rule has also been termed the *least* plausible premise rule (Rescher, 1976), the *weakest* link rule (Pollock, 1995, 1998, 2001), and the *MIN* rule (Walton, 1992).

Additionally, Walton also highlights a rule that mirrors the multiplication rule discussed in this thesis. It suggests multiplying the probabilities of a conditional premise and a factual premise taken from a knowledge base. This rule sets the conclusion's maximum probability to the value of the least probable premise, establishing an upper limit. Walton (1992, p. 34) names this the *product rule* and deems it compatible with plausible reasoning's core principles, though he does not elaborate on its coherence or reference it in subsequent works.

Pollock (1991, 2001) dismisses probabilistic approaches as overly complex for human cognition, challenging the notion that credibility or plausibility diminishes with each inferential step. He argues that deductive reasoning from justified premises should not yield an unjustified conclusion, criticizing probabilistic models for suggesting a weakening of conclusion strength despite strong initial reasons. Yet, Pollock doesn't clarify the threshold of plausibility for a conclusion to be deemed sufficiently strong, leaving his portrayal of these concepts vague.

⁶⁹ They assert that the conclusion should be, at a minimum, as plausible as the least plausible premise. The "at a minimum" suggests the possibility of it being more plausible than the least plausible premise. However, they do not specify the basis on which it could be more plausible. Certainly, it is not determined by a priori credibility, as this rule operates in a standard argumentative context where the a priori credibility of the conclusion is minimal. Consequently, these authors establish a lower limit but do not define an upper limit.

Critiques to the Theophrastus Rule from Cognitive Psychology

The Theophrastus Rule, under its various names, has faced criticism, especially from the perspective of cognitive psychology. Ulrike Hahn has questioned this rule in several works (e.g. Hahn & Oaksford, 2006b; Hahn et al. 2013), showing, through examples, that it can lead to counterintuitive and paradoxical results. Given the technical complexity of her arguments, I will briefly summarize them, without the pretence of addressing them in detail.

In an example presented by Hahn & Oaksford (2006b), two circular arguments are exposed:

- 1) God exists because the Bible says so, and the Bible is the word of God.
- 2) Electrons exist because we can see 3 cm traces in a cloud chamber, and the 3 cm traces in cloud chambers are the signature of electrons.

The authors argue that the second argument feels more intuitively convincing, highlighting that science often employs such reasoning. Despite their structural similarity—both presuppose their conclusions in the premises, offering no new rationale for belief—they elicit different intuitive evaluations. According to the Theophrastus Rule, there should be no apparent difference between the two arguments, and the plausibility of the conclusion should not increase after conditioning on the premises since, in these arguments, it will always be as low as the lowest plausibility of the premises. Yet, the second argument is seen as more plausible, interpreted as an inference to the best explanation rather than a fallacy, a distinction the Theophrastus Rule doesn't capture but which fits within a Bayesian framework, according to Hahn & Oaksford.

Hahn et al. (2013) challenge the application of the Theophrastus Rule to conjunctive arguments, which states that the plausibility of a conjunctive conclusion 'A & B' (derived from the conjunction introduction inference rule) should at least match the plausibility of its less plausible component. Yet, simple probabilistic scenarios, like the likelihood of a coin landing on heads on the first toss, on the second toss, and on both tosses, reveal the rule's limitations. In this scenario, the conjunctive conclusion is clearly less plausible than the plausibility of each set independently. Contradiction between events isn't required for a conjunctive conclusion to be less plausible than its individual components. Hahn and colleagues propose that the credibility of the least plausible or probable premise should serve as the maximum credibility threshold for conjunctive conclusions.

Moreover, from the perspective of plausible reasoning, it is argued that the message content and its communicative source should be regarded as integral components of a unified interconnected argument. In this plausibilistic framework, the conclusion inherits plausibility from the less plausible component. Therefore, variations in content strength become irrelevant if the source is perceived as less reliable, and vice versa. The conclusion reflects the degree of plausibility of the weakest part of the argument under consideration.

However, studies by Hahn et al. (2009) experimentally demonstrated that these predictions, derived from Theophrastus' rule, do not hold in the context of testimony. The experiment revealed that when the source was highly reliable, the less plausible assertion could be more plausible than

the most plausible assertion made by a less reliable source. According to Theophrastus' rule or MIN rule, equally implausible assertions should lead to equally implausible conclusions, regardless of the source's reliability. These arguments together with others presented in Hahn et al.'s (2013) article indicate that the Theophrastus Rule does not constitute an adequate approach to the natural evaluation of arguments.

Critiques to the Theophrastus Rule from the Perspective of this Thesis

Our critique of the Theophrastus Rule, while not explicitly stated, has been woven through our discussion. We revisit these points to focus on the issue at hand.

The controversy over the Theophrastus Rule doesn't primarily lie in valuing the less plausible premise, akin to the *de peiorem* rule, but in its assumption as the conclusion's lower plausibility limit. This stems from a specific interpretation of logical validity's truth preservation.

- *Plausibility and Probability in Belief and Argument*: Rescher (1976) posits Aristotle's works as the direct antecedents for his plausible reasoning theory. Although Aristotle closely ties the concepts of probability and plausibility, they aren't identical. Translation variations of the Greek terms have occurred. According to the references employed here, Aristotle mentions in *Rhetoric* and *Posterior Analytics* that necessary conclusions derive from necessary premises, and probable conclusions derive from probable premises. He adds that probable propositions are considered plausible when they align with general opinion. This interpretation might have led to the notion that plausible premises yield a necessarily plausible, i.e., credible, conclusion, preventing an implausible conclusion from plausible premises. To ensure this, it's established that the conclusion's plausibility lower limit is determined by the less plausible premise.

We clarify the difference between 'probable' and 'plausible' by viewing probability as a subjective certainty scale, ranging from absolute disbelief (impossibility) to complete belief (necessity). Within belief systems, probability indicates the level of an idea's subjective justification or acceptability. Hence, premises might be justified to some extent without reaching full certainty. When we combine the probabilities of two premises, the result may have some probability but still not meet the belief system's threshold for acceptance. For instance, in a scale where 0 denotes disbelief, 1 denotes belief, and 0.5 denotes uncertainty, combining two accepted premises each justified to 0.6 may yield a degree of justification for a conclusion insufficient for acceptance. Our view on validity suggests that justification degrees transfer differently than what plausible reasoning advocates suggest.

- *Incorporating Probabilistic Validity*: Authors like Adams (1998) and Evans & Over (2013) propose a form of deductive reasoning that integrates probabilistic validity. This approach ensures subjective probabilities are transmitted from premises to conclusion while specifying how this transmission occurs. According to Evans et al. (2015), an inference with multiple premises is considered *p*-valid when the conclusion's uncertainty does not exceed the sum of the uncertainties of its premises. The total incredibility of the premises sets the maximum incredibility for the

conclusion. If the combined incredibility reaches or surpasses 1, the conclusion's maximum incredibility is also total, setting its minimum credibility at 0. Consequently, such arguments are always valid. Therefore, p -validity becomes argumentatively relevant only when the incredibilities of the premises sum to less than 1.

As observed, the Theophrastus Rule, advocated by the plausibilistic perspective, establishes the least believed or plausible premise as the lower limit of the credibility of the conclusion. In contrast, the subjective probability approach of deductive reasoning suggests a lower credibility limit defined by $\sum P(\text{premises}) - 1$. Typically, this probabilistic lower limit falls below the threshold set by the Theophrastus Rule. While the latter relies on an intuitive understanding of logical validity, the probabilistic deductive approach, though less immediately intuitive, offers a more nuanced reflection of justification's gradations, grounded in the principles of mathematical probability.

The probabilistic perspective enables the construction of arguments with premises that, individually, have a sufficient degree of justification to become part of the belief system. However, taken together, they may lead to conclusions lacking such justification. For instance, two premises with a probability of 0.7 each result in a maximum incredibility of 0.6 for the conclusion, equivalent to a minimum credibility of 0.4. While a probability of 0.7 reflects a sufficient degree of justification to enter the belief system, 0.4 does not. Nevertheless, the inference remains valid within the framework of deductive but probabilistic validity, which adheres to the axioms of probability.

- *Consistency and the Multiplication as a Mechanism for Credibility Transmission*: Evans & Over (2013) introduce the notion of probabilistic consistency to the criteria for probabilistic validity, suggesting that believing in two statements with certain probabilities limits the permissible belief in a third statement logically connected to the first two. This corresponds with the concept introduced earlier that while premises may represent independent events, the conclusion's content is dependent on both, necessitating mutual consistency among the premises' subjective probabilities. For reasoning to be valid and consistent, the conclusion's subjective probability must align with that of both premises, ensuring a coherent combination of probabilities in the conclusion.

This thesis suggests a consistent approach to combining probabilities from premises to conclusion via multiplication, aligning with the principles of p -validity. It posits that multiplication, reflecting the joint effect of premises, adheres to p -validity's boundaries. Furthermore, considering the crucial role of both premises in establishing the conclusion's credibility, if the probability of one premise is 0, its effect on the conclusion is nullified. Conversely, a probability of 1 for one premise permits the complete transfer of the other premise's probability to the conclusion, provided that the latter's probability is greater than 0.

- *Reevaluating the Theophrastus Rule in Light of Our Model*: When the revised argumentation standard is met (particularly in the HH context), multiplication's result will not exceed the lower of the two premises' probabilities. Compliance with the standard ensures that the premises' joint incredibility is less than the conclusion's a priori incredibility. Here, multiplication

sets the *lower credibility limit* of the conclusion, with the least credible premise defining the *upper limit*, creating a credibility range for the conclusion different from Theophrastus Rule's prescription.

When the standard isn't met, resulting in a probability below the a priori level, the premises' credibility may still be considered. In such instances, multiplication outlines the method for combining and transferring premises' credibilities to the conclusion, defining its credibility's *lower limit* again. The credibility of the least believed premise significantly influences the distribution of conclusion credibilities, but not as a limit. Instead, a priori credibility serves as the *upper limit*. This diverges from Theophrastus' Rule in two key ways: the least credible premise no longer sets a limit, and Theophrastus' Rule lacks specifics on establishing an upper limit.

In our model, the multiplication of subjective probabilities sets the lower limit for the conclusion's credibility, directly influenced by the least believed premise. Contrary to Theophrastus' Rule, this premise defines the upper limit when argumentation standards are met, and it guides credibility transmission without imposing limits when standards are not met. Here, the multiplication's outcome is the lower limit, with a priori credibility defining the upper limit.

Moreover, this approach demonstrates that premises, even when justifiably credible, can lead to less credible conclusions without compromising validity. This probabilistic model aligns more closely with empirical data than Theophrastus' Rule, accommodating the accumulation of incredibility across premises of the same argument. Pollock's (2001) concerns about validity are addressed by embracing this probabilistic framework, marking a clear departure from Theophrastus' Rule.

5.11. Strengths and Methodological Innovations

About the Syllogism Use

In the present experiment, we focused on classical three-term affirmative, undefined hypothetical syllogisms, echoing the approach of our first experiment. This decision to limit ourselves to affirmative syllogisms stemmed from a desire to sidestep the complexity introduced by negations, typically more relevant in refutation contexts. Our syllogisms omitted quantifiers to mirror the implicit quantification common in everyday language, where subjective probabilities play a crucial role.

Additionally, our arguments are categorized as hypothetical due to the presence of a conditional in the major premise, reflecting the structure of topical relationships. Unique to our methodology is the use of three terms, including a 'middle term' present in both premises but not in the conclusion—an approach seldom seen in experimental studies on *argumentation* psychology or credibility dynamics. This design allows for the middle term's systematic manipulation across four credibility conditions (HH, HL, LH, LL), maintaining the same conclusion sentence. This innovation facilitates comparisons of credibility attributed to the conclusion based on varying premise credibility. It also prevents the explicit linkage of the conclusion's subject and predicate in any proposition or in the participant's mind until reaching the conclusion.

This methodology captures the complexity of everyday reasoning while retaining simplicity for analytical clarity. By choosing affirmative, indefinite, and hypothetical three-termed syllogisms, we examine the essence of inference-making in daily discourse, laying a solid foundation for understanding the dynamics of credibility transmission.

About the tasks design

Dividing the experiment into two separate tasks facilitates data analysis and interpretation. In the initial task, participants individually assess the subjective probabilities of the premises, separated into warrant and data premises across two blocks, randomized to mitigate biases such as the forewarning effect (McGuire & Papageorgis, 1962; Petty & Cacioppo, 1979) and my-side bias (Baron, 1995; Wolfe & Britt, 2008; Wolfe, Britt & Butler, 2009) among others. This approach helps prevent the anticipation of conclusions and ensures unbiased premise evaluation. Additionally, participants completed a distractor task for several minutes before starting the second task.

The second task involves presenting complete arguments and evaluating their conclusions. As conclusions depend on premises, the goal is to obtain the conditional credibility of conclusions given the credibility of their preceding premises. This experiment explores everyday qualitative reasoning with credibility magnitudes, steering clear of numerical or mathematical reasoning. Presenting the conclusion evaluation task separately proves advantageous, making it challenging for participants to precisely recall numerical ratings given to premises in the previous task.

In this second task, participants read both premises, interpret them qualitatively, and implicitly evaluate them quantitatively (though not numerically). In evaluating the conclusion, participants implicitly quantify their integrated perception of the premises' credibility, assigning it numerically to the conclusion without resorting to explicit mathematical reasoning. This method captures the essence of everyday qualitative reasoning with magnitudes of credibility and sidesteps potential biases introduced by explicit numerical calculations.

5.12. Results Summary

This study evaluated subjective probabilities for each individual premise and each conclusion derived from arguments formed by these premises. Arguments were categorized into four conditions based on the credibility of the premises (HH, HL, LH, and LL). By altering only the middle term, we maintained linguistic consistency across conditions while varying premise credibility. The analysis involved ANOVA, t-tests, correlations, and regressions, summarizing the key findings:

1. Critical assertions' average a priori credibility was below 0.5. High-credibility and low-credibility premises were rated as such, aligning with pre-classification.
2. ANOVA showed that premise credibility significantly affects conclusion credibility, underscoring their relationship.
3. In HH, a posteriori credibility surpassed a priori credibility, indicating a credibility boost from argumentative treatment, echoing Experiment 1 results.

4. Analyses confirmed both premises' equal importance in credibility transmission. Credibility patterns (HH vs. HL/LH and LL vs. LH/HL) illustrate this. High credibility in one premise leads to noticeable differences in conclusion credibility, which diminish significantly when the other premise is of low credibility.
5. Correlations between observed and predicted credibility, as determined by multiplication, were notably high and significant, achieving a strong correlation of 0.90 in the HH condition. This indicates that in standard argumentative contexts (where arguments are tailored to align with the audience's beliefs), the credibility of the premises nearly perfectly predicts the credibility of the conclusion. Such findings bolster the hypothesis that credibility is combined through multiplicative computation.
6. The experiment upheld the epistemic rule of the weakest link, especially in HH, where conclusion credibility never exceeded the least credible premise. In HL and LH, correlations with L premises affirmed the rule's applicability, though LL showed deviations.
7. The predicted credibility ranges for HH (between multiplication and the least credible premise) and for HL, LH, and LL (between multiplication and a priori credibility) was confirmed.
8. Contrary to expectations, a priori credibility impacted HH but not LL, influencing credibility variance. Its effect varied across conditions, prompting further questions about its role.
9. LL's reduced variability suggests a reason for lower correlations between observed and predicted credibility, affecting variance explained in regression.
10. Both multiplication and premise models explained variance similarly, emphasizing multiplication's role in integrating premise credibility and its transmission to conclusions.
11. Establishing an argumentation standard, we confirmed that premise incredibility should not surpass the conclusion's a priori incredibility for multiplication to enhance credibility. Correlations affirmed that lower joint premise incredibility relative to conclusion's a priori credibility leads to a higher posterior credibility.
12. The results validated the enthymematic credibility rule across all cases, indicating rule compliance is sufficient but not always necessary for belief acceptance. Further analysis is needed for precise prediction.

These findings reinforce the study's hypotheses, offering strong empirical backing for the discussed belief system and credibility transmission theory, and demonstrating participants' ability to adhere to probabilistic consistency in inferential processes.

DISCUSSION

1. Summary of the Thesis Goals and Objectives

In this thesis, the development of a complex theoretical model on beliefs, credibility, and the belief system from a cognitive perspective was proposed. During its development, the field of study was delineated and the research terrain identified, aiming to ensure the model was comprehensive in terms of definitions, components, relationships, and the processes of formation and dynamics of beliefs. Through this approach, the aim was to illuminate the origin of beliefs and their main sources, explain their function, and detail the nature of their interrelationships. Moreover, the model was intended to be empirically manageable, enabling precise predictions and generating a robust framework for future research.

The thesis also aimed to demonstrate the argumentative nature of the belief system and credibility, creating a bridge between such diverse fields as rhetoric, argumentation, the psychology of reasoning, and the psychology of discourse processing. This interdisciplinary link was intended to highlight the connections and mutual relevance of these areas in the study of credibility and belief change.

In experimental terms, the thesis focused on verifying key aspects of the theoretical model, particularly regarding the mechanisms involved in the deductive aspect of belief change. This included investigating the effectiveness of the prescriptions of argumentation and rhetoric within the context of psychological research on beliefs, as well as examining the systematic relationships among various degrees of credibility that combine in deductive reasoning to result in noticeable changes in credibility. The outlined objectives were addressed through a series of experiments designed to test the validity of the proposed theoretical model. The obtained results seek to advance our understanding of how beliefs are formed, sustained, and modified in argumentative and rhetorical contexts, as well as to offer new avenues for future research in this field.

2. General Discussion of the Theoretical Model

Throughout the presentation of this theoretical model, it has been intended to highlight how human belief systems are inherently adaptive and dynamic, continually adjusting in response to new information and environmental stimuli. The model emphasizes the interdependence between individual experience, communication, and inference as sources of belief formation, influenced by the cultural and social context. The model combines concepts of communication and reasoning to highlight the importance of the processes of transmission and interpretation of information in modifying or strengthening the existing beliefs of the receiver. Discourse is not merely a vehicle for information, but an active mechanism for belief formation. When this occurs, the credibility of the communicator, the prior beliefs of the listeners, and the characteristics of the message must be considered. This message will be evaluated through a validation process that optimizes the coherence of the credibilities involved.

An area that deserves additional attention is the relationship between individual and cultural beliefs, and how power structures and social dynamics influence the perception and acceptance of new beliefs. This is particularly relevant in a globalized world where intercultural interactions are frequent and can challenge pre-existing beliefs, leading to significant reconfigurations in individual and collective belief systems. Moreover, belief systems, although robust, are subject to fallacies and errors, such as confirmation bias and resistance to contradictions, which can compromise the adaptability and accuracy of the system. Recognizing these limitations can help develop strategies that promote critical thinking in the acceptance of new information.

2.1. About the Model Elements

The belief system model presented in this thesis offers an integrated view of how beliefs are formed, sustained, and fluctuate within a complex cognitive framework. Starting from traditional conceptions of beliefs, it modifies its components by detailing their nature and internal structure, resulting in an S – A – R scheme (Subjectivity – Adherence – Representation). This model describes subjectivity both as a repository where beliefs are recorded and as an active entity that interacts with and modifies its mental representations in response to new information. This process is bidirectional: subjectivity is shaped by the representations it integrates and, in turn, influences the formation of new representations. Essentially, subjectivity is the core of the belief system, a source of intuitions, a support for perceptions, and the recipient and interpreter of communications. It includes the cognitive mechanisms that enable the formation of beliefs and memory, where beliefs are recorded, organized, and adjusted in their updating. This dynamism underscores the adaptive capacity of human thought, necessary to navigate and respond to a constantly changing world.

In this context, the concept of representation is crucial. Mental representations must be complex and coherent to be considered credible within a belief system, integrating categories connected through justification relationships. This connection acts as an adhesive, essential for the solidity of the representation. Mental representations are understood as mental models of situations and general schemes. The former are dynamic and updated with changes in their

connections. These models are composed of various components that vary in concreteness, all interconnected. On the other hand, the schemes are stereotypical structures that evolve more slowly and include categories, generally more abstract, that are also interconnected. The present model proposes that what flows through the representational components of beliefs is justification. In linguistic terms, justification allows a predicate to be applicable to a subject.

Detailed discussions on adherence and topical relationships within the belief system aimed to highlight how beliefs are not maintained in isolation. The credibility of a belief is supported by the strength of the evidence backing it and by how it coherently connects with other beliefs within an individual's cognitive network. Justification and refutation relationships are crucial for this process. Thus, elements already justified in the system will transmit justification to a target belief by virtue of consistent and relevant connections with it. This justification will nourish the relationships between the components of the target representation, making it more credible. The fact that the justified information in the system takes the structure of an argumentative scheme allows for accepting or rejecting new information, depending on whether it can be derived from that.

2.2. A Balance Between Stability and Flexibility

The distinction between foundational and derived beliefs in the model provides a framework for understanding how some beliefs serve as basic pillars of thought, while others are more susceptible to revision and change. This underscores the importance of a balance between the stability needed for cognitive coherence and the flexibility required to adapt to new contexts, events, and evidence. The acceptance of the fallibility and revisability of beliefs, including foundational ones, is a distinctive feature of the model that enhances its suitability to everyday reality. The stability of a belief also depends on its centrality and the density of its connections within the belief system. Beliefs that are multiply connected tend to be more stable. Both foundational beliefs and schema-type beliefs exhibit high stability due to their central nature and extensive connectivity. However, it is important to highlight that, unlike foundational beliefs, schema-type beliefs are formed by inductive inference, through the abstraction of qualities from numerous informational inputs. These are beliefs that stabilize through the justification received from evidence and, in turn, serve to justify a multitude of other beliefs. In this way, their stability is reinforced within the system.

Unlike these, beliefs derived from deductive processes, which consist of applying general rules to specific cases, tend to be more unstable and subject to change. This is because specific cases can vary significantly, affecting the applicability of general rules in particular situations. As we have seen in the presentation of the model, the interaction between induction and deduction processes plays a fundamental role in the stability and adaptability of the representations within the belief system, reflecting how different types of inference contribute in various ways to its coherence and dynamism.

2.3. Explanatory Relationships

In analysing the model, it is essential to recognize that the predominant relationships within the belief system are explanatory, where justification flows through topical relationships instantiated in inductive and deductive processes. These processes connect new information with existing beliefs through an argumentative structure that allows for its interpretation and the assignment of a degree of credibility. In this way, an individual's belief system responds to information that confirms existing expectations as well as to that which challenges them. The human belief system manages contradictory information through a proactive process of seeking explanations. This process, supported by abductive, inductive, and deductive inferences, restores coherence and simultaneously expands the conceptual framework of the system, allowing it to adapt to the flow of information.

Topical relationships provide a framework that not only allows for the interpretation and justification of unexpected information, turning it into a new belief, but also facilitates a smooth and coherent integration within the individual's pre-existing belief system. These inferences act as a bridge between the unknown and the known, enabling both the prediction and coherent interpretation of new data. Moreover, the argumentative structure emphasizes the importance of clear articulation to support adaptation and change within the belief system, justifying, validating, and consolidating new beliefs based on those already accepted.

In the present model, a central feature of the human belief system is considered to be its ability to maintain a balance between the stability of knowledge, necessary for prediction, and the adaptability required by environmental changes. Explanatory inferences, by facilitating the management of contradiction and uncertainty, preserve the coherence of the system and ensure its adaptation and modification, reflecting a capacity for learning and evolution essential for individual success.

2.4. Types of Explanatory Relationships in the Belief System

As just announced in the previous section, explanatory relationships are classified into three main types, each with a specific role in the system:

Abduction: An initial method for formulating explanatory hypotheses in response to unexpected data or events, using prior knowledge to postulate causes for surprising phenomena.

Induction: Crucial in the formulation of general rules from observed regularities and strengthens the credibility of explanations, facilitating the validation and integration of new beliefs within the existing cognitive framework.

Deduction: Uses the product of induction to structure reasoning that not only integrates new beliefs but also transmits their justification to these new representations, ensuring consistency within the belief system.

2.4.1. Induction: Deepening the Process

2.4.1.1. *Hypothesis Formation*

A fundamental aspect of this process involves the generation of general rules from the observation of regularities. This approach transforms direct and specific observations into broader generalizations, crystallizing into schemas within long-term memory. These schemas serve as frames of reference that guide our perception and interpretation of new data and situations. For example, repeated observation that a certain behavior leads to specific results can lead us to formulate a general rule that predicts these results under similar conditions. This process strengthens our existing knowledge and expands our ability to apply it to contexts not directly experienced.

By transforming specific observations into broad generalizations, induction facilitates the incorporation of new incoming information into the belief system, aligning it with pre-existing structures and enabling rapid, smooth, and coherently organized integration. This constitutes one of the fundamental processes in the system's evolution. Through the processes of abstraction applied to concrete examples, induction constructs the general rules and schemas, that is, the stable beliefs that we call knowledge. These elements provide the necessary infrastructure for the evaluation of the credibility of other informational fragments, their prediction, or their explanation.

2.4.1.2. *Exemplification*

Continuing the discussion on induction, we now highlight how specific and relevant examples are applied to reinforce general hypotheses, as well as to project these hypotheses onto new, particular instances that are unknown in some relevant aspect. This method employs paradigmatic cases that serve a dual purpose. First, they influence the perception of the credibility of a general hypothesis: since they are representative of the hypothetical rule, they demonstrate its confirmation. Second, they affect the perception of the credibility of conjectures about particular instances. If a new instance is similar to the paradigmatic cases, then, by virtue of the rule, it can be stated that what applies to the paradigmatic cases also applies to the new instance. Furthermore, it illustrates how variability among typical examples can increase the credibility of inductive conclusions.

This approach to exemplification is particularly applicable in contexts such as the legal field, where careful selection of relevant examples is crucial for applying general theoretical rules to new situations, moving from the particular to the particular through the general rule. Exemplification not only supports the hypotheses within the belief system but also reinforces confidence in them, demonstrating their applicability in real-world contexts.

2.4.2. Deduction: Adaptation and Evolution in the Belief System

Exploring the various aspects of deduction, we have identified some elements of its classical conception that persist and adapt in probabilistic and psychological models to address the subjectivity and uncertainty of our everyday experience. These elements demonstrate how certain interpretations of deductive logic are intrinsically linked to the explanation and justification of phenomena, as well as to the organization and understanding of our beliefs. Classical deduction,

traditionally marked by a rigid focus on truth and certainty, entails significant limitations when applied to the complexity of human reasoning and real-life situations where certainties are rare. Integrating elements of subjective probability into the deductive process allows the system to have the necessary flexibility to derive fallible or retractable conclusions more suited to the limitations of human knowledge. This new form of deduction takes into account the importance of personal interpretations and individual experiences in judgment formation.

At the core of this model, the transmission of belief justification is presented as a process as crucial as the preservation of the truth of propositions in older approaches. The degrees of subjective justification of beliefs, resulting from foundational beliefs, inductive processes, and epistemic validations, are the basis on which deductive reasoning processes operate. From there, deductive processes focus on evaluating the reliability of inferential calculations to determine the credibility of new information. Exploring the interrelationship between deduction and induction in this context reveals that both modes of reasoning not only coexist but also complement and intertwine within the belief system. Together, they allow for adjusting the credibility to information stemming from new evidence, as well as to new interpretations of that evidence, promoting the proper accommodation of the belief system to reality.

2.5. On the Measurement of Credibilities as Probabilities

Following the steps of the New Paradigm, measuring credibilities as subjective probabilities represents a significant advance in our psychological understanding. This approach provides valuable insights into how people formulate, sustain, and adjust their beliefs within the system. The use of a quantitative framework allows for a systematic and precise evaluation of credibilities. This facilitates the analysis of how individuals adjust the connections between the components of their representations and with the active belief system in working memory at any given time. Additionally, it quantifies the intensity with which beliefs are held, respecting their subjective nature and allowing for a more detailed description of their adaptation to new information or critical revisions.

In our model, we establish clear correspondences between modal terms—impossibility, necessity, and contingent possibility—and specific probabilistic values such as 0, 1, and others within the range. This relationship is fundamental to understanding the structure of certainty and uncertainty in the belief system and what credibility measurements truly mean. Adopting a probabilistic perspective to measure credibilities facilitates their treatment as both an independent and dependent variable in empirical research. Moreover, integrating the theory of mental models with probabilistic measurement offers a dual, quantitative and qualitative approach to exploring beliefs. This methodological duality could open new avenues of research to explore how these two approaches can complement each other and provide a more comprehensive understanding of the functioning of beliefs within the cognitive system.

It is essential to recognize the limitations of the model, especially the assumptions inherent in the use of subjective probabilities. Probability estimates vary significantly among individuals,

presenting challenges for standardization and interpretation of results. This variability points to the possible need to develop methods to calibrate and validate credibility measurements, ensuring their applicability and accuracy in various contexts. In any case, this merits a detailed conceptual analysis. Finally, while our approach aligns with the trend towards quantification and mathematical modeling of psychological processes, it is important to maintain an open dialogue about the advantages and disadvantages of this approach compared to traditional qualitative models. This debate will continue to enrich both theoretical development and methodological innovation in the study of beliefs.

2.6. Integration of Probabilistic Validity and Consistency in Deduction with Beliefs

In the development of our theoretical model of the belief system, we aim to understand how people reason and justify their beliefs in contexts of uncertainty. By integrating probabilistic validity and consistency as fundamental principles, our model ensures the rationality of deductive operations related to the magnitudes of credibility. This model emphasizes the importance of effectively managing the credibilities assigned to our representations of the world, offering a structured framework to guide this management coherently. Probabilistic validity, or p -validity, adjusts deductive logic to the computation of degrees of belief that arise in everyday thinking, providing a certain normative framework. This framework ensures that the incredibility of conclusions does not exceed the joint incredibility of the premises. This graduated way of understanding credibility seems to better fit with how people form and sustain beliefs under conditions of uncertainty. Moreover, by dictating that the incredibility of conclusions does not exceed that of the premises, this approach also supports the consistency and analyticity inherent in deduction.

Probabilistic consistency is conceived here as the guiding characteristic of changes in credibilities. This is based on the theoretical assumption that the belief system tends to remain free of contradictions. However, it is important to understand that probabilistic consistency does not apply to the entire belief system due to computational limitations. Instead, it focuses solely on the *local belief system*, that is, on the elements active in working memory at any given time, where the system can detect and resolve contradictions. In this context, the multiplicative hypothesis for combining credibility in deductive arguments, especially in enthymemes, is introduced. This is the core hypothesis of the model, and it is based on the properties of p -validity, probabilistic consistency, and the consideration of credibility as a diminishing magnitude in its combination. This hypothesis aims to describe how premises transmit their credibility to conclusions while respecting the axioms of probability.

One consequence of the multiplication hypothesis is the epistemic rule of the weakest link, whereby, in ideal argumentation situations, the posterior credibility of the conclusion should not be greater than that of the least believed premise. However, when the credibility derived from the multiplication is lower than the a priori credibility in the critical idea, its credibility can oscillate between the multiplication itself as the lower limit and the a priori credibility as the upper limit. The introduction of the argument strength function and the enhanced argumentation standard to

evaluate the effectiveness of arguments marks a significant advance in the application of the multiplication hypothesis. This criterion measures the minimum joint credibility of the premises necessary to increase the credibility of an idea or conclusion within reasonable limits. It thus provides a suitable tool for accurately analysing how premises influence conclusions.

This model offers a detailed description of the dynamics of belief formation and maintenance within the local belief system. It details a clear and straightforward method for integrating the credibility of premises, focusing on how the system maintains internal coherence at the focus of attention at any given time. The multiplicative hypothesis provides an adequate theoretical foundation for understanding how credibilities function in the deductive part of the system, both under conditions that meet the argumentation standard and in conditions where argumentation is based on premises not believed by individuals. This approach allows for a detailed understanding of how variations in the credibility of the premises can affect the conclusions.

Therefore, the implementation of the multiplicative hypothesis represents a significant advance in our understanding of the computation of credibilities through enthymemes in everyday reasoning. It shows how probabilistic deduction modifies credibilities, measured as subjective probabilities, within the belief system through argumentative structures.

3. General Discussion of the Experiments

3.1. Experiment 1

In 'Experiment 1', the structuring and justification of beliefs were investigated upon the presentation of arguments, testing some rhetorical concepts. This study integrated the psychology of reasoning with practical assessments of enthymemes and their impact on the credibility of controversial claims. The experiment examined the hypothesis that warrants, by expressing hypothetical rules, can transfer credibility from data premises, which represent knowledge about accepted facts, to new critical assertions about future events. These assertions, framed within argumentative structures, function as conclusions derived from the interaction between warrants and data. The study proposed that if new information can be derived from previous information, then it can be explained by it and, therefore, integrated into the belief system. Thus, the experiment sought to demonstrate how the relationships of deductive explanation operate in the system through the cause-effect argumentation scheme.

Having established the central hypothesis of the study, let us examine more closely the main independent variable and its impact on the experiment. The primary independent variable of the study was the credibility of the warrant premises, which presented hypothetical binary relational rules perceived as causal. These rules were articulated through conditional sentences, with a cause in the antecedent and an effect in the consequent. To vary the credibility of the warrants, the strength of the association perceived between the cause and the effect by the participants was

manipulated; a greater strength of association increased the credibility of the rule, enhancing its capacity to transfer credibility from the data premises, which were consistently held to be credible at high levels. Thus, variations in the credibility of the warrants explain the differences observed in the critical assertions, which functioned as conclusions in the arguments.

This experiment evaluated the effect of enthymemes on the credibility of ideas initially disbelieved by the participants, using a pretest-posttest design with a control group. Both the experimental and control group participants assessed the same critical assertions in the pretest, presented in isolation. In the posttest, while the control group continued to evaluate these assertions in isolation, the experimental group was presented with them framed within arguments that included the data premises and the warrants. These arguments varied in the strength of association of the warrants. An increase in credibility was expected only in cases where the warrants had a strong association, with no significant differences anticipated in other comparisons, either by measurement timing (pretest versus posttest) or by group (experimental versus control).

The results confirmed the expected differences and provided clear insights into the hypotheses posited, as detailed below. In the control group, the absence of variations reaffirmed that the differences observed in the experimental group could not be attributed to external variables, such as the mere passage of time. Within the experimental group, the differences demonstrated that the increase in credibility was directly related to the strength of association in the warrants, and consequently, to their own credibility. Thus, the impact of the argumentative structure and the strength of the premises on the transfer of credibility is clearly demonstrated in the results of the experiment. It was observed that the credibility of the conclusions could be significantly increased through warrants with strong association, which provides valuable insights into the mechanisms of persuasion and conviction.

Furthermore, the experimental structure required dividing the study into two parallel but complementary experiments due to the manipulation of materials. The results obtained in both experiments corroborated the initial hypotheses, reinforcing the validity of the findings. However, significant differences were observed in the response to warrants of weak association between the two experimental groups. This phenomenon illustrates another prediction of the model: if the credibility of the premises results in a posterior credibility lower than the initial, individuals might ignore these premises or use them to evaluate the conclusion. In group G2, ignoring the premises maintained their subsequent credibility unchanged from the initial, whereas in group G1, considering them reduced their subsequent credibility, as expected by the model.

This study also emphasises the need for a more integrated and multifaceted approach in the study of the psychology of argumentation and rhetoric. Testing the effectiveness of enthymemes in a controlled experimental environment contributes to the theoretical development of the field while offering practical guidelines for formulating more persuasive and credible arguments, applicable in public spheres. "Experiment 1" enhances our understanding of how argumentative structures affect the credibility and acceptance of conclusions within a belief system. It validates

and expands the proposed theoretical model, demonstrating how the principles of argumentation and rhetoric are not only applicable in theory but also have a direct and measurable impact on how people accept new information.

3.2. Experiment 2

In Experiment 2, a large number of hypotheses concerning very precise mechanisms of operation within the proposed belief system were tested. For this reason, we will extend the general discussion and final remarks about what these results tell us about the proposed theoretical model.

3.2.1. Influence of the Credibility of Premises on the Evaluation of the Conclusion

In this study, an effort was made to eliminate any effect of the participants' context model by maintaining uniformity in situational elements during the individual presentation of the experiment. Additionally, the arguments were presented neither preceded nor followed by any text. They were not presented within the scope of a broader discourse or conversation (there was no dialogic context). Likewise, the form of the arguments remained constant: all arguments were syllogisms of the first figure. However, the contents of the premises were modified to yield different evaluations of their credibility based on the individuals' prior knowledge and beliefs, thus obtaining the four conditions of credibility. On the other hand, the content of the conclusions remained constant across these four conditions. Any difference in their resulting credibility could not be due to their content, but rather to their interpretation based on the framework designated by the premises.

3.2.2. On the Hypothesis of Equitable Transfer and Multiplication

The results obtained in this experiment provide significant empirical validation for the hypotheses proposed in the theoretical model of credibility transfer and dynamics of deductive reasoning. Essentially, the experimental findings corroborate the theory that the credibility of both the data premises and the warrant premises is fundamental in influencing the credibility of the conclusions. This aligns with our hypothesis of Equitable Transfer, which presupposes an equally important role for both premises in the argumentation process.

The significant effect of the credibility factor of the warrant premise and the data premise, as well as the interaction between them, suggests that the credibility of the conclusion is affected by the variability in the credibility of both premises. This observation is consistent with the Multiplicative Combination Hypothesis, which suggests that credibility is transmitted and combined through a multiplication operation in deductive reasoning. The specific results, where the subsequent credibility of the conclusion not only depends on the premises but can also be influenced by the prior credibility of the conclusion, reaffirm the flexibility and realism of our theoretical model, which considers the adaptability of the belief system in response to variable information. The finding that the order of premises does not have a principal effect confirms the stability of the proposed model, demonstrating that the sequence of presentation does not influence the credibility assigned to the conclusions. This reflects the appropriate conceptualisation of the belief system as a network of argumentation where the sequence of presentation is less

relevant than the strength of the argumentative connections and the inherent credibility of the representations.

Finally, the significant correlation between the observed credibilities and those predicted by the multiplication model, particularly high in the HH condition where argumentation standards are met, validates the application of this model in assessing credibility in real deductive reasoning contexts. In addition to confirming the empirical validity of the model, it also demonstrates its applicability in practice, providing an effective tool for evaluating and understanding how credibility is manipulated and transformed through deductive reasoning in everyday situations.

These experimental results provide solid support for the developed theoretical model, demonstrating that the dynamics of credibility in deductive reasoning can be understood and predicted using a framework of credibility transfer based on principles of probabilistic validity and consistency.

3.2.3. On the Validation of the Limits of the Range of Posterior Credibility

The results obtained in this experiment provide empirical validation of the hypotheses posited in our theoretical model concerning the limits within which posterior credibility can fluctuate. In particular, the Hypothesis of the Weakest Link receives robust empirical support, demonstrating that the credibility of the conclusion is predominantly determined by the least believed premise. This observation is perfectly met under conditions where both premises have high credibility (HH). In these cases, the credibility of the conclusion does not exceed that of the least believed premise. This corroborates that the multiplication operation respects an upper limit, determined by the credibility of the least believed premise.

The experimental results confirm that the weakest link rule is met even when one of the premises has low credibility (HL and LH cases), albeit imperfectly. Under these conditions, the posterior credibility was equal to or lower than the prior credibility. The correlations show that, although the posterior credibility is strongly influenced by the credibility of the least credible premise, it tends to exceed it, suggesting a possible attraction effect by the prior credibility. In those cases, the posterior credibility always exceeded that of the least believed premise, indicating that the compliance with the rule was lax. Moreover, the posterior credibility also always exceeded the prediction obtained through multiplication, although it did not surpass the prior credibility. Thus, the posterior credibility met the model's expectations, remaining within the established limits: multiplication as the lower limit and prior credibility as the upper limit.

From a theoretical standpoint, the confirmation of the weakest link rule in contexts where the premises have high credibility strengthens the integration of probability theory into deductive reasoning compared to other plausibilist proposals. This demonstrates that the model is not only theoretically coherent but also empirically validatable, enhancing our ability to predict and explain changes in credibility within the belief system. These experimental results bolster the assumed precision of the proposed theoretical model and also highlight its empirical relevance for

understanding the transmission and modification of credibility in deductive reasoning. The model proves to be an effective tool for predicting how different levels of premise credibility influence the final credibility of the conclusions.

3.2.4. On the Influence of Prior Credibility and the Multiplication of Premises on the Variability of Posterior Credibility

The discussion regarding the influence of prior credibility and the multiplication of premises on the variability of posterior credibility reflects important aspects of the proposed theoretical model. In the HH condition, it is observed that the increase in the credibility of conclusions is primarily attributed to the credibility of the premises, with a lesser, though significant, role of prior credibility. This result supports to some extent the hypothesis that in contexts where the premises are highly credible, prior credibility plays a minimal (or null) role in explaining the variability observed in the conclusions. This finding validates the theoretical model, demonstrating that the belief system adheres to the rules of multiplication in standard argumentation situations, maintaining probabilistic consistency that reinforces the structure of the argument and the transfer of credibility. Despite this, it is acknowledged that prior credibility was expected to exert no influence in this condition.

On the other hand, in the L conditions (HL, LH, and LL), the posterior credibility of the conclusions is equal to or lower than the prior credibility, suggesting that the influence of prior credibility could be more notable in scenarios with premises of low credibility. In these contexts, the posterior credibility may exceed that of the least believed premise due to a possible attracting effect of the prior credibility. This reflects how the model adapts to the variability of situations where reasoning must operate outside of standard argumentative conditions. Unlike in HH, in HL and LH, the prior credibility exhibited a considerable influence, although less than the combined action of the premises.

The multiple regression showed that the multiplication of the credibilities of the premises is a more consistent and robust predictor of the credibility of the conclusions than the credibility of the premises taken individually. It also demonstrated that the multiplication of the credibility of the premises is a predictor equivalent to that of the credibility of the premises acting together, which confirms another key aspect of the proposed theoretical model: that multiplication is simply the combination of the credibility of the premises. This integrative approach not only confirms that multiplication acts as a unified and coherent combination of the credibility of the premises, but it also highlights its ability to effectively predict the behaviour of posterior credibility in various argumentative contexts.

3.2.5. Conclusions

The general discussion and final reflections on the experimental results broadly reflect the validity and applicability of the proposed theoretical model, which addresses the transfer of credibility and the dynamics of deductive reasoning. The empirical findings not only confirm the strength of the model in standard argumentative conditions but also its efficacy in complex and variable scenarios,

with less credible premises. The study demonstrates that the operation of multiplication is essential in modulating credibility for reasoning with credibilities in the deductive aspect of the system. This experiment has shown how, given equal conditions of formal structure and context, differences in the credibility of the same conclusion can be achieved based on differences in the credibility of the premises. Furthermore, the correspondence between the credibilities of the conclusions and those predicted by the proposed computation of the premises' credibilities has been verified.

This integrative approach confirms that multiplication acts as a unified and coherent combination of the credibility of the premises, and it also highlights its ability to effectively predict the behaviour of posterior credibility in various argumentative contexts. The experimental results validate that the credibility of the conclusions is predominantly determined by the least believed premise. When an argument raises the posterior credibility of a conclusion above its prior credibility, the credibility of the least believed premise acts as an upper limit on the credibility assignable to that conclusion.

In conditions where the premises have low credibility, this research has shown that the posterior credibility does not exceed the prior credibility, thus confirming the theoretical expectations about the limits of the range of posterior credibility. This behaviour demonstrates the model's ability to adjust the credibility assessment conservatively, avoiding overestimations that could compromise the integrity of the belief system. Moreover, consistency with probability theory and the integration of the multiplication of credibilities highlight the probabilistic coherence of the model, enhancing the accuracy in predicting and explaining changes in credibility within the belief system.

The research also reinforced the "Enhanced Argumentation Standard" and the "Criterion for Becoming a Belief". It was validated that for a conclusion to effectively increase its credibility, the combination of the incredibility of the premises must be lower than the incredibility of the conclusion a priori. This finding was supported by the use of the Relative Disbelief Index, which correlates inversely with an increase in the credibility of the conclusions. Additionally, the results highlight that a combination of premises whose combined incredibility is less than 0.5 generally results in a conclusion credibility that exceeds this threshold, which is essential for persuasion and acceptance within the belief system. These results demonstrate the accuracy of the theoretical model and offer clear guidelines for formulating effective arguments in deductive reasoning with beliefs. These guidelines ensure that the premises are solid enough to increase the credibility of the conclusions. The proposed model is an effective tool for predicting the impact of arguments on the credibility of statements, facilitating a coherent adaptation to new evidence and contexts.

4. Theoretical and Applied Implications of the Argumentative Model of the Belief System

4.1. Adaptability and Dynamics of Belief Systems

This research explores the inherent adaptability of human belief systems, which are dynamic and continuously adjust to new challenges and data, both external and internal. The study details the specific mechanisms by which beliefs are modified, highlighting the crucial interaction between different types of inference—abduction, and especially induction and deduction—and how these contribute to the formation and evolution of beliefs. Additionally, it explores how the social and cultural context influences the evaluation, acceptance, and modification of beliefs.

This work conceptualizes inferential processes as adaptive mechanisms that construct and also facilitate the evolution of the belief system. By analysing how these inferences enable the belief system to respond flexibly and coherently to new interpretive challenges, the model underscores the critical function of these processes in the adaptation and evolution of the belief system. This approach provides an enriched understanding of the cognitive flexibility that characterizes human belief systems and how they are shaped by, and in turn shape, their social and cultural context.

4.2. Advances in Understanding the Relationships Between Deductive Reasoning and Belief Formation

This work represents an advancement in understanding the deductive component of the combination of credibilities through the integration of principles of validity and probabilistic consistency. This approach has enabled the development of a more realistic theoretical framework for assessing how people manage and justify their beliefs in situations of uncertainty. By incorporating these metrics into the analysis of reasoning, this research offers a new foundation for understanding the fluctuations in the credibility assigned to different types of information and how these are processed within a belief system. The proposed model facilitates a more nuanced understanding of how deductive reasoning with varying intensities of credibility is conducted.

This work aims to contribute to the understanding of human reasoning in everyday life. In our daily lives, we tend to reason about thematic and everyday content rather than artificial symbols such as letters, geometric figures, or colours. We generally hold these contents with some caution, not with the certainty that they are true. And we do so based on our knowledge and from a personal perspective, not as if we were characters transported to fictitious contexts. This approach aligns with the spirit of other works in the psychology of reasoning and discourse comprehension, which have delved into exploring the "core processes" of cognition. Thus, this study seeks to reclaim the approach of pioneering research in logic, dialectic, and rhetoric, focused more on human beliefs and opinions than on mathematical or stylistic details. This integrative approach places at the base of the model the tendency towards unity and consistency of the human cognitive system.

A central aspect of this work is the S – A – R (Subjectivity – Adhesion – Representation) scheme, which represents an evolution from traditional approaches to studying the necessary characteristics

of the subject and representation for belief formation. This model addresses subjectivity not only as a repository of beliefs but also as an active entity that constantly modifies its mental representations. It is considered that representations must comprise interrelated coherent categories, thus strengthening the belief structure. Furthermore, the model seeks to define credibility and explore how to address it empirically, offering a comprehensive framework for analysing the dynamics of beliefs. The model redefines credibility as a relational magnitude that links representation with subjectivity at various degrees of intensity. Inspired by rhetorical traditions, we refer to this variable intensity link as 'adhesion'. This adhesion is produced by justification. By adhering a representation to subjectivity, justification is transmitted to the representation, strengthening the internal bonds among its components. This enhances the perceived coherence of the representation and the consistency of the system. This process highlights the model's ability to capture the dynamic interactions between old beliefs and new information, offering a detailed framework for understanding how beliefs are continuously adjusted and validated through a process of evaluation and re-evaluation.

The introduction of the argument strength function as a revised standard of argumentation to measure the effectiveness of arguments and evaluate the credibility of conclusions within a probabilistic framework is another methodological innovation of this research. This function quantifies how variations in the strength of the premises affect the acceptability of conclusions in argumentative practice. This allows for the evaluation of argumentative efficacy and enhances the ability to predict and explain how variations in the credibility of the premises influence the conclusions.

4.3. The Rule of Theophrastus in Deductive Reasoning: Evaluation and Proposal for Adaptation

A critical and comprehensive analysis of the Rule of Theophrastus has been conducted, exploring its evolution from its roots in modal and assertoric logic to its adaptation in the frameworks of contemporary plausibilistic reasoning (e.g., Pollock, 2001; Rescher, 1976; Walton, 1992). Its historical origin has been presented, and it has been shown how it has been interpreted, integrated, and modified in current plausibilistic deductive models. This study has attempted to provide a perspective on the interaction between historical logical traditions and modern argumentative theories. Additionally, it has critically examined the contemporary adaptations of the rule in the context of informal logic and argumentation.

In terms of emerging criticisms, especially from cognitive psychology, the suitability of the Rule of Theophrastus for assessing plausibility is questioned. These criticisms highlight possible limitations of the current interpretation of the rule, suggesting that it may not adequately capture the complexity of human reasoning and the attribution of credibility in scenarios of seemingly contradictory information. The reinterpretation of the Rule of Theophrastus in plausibilistic logic resembles the deductive validity of classical logic, which guarantees that if the premises are true, the conclusions will also be true. Similarly, in plausibilistic logic, the validity of an argument ensures

that if the premises are plausible, the conclusion must also be so. Thus, it is not possible to obtain an implausible conclusion from plausible premises. In this framework, the least plausible premise sets the lower limit of plausibility for the conclusion: the conclusion cannot be less plausible than the least plausible premise.

In this thesis, the Hypothesis of Multiplicative Combination for the credibility of premises has been defended. This hypothesis is based on three key assumptions of the model: credibility is a decreasing magnitude, and it respects both validity and probabilistic consistency. According to probabilistic validity, the credibility of a conclusion must accumulate the incredibility of its premises. This determines that it is a decreasing magnitude. Adding the requirement of consistency, it turns out that the operation that best describes the combination of the credibilities of the premises is multiplication. This form of combination predicts that it is possible to obtain disbelieved conclusions from believed premises. Experiment 2 has demonstrated the validity of this claim, showing the effectiveness of the revised argumentation standard. Additionally, a corollary of multiplication is that, in standard argumentation, the credibility of the conclusion will not exceed that of the least believed premise. This study has also found empirical support for this corollary. Therefore, based on these definitions and the data presented, the Rule of Theophrastus is unsustainable from this model.

This is the contribution that this model has made to the debate on the Rule of Theophrastus among proponents of plausibilistic and probabilistic logic in argumentation and psychology. This innovation provides a potentially suitable framework for assessing plausibility and credibility in deductive reasoning with rhetorical or everyday arguments. In the model presented here, we adapt the Rule of Theophrastus by recovering the rule of *de peiorem* and converting it into the epistemic rule of the weakest part. This part of "Experiment 2" enriches the theoretical discussion on certain combination rules debated in some current theories of argumentation, providing empirical data. By challenging and reevaluating the Rule of Theophrastus, a critical dialogue on plausibility and credibility is encouraged. This shows the appropriateness of adapting plausibilistic theories of argumentation to cognitive perspectives on information processing, thus providing a broader understanding of the qualitative and quantitative nature of credibility.

4.4. Interdisciplinarity and Empirical Applicability of the Thesis

This research has successfully bridged significant gaps between disciplines that traditionally do not interact extensively, such as rhetoric, dialectics, and various areas of psychology. We believe this interdisciplinary approach enhances our understanding of belief systems and the dynamics of change and credibility. Moreover, it allows for the improvement of the methodologies used for their study. The connection formed between these fields provides theoretical support to the model and points towards practical relevance. After all, rhetoric and dialectics were (or are) eminently practical disciplines. Indeed, the findings of this study could potentially be applied in practical contexts such as education, negotiation, advertising, and even cognitive restructuring therapy.

This research advances the renewed trend towards integrating dialectic and rhetorical argumentation, probabilistic reasoning, and discourse to comprehend how credibility functions (e.g., Krzyżanowska, Collins, Hahn, 2017; Oaksford & Hahn, 2012; Tešić & Hahn, 2023). While many studies under the New Paradigm related to argumentation have extensively covered the inductive aspects of credibility dynamics (e.g., Hayes & Heit, 2018; Oaksford & Hahn, 2007; Rehder, & Hastie, 2001), the deductive side has been less explored. This study aims to fill some of that gap, both theoretically and empirically. To achieve this, it revisits the rhetorical concept of the enthymeme, understood as a deductive argument about probabilities designed to enhance the credibility of its conclusion for a specific audience. Based on this concept, the experimental conditions required to perform comparisons aimed at demonstrating the effects of the premises' credibility are established. This has led to the development of specific methodologies to gain a deeper understanding of the cognitive processes and computations involved in the formation and modification of beliefs. The interaction between different disciplines can deepen our understanding of the dynamics of beliefs and justification.

4.5. Methodological Contributions to Belief Research

This thesis introduces an innovative methodology for the empirical analysis of beliefs and their dynamics, enabling detailed examination of the cognitive computations involved in the combination of credibilities. This approach facilitates the replication of results and the development of future studies while ensuring precise predictions and rigorous control over the variables of credibility computation. Specifically, in Experiment 2, only the middle term in the argument varies across four conditions of credibility, allowing for the systematic manipulation of the credibility of each premise to achieve all four possible combinations. The experimental design, similar to a Latin square, ensures that each item is evaluated in all combinations of premise credibility across different subjects, and that each subject experiences all credibility conditions through different items.

It is also important to highlight that the *first-person methodology* used, which employs the credibility ratings provided by the participants themselves, enhances the external validity of the results obtained. Added to this is the construction of arguments with realistic and everyday themes, about which the participants may hold beliefs. This combination of methodological features provides rigorous control of the variables, suitable for studying the dynamics of credibility through argumentation. Indeed, these form a valuable contribution to the field of belief research.

4.6. Contributions of the Research to the Understanding of Persuasion and its Multidisciplinary and Applied Implications

4.6.1. Understanding Persuasion

This research contributes to the study of the psychology of persuasion. In particular, it could serve to expand certain aspects of the Elaboration Likelihood Model (ELM) by Petty and Cacioppo (1986). The ELM distinguishes two routes of persuasion: the peripheral route, which uses external elements such as the attractiveness of the source and music to increase the attractiveness of the message in a non-rational manner; and the central route, which involves a logical and rational analysis of the

message, assessing its merits and how these align with the prior knowledge of the receiver. For this route to be effective, the receiver must be sufficiently motivated and capable of processing the message. The authors argue that the quality of the message directly influences its effectiveness, being integrated into the individual's belief network if it is high, or discarded if it is low. However, they acknowledge that it is still not well understood what makes an argument persuasive, an area that has motivated numerous recent studies in psychology and argumentation theory (e.g., Areni & Lutz, 1988; Areni, 2003; Hahn et al., 2009; Mercier, 2009; Van Enschot-Van Dijk, Hustinx, & Hoeken, 2003).

This research contributes to understanding what makes an argument persuasive, demonstrating that the premises employed must align with the audience's prior beliefs. This research has revealed that in arguments of the first figure, the credibility of the premises is combined through an operation similar to multiplication and must adhere to the revised standard of argumentation. During comprehension, incoming information activates relevant knowledge and undergoes a validation process where some elements justify the information while others contradict it, thus determining its credibility. For a message to be persuasive, it must present credible premises as reasons. It is crucial to know the audience's prior beliefs to properly configure these premises.

Unlike what the central route of the ELM prescribes, this work suggests that the validation of premises is routinely and efficiently carried out through *epistemic monitoring*, consuming few working memory resources (Richter, 2011; Richter et al., 2009). Furthermore, studies such as those by Lea (1995) and Lea, Mulligan & Walton (2005) could be interpreted as indicators that the computation of credibilities with enthymemes might be a routine process. Their studies showed that in simple reasoning, such as *modus ponens* (MP), people routinely move to the conclusion as soon as they process the relevant information from the premises. Thus, it could happen that with enthymemes, credibility is computed as soon as the necessary information is obtained. This, in turn, would be compatible with other studies in the field of discourse (e.g., Wiley & Myers, 2003).

All this does not mean that strategic scrutiny, which depends on the individual's motivation and capacity, does not occur. It might take place later, in a process called *epistemic elaboration* (Richter, 2011). In this sense, the reasoning that occurs routinely would not depend on skill or motivation, but the generation of supportive thoughts in favour of the message would. Therefore, in the central route, there must be degrees of intensity in the processing. Studies by Briñol and Petty (e.g. 2015) would support this interpretation, speaking of a continuum in levels of elaboration.

This model may be relevant to the psychology of persuasion and practical applications in marketing, advertising, journalism, and political propaganda. Its ability to precisely predict the credibility of an idea based on the multiplication of prior premises makes it a valuable and applicable tool, especially considering the enhanced standard of argumentation. Advertising professionals, who commonly use techniques from the peripheral route due to a lack of knowledge about the keys of the central route (Apter, 2007; León Sáez de Ybarra, 2014; O'Shaughnessy & O'Shaughnessy,

2004), can now use this model to present statements based on ideas already accepted by the audience. Market research plays a crucial role in revealing the beliefs and motivations of specific market segments, allowing for the identification of elements that can function as highly credible premises. This approach is equally applicable to political propaganda and debates in the media.

4.6.2. Practical Applications

On the other hand, this study could be relevant to research in education due to the similarities between enthymematic argumentation and the Socratic or maieutic educational method (e.g., Lee, Kim, & Kim, 2014), albeit without the use of explicit questioning. This method facilitates the acquisition of new beliefs or the modification of the credibility of existing ones. From two propositions considered true by the subject, a new conclusion is generated through reasoning. Similarly to how knowledge is constructed, where already understood concepts are used as a basis to understand new ones, in argumentation, ideas previously accepted as true are used to adopt new beliefs. Thus, accepting a new idea is akin to understanding something previously unknown.

The research conducted here could also be relevant for certain applications of cognitive restructuring in clinical psychology. It may be that both cognitive restructuring and belief change via argumentation operate on similar principles. Ultimately, in both cases, the change occurs by inducing a certain interpretation of reality in the listener. Indeed, it is possible to start from healthy and adaptive beliefs to modify the interpretation of the same fact of reality so that the new interpretation inherits those adaptive characteristics.

5. Limitations

5.1. Scope Limitations and Unaddressed Concepts

In the theoretical framework of this research, the concept of local belief system was introduced, understood as the set of closely related beliefs that are activated by new information and that can either justify or refute its credibility. Although the existence of the local belief system is suggested by numerous studies in discourse psychology (e.g., Halldorson & Singer, 2002; Singer, 2013) and reasoning (e.g., Brisson & Markovits, 2020; Cummins et al., 1991; Cummins, 1995; Quinn & Markovits, 1998; De Neys, 2010), this concept was not explored in the current experiment. It would be desirable to conduct research focused directly on detecting the activation of both refutational and supportive elements triggered by the content of the premises.

On the other hand, although the study demonstrated how the credibility of data and warrant premises is integrated to justify a conclusion, specific data were not collected to confirm significant differences in the *functional nature* of these types of premises at an empirical level. The data premise refers to concrete facts and acts as the original source of credibility, whereas the warrant premise consists of a general rule of relation and transmits this credibility. This theoretical

distinction was not a main focus of the empirical research, hence future studies should explore these differences to strengthen the theoretical foundations of the study.

Another area related to the distinction between the two types of premises that warrants attention is the effect of the premises' structure on the computations of credibility integration. If integration operates through multiplication with commutative properties, the differences between the types of premises might not be evident in the results. However, a deeper analysis of each type of premise could reveal differences in their cognitive processes. Perhaps, converting the conditional premise into an assertive Subject-Verb-Object structure could allow for more direct comparisons in relation to processing in future experiments, by standardising the form of presentation of both premises.

5.2. Methodological and Generalisability Limitations

This study faces several methodological limitations and challenges to the generalisation of its findings. The focus on the first syllogistic figure and the cause-effect argumentation scheme, although central to the most common modes of reasoning, limits the applicability of the results to other types of reasoning and argumentative structures. This may question whether the efficacy of the multiplicative combination model is universal or limited to the specific contexts addressed. The methodology employed, keeping all variables constant except for the credibility of the premises, strengthens the internal validity by effectively isolating this independent variable. However, the decontextualised presentation of the arguments and the specific manipulation of the variables may negatively affect the external validity of the study. Nevertheless, given that the participants were free to interpret the materials and that the topics addressed were contemporary and relevant, these factors indirectly contributed to partially mitigating this effect.

The specific content of the experimental materials, designed to be relevant in a particular context, faces obsolescence as cultural and temporal conditions change, limiting the ability to perform exact replications of the study in the future. This obsolescence reflects the adaptability of the belief system to a changing world and highlights the difficulty of creating long-lasting credibility databases. Finally, it is true that the credibility of the premises was effectively manipulated in each experimental condition. However, the indirect manipulation through the strength of association in Experiment 1 and the absence of a pretest to validate this manipulation in Experiment 2 could limit the interpretation of the results. Increasing the number of items could facilitate interpretation in the LL condition (with both premises of low credibility), but it could also induce fatigue in the participants.

It is essential that future research adopts broader experimental designs that expand the argumentative structures and populations studied. This will allow for a more precise capture of the credibility dynamics across a variety of argumentative contexts and produce results more generalizable to diverse situations. Moreover, it is important to incorporate a variety of formal, pragmatic, and emotional variables to enrich our understanding of the credibility process and its influence on the evaluation of arguments and the formation of beliefs in real-world contexts.

Although the current findings are significant, they represent only a part of an extensive field that merits more detailed explorations in future research on the dynamics of beliefs.

5.3. Demographic Limitations

This study predominantly focused on young university students, mostly women from the province of Tarragona. This specific demographic might limit the generalisation of the results to more varied populations, given that factors such as age, educational level, and geographical location could influence the evaluation and computation of credibilities in unforeseen ways. Moreover, it is prudent to acknowledge that cultural or geographical differences and demographic variables could affect the results in unexpected manners. Some proposals support this notion (Nisbett, 2004), while others do not (Johnson-Laird & Lee, 2006; Mercier, 2011).

This underscores the importance of maintaining a critical perspective towards the influence of unknown variables, including individual differences such as working memory capacity, which might also have a significant impact (e.g., De Neys et al., 2005). Although the combination of subjective probabilities is a basic and, we believe, automatic process, the specificity of the sample indicates the need to conduct future studies with more diverse populations to confirm the generalisability of our findings.

5.4. Limitations in the Unit of Measure

A limitation of these types of studies is the quantification of credibility using subjective probabilities, as discussed in the Part II or the thesis. This quantification, although useful for converting credibility into operational numerical values, introduces potential constraints.

Relative Nature of the Credibility Measure: Unlike standard measurement units such as the metre or kilogram, which are absolute and universal, subjective probability is relative to the subject and context. While a metre or a kilogram maintains its value consistently irrespective of the observer or context⁷⁰, subjective probability varies significantly among individuals and for the same individual at different times. This is because credibility depends on long-term memory and the local belief system activated by the context, presentation, and source of the information. In this sense, asking "What credibility does this statement have?" is less appropriate than asking "What credibility does this statement hold for you here and now?", highlighting its subjective and contextual nature, rather than being an objective property of the object.

Limitations of Subjectivity in Measuring Credibility: Although subjective probability is fundamental to our study, its inherent subjectivity introduces ambiguity. Differentiating, for example, between a credibility value of 0.7 and 0.8 could vary significantly between different individuals or even for the same individual at different times. There is no universal standard that ensures that intervals between numerical values are constant or that the same magnitude of credibility is consistently represented by the same subject on different occasions. In Part II, it is

⁷⁰ To be precise, the kilogram is a *relational measure*, which relates the mass of a body to a gravitational field. In that sense, it is similar to credibility, which relates a representation to an individual's subjectivity.

proposed that magnitudes of credibility are formed through the activation of intuitive theories and the simulation of contents in mental models, which are then expressed through modal qualifiers or quantified numerically in experiments. However, the transition from the perception of credibility to the assignment of a numerical value is subject to variability and error due to emotional, attentional, and memory factors, among others. This error affects the reliability of the mathematical operations performed and the predictions based on these measures, introducing a level of uncertainty into the results.

Implications for Result Interpretation: These features highlight the importance of further investigating the factors that influence the perception of credibility and the assignment of values, that is, the explicit representation of the underlying magnitudes. It is crucial that future research develops methods that can adequately capture this subjective complexity, which would allow for more accurate interpretation and data analysis.

Despite these limitations, the study ensures internal reference to the scale that appears to be consistently interpreted by all individuals, with limits of 0 and 1 and the midpoint at 0.5 representing the impossible, the necessary, and the uncertain, respectively. This anchoring, reinforced by the experimental design and the instructions provided, mitigates the potential adverse effects of the lack of equality in intervals or in the assignment of values to credibility magnitudes. Although subjective probability has limitations as a unit of measure for credibility, it may be the best available option for studying an inherently subjective phenomenon. It is therefore crucial to interpret the results within the context of these inherent limitations, since the subject of study is subjective.

5.5. Limitations in the Interpretation of Experiment 2 Results

Differences in HL and LH Averages: We observed an unexpected discrepancy in the averages of HL and LH, where it was anticipated that $HH > HL = LH > LL$. However, the results showed that $HL > LH$. It was initially thought that this could be due to variations in the averages of H and L between HL and LH, but post-hoc inspections were inconclusive regarding this hypothesis. These unexpected differences suggest that the relationship between the degree of credibility of the premises and their functional nature (guarantor or data-premise) might influence outcomes in ways not yet understood. This observation underscores the need for future research to explore how specific functional properties of the premises can affect the results (as previously noted).

Decrease in Correlations under L Conditions: When at least one premise is L, we noticed a decrease both in the average credibility of the conclusions and in the correlation between observed and expected scores. Although this is consistent with the multiplication model, the decrease in correlation under L conditions was not anticipated and requires a more detailed explanation. A possible cause could be a reduction in the variability of the conclusions due to the lower credibility of the premises, which in turn could reduce the ability to detect systematic patterns in the correlations. This interpretation, although consistent with some observed data, remains speculative and suggests the need for a more detailed analysis in future studies.

Difference between Conclusions and the Least Credible Premise under L Conditions: Another significant limitation is the discrepancy between the rule of the weakest link and the observed results, where the conclusions in L conditions consistently exceeded the credibility of the least credible premise. This could be due to variations in how participants evaluate the credibility of the conclusions, potentially influenced both by prior credibility and by the multiplication of the premises' credibilities. Some participants might be evaluating the conclusions based solely on their prior credibility, while others may be influenced by the combined credibility of the premises. These differences in evaluation criteria complicate the interpretation of the results and highlight the need for more research to better understand these patterns.

The discussion we have conducted has attempted to highlight the complexities and limitations inherent in studying the dynamics of credibility in reasoning. The variability in the interpretation of credibility values underscores the need for a broader and more nuanced approach to explore these phenomena in future research.

6. Future Research Directions

This section outlines potential research directions arising from this study, starting with ideas generated from discussing the limitations and extending to new areas that deepen the understanding of the cognitive dynamics of belief systems.

6.1. Methodological Improvements and Exploration of Epistemic Communities in Credibility Research

Increasing Items in Conditions of Low Credibility: Although the stability and statistical power of our analyses were generally adequate, it was observed that in conditions of low variability, such as the LL condition, the patterns were not clear. To address this issue, it is recommended to increase the number of items in these specific conditions. This would facilitate a more precise detection of patterns through regression analysis, improving the robustness of our findings.

Experimentation with Different Epistemic Communities: The belief system functions as an interconnected network of informational units in memory, causing the credibility of a statement to vary among different audiences. We propose conducting studies that use various epistemic communities to examine how pre-existing beliefs influence the acceptance of new information. This could involve providing arguments with *premises* that are accepted by one group but not by another, allowing for significant differences in the acceptance of conclusions to be observed. Such an approach would have direct applications in areas such as propaganda and advertising, tailoring messages to the target audience. Additionally, the role of prior credibility on posterior credibility, which was not manipulated as an independent variable in previous studies, should be explored more deeply. Systematically manipulating the *prior credibility of conclusions*, using groups with different baseline values or assumptions, such as environmentalists versus motor enthusiasts, could

provide new insights into how initial credibility influences the final credibility assigned to the same conclusion.

Exploration of Differences in the Impact of Premises: An intriguing finding has been the observed difference between the HL and LH conditions. Future studies could focus on determining whether this difference is consistent and what causes it. We suggest as a possibility that it could be due to the difference in the nature of both types of premises, one being the guarantor (i.e., a general rule) and the other the data premise (i.e., a more particular event). Uncovering the causes behind this phenomenon could offer fundamental insights into the functional differences of the system structures and how people process and integrate different types of information into their belief systems.

6.2. Exploration of Diversity in Argument Structures and Argumentation Schemes

This study has focused exclusively on syllogisms of the first figure, considered canonical and relatively straightforward. However, in everyday life and common reasoning, a variety of argumentative structures are employed which can significantly influence the credibility assigned to ideas. The third figure syllogism, for example, where the middle term has a lesser degree of generality, presents distinct challenges that can alter both the interpretation and acceptance of conclusions. Additionally, the use of quantifiers and negations in various argumentative structures deserves detailed exploration, as these variations could reveal significant differences in how conclusions are processed and accepted. A study that includes these variations, especially in the use of quantifiers and negations, could provide a deeper understanding of credibility perception.

Additionally, topical schemes have been discussed throughout this thesis, which function as structural templates facilitating binary connections between relevant variables in a conditional form. While the simpler schemes have been studied, arguments a fortiori and arguments of contraries present additional complexities that could impact how conclusions are evaluated and accepted. For example, an a fortiori argument that suggests, 'if a less likely event occurs, then a more likely one should also occur,' calls for a detailed review of the underlying credibility computation, as multiplication may not be applicable in this case. Moreover, it would also be interesting to systematically examine the presence or absence of modal qualifiers like 'probably,' 'certainly,' etc., before the sentences to be evaluated by participants, as this could affect the perceived credibility.

Investigating these variations in argumentation schemes and argument formulation could provide valuable insights for fields such as persuasion, informal logic, and argumentation theory. It is crucial to delve deeper into these areas to identify common patterns and differences in the processing of these schemes, which could enrich our understanding of the dynamics of beliefs and enhance our ability to design more effective and persuasive arguments. This integrative approach expands the scope of the study of credibility in different types of arguments and syllogistic structures and also opens new directions for research into the interaction between the form of the arguments and their persuasive effectiveness.

6.3. Approach to Refutation

The approach to refutation offers ample opportunities for future research, especially in developing a comprehensive psychological theory that explores how beliefs are challenged and modified within cognitive psychology, as well as in the practice of argumentation in debates and public discourse. In the theoretical model presented in this thesis, a distinction was made between justification and refutation relations within the belief system, highlighting the influence of refuters in determining the credibility of representations (e.g., De Neys et al., 2002, 2003ab; Geiger, & Oberauer, 2007; Quinn & Markovits, 1998). However, the process of how a belief can be refuted was not deeply explored, despite this process being essential for understanding how beliefs are modified or abandoned.

Refuting a belief involves more than merely reversing the acceptance process; it requires presenting compelling arguments to make an individual stop believing in something previously accepted. This is complicated by the interconnection of beliefs within the system, which form part of our worldview. Additionally, as Schröder et al. (2008) pointed out, once a belief is validated and integrated, it is rarely re-evaluated unless significant contradictions arise. Various conceptual and methodological aspects presented in this thesis could be combined with studies on refutation in discourse, stemming from the framework of Knowledge Revision Components (e.g., Kendeou, Smith, & O'Brien, 2014; Kim & Kendeou, 2021; Lassonde, Kendeou, & O'Brien, 2016), establishing direct links between studies of reasoning, argumentation, and discourse processing. Undoing an entrenched belief could be more complex and involve significant emotional or identity costs. Emotional resistance to changing beliefs is a well-documented phenomenon in social psychology (Knowles & Linn, 2004; Pfau, Szabo, Anderson, Morrill, Zubric, & H-Wan, 2001) and understanding the role of cognitive dissonance in this process could be crucial.

6.4. Approach to Emotion

Emotions play a crucial role both in the formation and refutation of beliefs. Not only do our emotions affect the interpretation of everyday events—for instance, an anxious person may perceive neutral messages as threatening, influencing their beliefs about others' intentions—but they also influence our future expectations, biasing our beliefs towards optimism or pessimism (Calvo, Eysenck, & Castillo, 1997; Gutiérrez-García, & Calvo, 2017; Miceli, & Castelfranchi, 2014). This phenomenon underscores the importance of studying pathos in classical rhetoric. Furthermore, the esteem and trust we hold towards an information source significantly impact our willingness to accept their claims and resist the refutation of previously established beliefs (Van Boekel, Lassonde, O'Brien, & Kendeou, 2017), highlighting the importance of ethos in persuasive communication. Even motives for social affiliation can lead to the adoption of beliefs without the presentation of considered solid arguments, merely to align with a reference group (Iyer, & Krishnan, 2002; Pandelaere, 2022).

Thus, a proposal for future studies would be to explore how the mechanisms of the belief system affect the dynamics of emotions and vice versa. According to theories addressed in App. C,

enthymemes and argumentation schemes not only convey credibility but may also transfer the hedonic valence from the premises to the conclusion. For instance, an initially neutral statement could acquire a positive or negative emotional connotation depending on its presentation in the argument. This might manifest in the subject of a conclusion acquiring a certain emotional charge derived from the hedonic tone of the predicate, provided the latter is previously valued as pleasant or unpleasant.

Considering a hypothetical scenario, if a speaker wishes to induce a negative attitude towards "medical advancements" within a community that values them positively, they could attempt to associate these advancements with negative aspects, such as an increase in diseases, proposing that such advancements will result in undesirable consequences. If the audience accepts this connection, the emotional valence towards "medical advancements" could shift from positive to negative.

This approach suggests that altering beliefs about a subject can directly change its emotional perception, offering a fertile ground to investigate the interplay between credibility and emotion. The foundations of these ideas can be traced back to the theories of Fishbein (1963; Fishbein & Ajzen, 1975), which link the attitude towards an object with the belief that it possesses certain emotionally valued attributes.

6.5. Approaching the Dynamics of Credibility

Research into Validation and Integration Processes of Credibility: A promising line of research focuses on the processes of validation and integration of credibility within the realm of discourse processing and comprehension. Although there are advanced studies on epistemic validation in sentence processing (e.g., Isberner & Richter, 2014a; Richter & Maier, 2018; Schroeder et al., 2008, Singer, 2019), research on argument processing remains limited. This work proposes using the theoretical and methodological apparatus of these disciplines to explore the processing of credibility. The manipulation of the credibility of premises by modifying the middle term allows the linguistic formulation of the conclusions to be identical across the four credibility conditions of the argument. This would facilitate further studies to explore online processes in credibility calculations, allowing precise comparisons of reading times between conclusions and specific parts of the premises.

Implicitness and Explicitness of Premises in Enthymemes: Another intriguing approach is the study of the enthymeme from its dual aspect as a rhetorical argument aimed at changing credibilities and as an argument with an implicit premise. Manipulating the credibility of premises, along with varying their degree of implicitness or explicitness, offers a fertile field for research. This approach would advance understanding of the belief system model presented. Moreover, it would contribute to the psychology of reasoning and the processes of discourse validation. The central focus of this research would be the inference of implicit content and the computation and integration of its credibility within the context of the argument. This type of study, which would

directly connect reasoning and discourse studies, would be highly informative about the structure of the belief system in memory.

Relation to Memory and Processing Automation: Closely related to the previous points, these studies could also delve into how the belief system interacts with long-term memory and working memory, assessing whether credibility processes are automatic and routine, or whether they are central and conscious processes. Additionally, it would be relevant to investigate whether certain contexts promote a mindset that facilitates the automatic processing of credibility, similar to what occurs with the processing of figurative language observed in metaphor research (e.g., Gibbs, 1994). This would link to some relevant studies in the field of discourse validation (Isberner & Richter, 2014b).

Multimodal Modalities in Belief Research: Finally, considering that beliefs are formed on multimodal mental representations and not merely on propositions in their traditional sense, it would be innovative to conduct experiments exploring belief change using different modalities, such as images or music. This multimodal approach could unveil new insights on how various forms of stimulus influence the reasoning involved in belief modification, broadening the spectrum of research into the dynamics of credibility and its sources of formation.

This outline for future research lines in the dynamics of beliefs continues with the work presented in this thesis and also opens new directions that can significantly enrich the field of credibility study and the psychology of beliefs.

7. Conclusion

The thesis discussion deeply addresses the interaction of beliefs, credibility, and belief systems from a cognitive perspective, emphasising their argumentative nature and the connections with the psychology of reasoning, discourse processing, and traditional rhetoric and argumentation. Through the experiments presented here, key aspects of the theoretical model were validated, demonstrating how argumentation techniques directly affect credibility and the modification of beliefs in the deductive aspect of the system, reinforcing the theoretical framework and highlighting the adaptability and dynamism of belief systems in response to new arguments, evidence, or approaches. The proposed Subjectivity – Adherence – Representation schema stands out for presenting a renewed vision of belief systems, emphasising their adaptability and dynamism in the face of new information. This model describes subjectivity as the starting point of the system, which stores beliefs and actively modifies its mental representations, in a bidirectional exchange with the perception and interpretation of information. The mental representations, which must be coherent and justified, form an interconnected network through relationships of justification, essential for the maintenance of the cognitive system.

The model articulates a balance between stability and flexibility, distinguishing between core beliefs (whether foundational or schematic), which provide a relatively stable framework for the interpretation of new information, and other more circumstantial beliefs, therefore more susceptible to changes and revisions. This balance is vital for the relevance and adaptability of the system, allowing adjustments in response to new contradictions or data, and highlights the importance of inductive and deductive processes in maintaining the system's consistency, coherence, and adaptation. The S – A – R schema has significantly advanced our understanding of the structure and function of belief systems, showing how beliefs interact with and adapt to information in a world where there is always something changing and something remaining constant.

This approach emphasises the need to regard beliefs as fundamental phenomena for human cognition. On one hand, the justification of stable beliefs in space and time provides security to the individual for decision-making and action, offers predictive capacity, and thus, a sense of understanding and control over the environment. These types of beliefs are essential for not getting lost in diversity, and for understanding the world, and oneself within it, as an integrated unit, with coherence and continuity. On the other hand, beliefs concerning specific issues are changeable and adapted to specific contexts, often linked to more general and stable beliefs. These particular beliefs are more dynamic and variable, and often serve as informational inputs for the more general ones. This thesis seeks to enrich academic literature by establishing a starting point for future research that could further explore the stability and dynamism in the adaptability and functions of beliefs in various cognitive and social interactions.

The discussion in this thesis has delved into the complex interaction between belief systems in human cognition from a multidisciplinary perspective, encompassing aspects of the psychology of reasoning, discourse processing, and rhetorical argumentation. The experiments conducted validate key aspects of a theoretical model focused on the dynamics of belief systems, rather than their stability. These experiments have addressed how argumentation with enthymemes influences the credibility and modification of credibilities. This approach has developed a precise understanding of argumentation with enthymemes through probabilistic models. The discussion also critically reviews the Rule of Theophrastus, proposing its adaptation to reflect the necessary flexibility in contemporary reasoning logic. Furthermore, the research has effectively integrated quantitative and qualitative approaches in an innovative methodology for the empirical study of beliefs, enhancing the understanding of the configuration and justification of credibility within belief systems. This methodology provides a solid foundation for the replication and verification of findings, strengthening control over the variables involved in the computation of credibility.

The section on Future Research Directions in this thesis discussion addresses the need to expand and improve the methodologies for studying the dynamics of beliefs, emphasising how different epistemic communities interpret and accept information based on their prior beliefs. It proposes increasing the number of items in studies with low variability to enhance analytical

precision, as well as exploring the impact of initial credibility on conclusions through the manipulation of a priori credibility. Additionally, it suggests investigating the differences in the impact of data premises versus warrant premises and expanding the study on the diversity of argumentative structures to understand how they affect the perception and acceptance of credibility.

It is also proposed to develop a model of the mechanisms involved in refutation to understand how beliefs are dismantled within interconnected systems. This approach will consider the emotional cost of changing or discarding beliefs, using cognitive dissonance as a theoretical framework. It is recommended to examine the role of emotions and esteem towards information sources in the formation and refutation of beliefs. Finally, it suggests delving deeper into the processes of validation and integration of credibility in discourse through advanced tools from psycholinguistic studies, addressing the interaction of cognitive, emotional, and argumentative aspects in the study of beliefs.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

APÉNDICE A: EXPERIMENTO 1

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

Appendix A-I – Warrant Generation - Sample from one of the Questionnaires

A1

EDAD:

Instrucciones: en las siguientes páginas encontrarás la parte inicial de una frase en la que aparece una causa, y se te pide que la completes con todos los efectos que se te ocurran; por ejemplo:

‘Si [causa], entonces [¿?]’.

Por favor, completa la frase haciendo una lista de todos los efectos que se te ocurran para cada una de las causas que se te presentan. No emplees demasiado tiempo para cada una de ellas (un minuto que será señalado por el investigador) procurando no dar distintas versiones de una misma idea y que lo que dices sea de carácter general y realista. A continuación, se te presenta un ejemplo:

Escribe todas las consecuencias (efectos) que se te ocurran que podrían derivarse de la siguiente situación (o causa):

CAUSA: ‘Si la contaminación marina continúa, entonces...’.

EFFECTOS:

- Las ballenas desaparecerán.
- El pescado que comemos también estará contaminado.
- Dejaremos de bañarnos en el mar.
- La vida marina desaparecerá.
- El mar se quedará sin oxígeno.

Procura **no dar respuestas del tipo:*

CAUSA: ‘Si la contaminación marina continúa, entonces...’.

EFFECTOS:

- Las ballenas morirán.
 - Muchos peces morirán.
 - Muchos animales morirán.
- } Por ser variaciones de una misma
-
- Mi abuela ya no se meterá en el agua.
 - El delfín Flipper será desgraciado.
- } Por ser casos muy concretos y
-
- El mar se revelará contra nosotros en forma de Tsunami.
 - Seremos castigados por Neptuno.
 - Nos iremos a Marte a empezar de nuevo.
- } Por no ser

MUCHAS GRACIAS POR TU COLABORACIÓN

Por favor, escribe todas las consecuencias (efectos) que se te ocurran que podrían derivarse de las siguientes situaciones (o causas):

CAUSA: Si se extiende el libre mercado, *entonces...*

EFFECTOS:

Se les dejaba un espacio para contestar como este. Había siete ítems por página.

CAUSA: Si se encuentra la cura para el cáncer, *entonces...*

EFFECTOS:

CAUSA: Si se preservan los hábitats naturales, *entonces...*

EFFECTOS:

CAUSA: Si aumenta el consumo de masas, *entonces...*

EFFECTOS:

CAUSA: Si se endurece la ley del menor, *entonces...*

CAUSA: Si te quemas mucho al sol, *entonces...*

EFFECTOS:

CAUSA: Si disminuye la tasa de nacimientos de un país, *entonces...*

EFFECTOS:

CAUSA: Si las universidades bajan su nivel de exigencia, *entonces...*

EFFECTOS:

CAUSA: Si alguien fuma muchos porros, *entonces...*

EFFECTOS:

CAUSA: Si aumentan las zonas de construcción o 'edificables', *entonces...*

EFFECTOS:

CAUSA: Si aumenta la población del mosquito tigre, *entonces...*

EFFECTOS:

CAUSA: Si no se actúa pensando en el futuro, *entonces...*

EFFECTOS:

CAUSA: Si aumenta la obesidad infantil, *entonces...*

EFFECTOS:

CAUSA: Si aumenta el número de nacimientos no deseados, *entonces...*

EFFECTOS:

CAUSA: Si en una sociedad aumenta la proporción de hombres respecto a la de mujeres, *entonces...*

EFFECTOS:

CAUSA: Si los supermercados empiezan a emplear bolsas reutilizables o de papel reciclado, *entonces...*

EFFECTOS:

CAUSA: Si disminuye el uso de preservativos, *entonces...*

EFFECTOS:

CAUSA: Si aumenta la diferencia entre ricos y pobres, *entonces...*

EFFECTOS:

CAUSA: Si se ratifica el protocolo de Kyoto, *entonces...*

EFFECTOS:

CAUSA: Si se permite que las jóvenes de 16 años aborten sin conocimiento paterno, *entonces...*

EFFECTOS:

CAUSA: Si aumentan los puestos de trabajo, *entonces...*

EFFECTOS:

CAUSA: Si te pasas todo el día delante de una pantalla (ordenador, televisión, etc.), *entonces...*

EFFECTOS:

CAUSA: Si en un país desarrollado desciende el número de nacimientos, *entonces...*

EFFECTOS:

CAUSA: Si se abusa del empleo de papel, *entonces...*

EFFECTOS:

CAUSA: Si se produce el trasvase del Ebro, *entonces...*

EFFECTOS:

B1

EDAD:

Instrucciones: en las siguientes páginas encontrarás la parte inicial de una frase en la que aparece una causa, y se te pide que la completes con todos los efectos que se te ocurran; por ejemplo:

‘Si [causa], entonces [¿?].’

Por favor, completa la frase haciendo una lista de todos los efectos que se te ocurran para cada una de las causas que se te presentan. No emplees demasiado tiempo para cada una de ellas (un minuto que será señalado por el investigador) procurando no dar distintas versiones de una misma idea y que lo que dices sea de carácter general y realista. A continuación, se te presenta un ejemplo:

Escribe todas las consecuencias (efectos) que se te ocurran que podrían derivarse de la siguiente situación (o causa):

CAUSA: ‘Si la contaminación marina continúa, entonces...’.

EFFECTOS:

- Las ballenas desaparecerán.
- El pescado que comemos también estará contaminado.
- Dejaremos de bañarnos en el mar.
- La vida marina desaparecerá.
- El mar se quedará sin oxígeno.

Procura **no dar respuestas del tipo:*

CAUSA: ‘Si la contaminación marina continúa, entonces...’.

EFFECTOS:

- Las ballenas morirán.
 - Muchos peces morirán.
 - Muchos animales morirán.
- } Por ser variaciones de una misma
-
- Mi abuela ya no se meterá en el agua.
 - El delfín Flipper será desgraciado.
- } Por ser casos muy concretos y
-
- El mar se revelará contra nosotros en forma de Tsunami.
 - Seremos castigados por Neptuno.
 - Nos iremos a Marte a empezar de nuevo.
- } Por no ser

MUCHAS GRACIAS POR TU COLABORACIÓN

Por favor, escribe todas las consecuencias (efectos) que se te ocurran que podrían derivarse de las siguientes situaciones (o causas):

CAUSA: Si te tomas a la ligera tus estudios, *entonces...*

EFFECTOS:

CAUSA: Si uno malgasta alimentos, *entonces...*

EFFECTOS:

CAUSA: Si se aprueba la ley del aborto, *entonces...*

EFFECTOS:

CAUSA: Si se es madre primeriza a los 40 años, *entonces...*

EFFECTOS:

CAUSA: Si alguien se droga frecuentemente, *entonces...*

EFFECTOS:

CAUSA: Si aumenta la inmigración, *entonces...*

EFFECTOS:

CAUSA: Si se extiende el uso del coche eléctrico, *entonces...*

EFFECTOS:

CAUSA: Si se mantiene la quema de combustibles fósiles, *entonces...*

EFFECTOS:

CAUSA: Si se incrementan los planes de obras públicas de gran envergadura, *entonces...*

EFFECTOS:

CAUSA: Si la mayor parte del tiempo de ocio se emplea en jugar a la consola, al ordenador y en los chats, *entonces...*

EFFECTOS:

CAUSA: Si uno fuma mucho, *entonces...*

EFFECTOS:

CAUSA: Si se suben los impuestos, *entonces...*

EFFECTOS:

CAUSA: Si aumentan las zonas edificables y la destrucción de parajes naturales en España, *entonces...*

EFFECTOS:

CAUSA: Si, entre la gente, aumenta el sentimiento de soledad, *entonces...*

EFFECTOS:

CAUSA: Si cada vez se ve más telebasura, *entonces...*

EFFECTOS:

CAUSA: Si aumenta la población en paro, *entonces...*

EFFECTOS:

CAUSA: Si aumenta el uso de papel reciclado, *entonces...*

EFFECTOS:

CAUSA: Si se les pone ruedas a las mochilas escolares, *entonces...*

EFFECTOS:

CAUSA: Si se siguen emitiendo gases invernadero, *entonces...*

EFFECTOS:

CAUSA: Si disminuye la natalidad y aumenta el número de receptores de ayudas sociales, *entonces...*

EFFECTOS:

CAUSA: Si aumenta la proporción de hombres que no encuentran pareja, *entonces...*

EFFECTOS:

CAUSA: Si la población de un país envejece, *entonces...*

EFFECTOS:

CAUSA: Si continúa la demanda de abrigos de piel, *entonces...*

EFFECTOS:

CAUSA: Si la contaminación de los ríos continúa, *entonces...*

EFFECTOS:

CAUSA: Si aumenta la libertad sexual entre adolescentes, *entonces...*

EFFECTOS:

Appendix A-II – Pretest Phase: Evaluation of Isolated Conclusions

Experimental Items Group 1 (G1) – Control Group 1 (C1)

1. El trabajo que se hace en la universidad logrará que muchas mujeres se sientan esperanzadas.
2. Las empresas automovilísticas contribuirán a mejorar el medio ambiente.
3. Los estereotipos sociales presentes en ciertos programas de televisión pueden contribuir al aumento de la homosexualidad.
4. Una importante proporción de la juventud española de hoy sufrirá daños cerebrales y psicológicos mañana.
5. El reciclaje contribuirá fuertemente a fomentar el turismo de montaña y rural.
6. Los partidos políticos progresistas facilitarán una mayor diversidad cultural en esta sociedad.
7. Una importante proporción de los niños de hoy podrá tener ataques de epilepsia mañana.
8. El PP destruirá muchos bosques y espacios de valor ecológico.
9. El parlamento español producirá indirectamente una disminución de la natalidad.
10. La ley del aborto producirá un aumento de enfermedades de transmisión sexual.
11. En nuestra sociedad aumentarán las muertes por alcoholismo, drogadicción y suicidios.
12. En los países ricos la población envejecerá.
13. En la sociedad española puede aumentar el número de familias con un solo hijo.
14. En China bajará el índice de matrimonios.
15. En la sociedad española claramente aumentarán las campañas en favor de hábitos alimentarios saludables.
16. China es un país que se verá en graves problemas económicos.
17. En nuestra sociedad aumentará el número de niños con la columna desviada.
18. Los niños y adolescentes de nuestra sociedad gozarán de relaciones sociales de menor calidad y más soledad.
19. El Partido Popular reducirá la producción de arroz en el conjunto del Estado.
20. La ley del aborto aumentará la probabilidad de que los niños sean más felices.
21. Los supermercados Carrefour contribuirán a la disminuir la contaminación a largo plazo.
22. Los medios de transporte contribuirán a la extinción de los osos polares.
23. Con el nuevo sistema universitario de grados los alumnos verán menos materia y sabrán menos que antes.
24. En verano mucha gente pasa frío.
25. Muchas mujeres de esta sociedad de consumo perderán peso en un futuro próximo.
26. Los políticos son presumidos.
27. En verano la gente se rasca más de lo normal.
28. El coral dispone de algún mecanismo respiratorio.
29. Muchos niños tienen (y tendrán) calambrazos.
30. Valentino Rossi afrontará su día a día con optimismo y energía.
31. Marcas de ropa deportiva como Adidas o Nike aumentarán sus ventas respecto a otras marcas.

Filler Items of High Credibility Group 1 (G1) – Control Group 1 (C1)

1. El PSOE ganó las últimas elecciones al Gobierno de España.
2. En nuestra sociedad aumentará la edad de jubilación a 67 años de edad.
3. El descubrimiento oficial de América fue el 12 de octubre de 1492.
4. Las manzanas caen de los árboles debido al efecto de la gravedad.
5. Las personas que fuman aspiran humo de sus cigarros.
6. Un cuadrado tiene cuatro lados iguales unidos por esquinas de 90 grados.
7. La sociedad mediterránea se caracteriza por ser muy alegre y abierta.
8. Al meter una cerveza en el congelador durante más de 4 horas, se congelará y se estropeará.
9. De aquí a dos años muchos países del continente africano seguirán siendo pobres.
10. El año que viene la Sagrada Familia de Barcelona seguirá estando en obras.
11. La industria de la automoción fabricará coches que sean cada vez más seguros.
12. Varios islotes del archipiélago de las Baleares terminarán siendo de propiedad inglesa o alemana.
13. En algún momento de la historia de la humanidad, EE.UU. dejará de tener la hegemonía económica mundial.
14. La mayoría de grandes mamíferos salvajes se habrán extinguido ya en el siglo XXII.
15. Las reservas petrolíferas se agotarán en menos de 200 años.
16. Durante los próximos dos años Oriente Medio seguirá en conflictos armados.
17. En otoño del 2011 los árboles caducifolios perderán sus hojas.
18. En un futuro próximo todas las películas que se expongan en los cines serán en 3-D.
19. Aunque prohíban las páginas web de descargas ilegales (música, películas, etc.), el impuesto del canon digital seguirá existiendo.
20. Algún día el Sol dejará de existir.

Experimental Items Group 2 (G) – Control Group 2 (C2)

1. El trabajo que se hace en la universidad salvará la vida de muchas mujeres.
2. Las empresas automovilísticas contribuirán a la disminución del consumo de petróleo y carburante.
3. Los estereotipos sociales de los medios de comunicación pueden contribuir al aumento de embarazos no deseados.
4. Una importante proporción de la juventud española de hoy tendrá un peor rendimiento académico o laboral mañana.
5. El reciclaje contribuirá fuertemente a salvar la vida de muchos animales salvajes.
6. Los partidos políticos progresistas facilitarán un aumento del número de delitos.
7. Una importante proporción de los niños de hoy acabará con problemas visuales.
8. El PP disminuirá el precio de la vivienda.
9. El parlamento español producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.
10. La ley del aborto producirá un crecimiento del número de embarazos.
11. En nuestra sociedad aumentarán las depresiones.
12. En los países ricos muchos colegios tendrán que cerrar.
13. En la sociedad española puede aumentar el número de recién nacidos con deficiencias.
14. En China aumentará la prostitución y los crímenes sexuales.
15. En la sociedad española claramente aumentará el riesgo de padecer enfermedades cardiovasculares.
16. En España aumentará la proporción de personas con enfermedades degenerativas.

17. En nuestra sociedad disminuirán los problemas infantiles de espalda.
18. Los niños y adolescentes de nuestra sociedad perderán habilidades de creatividad e imaginación.
19. El Partido Popular contribuirá a la destrucción de las reservas naturales del país.
20. La ley del aborto disminuirá las rupturas familiares.
21. Los supermercados Carrefour dañarán al sector comercial del plástico.
22. Los medios de transporte contribuirán a la destrucción de los polos.
23. Con el nuevo sistema universitario de grados los alumnos tendrán más tiempo de ocio y diversión que antes.
24. En verano mucha gente pasa frío.
25. Muchas mujeres de esta sociedad de consumo perderán peso en un futuro próximo.
26. Los políticos son presumidos.
27. En verano la gente se rasca más de lo normal.
28. El coral dispone de algún mecanismo respiratorio.
29. Muchos niños tienen (y tendrán) calambrazos.
30. Valentino Rossi afrontará su día a día con optimismo y energía.
31. Marcas de ropa deportiva como Adidas o Nike aumentarán sus ventas respecto a otras marcas.

Filler Items of High Credibility Group 2 (G2) – Control Group 2 (C2): The filler items used for Group G2 were the same as those used for Group G1.

Appendix A-III – Pretest Phase: Evaluation of Data Premises

Experimental Data Premises Group 1 (G1) and Group 2 (G2)

1. El trabajo que se hace en la universidad puede encontrar la cura del cáncer de mama.
2. Las empresas automovilísticas contribuirán a que la mayoría de la gente use el coche eléctrico.
3. Los estereotipos sociales presentes en ciertos programas de televisión promueven el sexo entre adolescentes.
4. Una importante proporción de la juventud española de hoy se droga pronto y frecuentemente.
5. El reciclaje contribuye fuertemente a preservar los hábitats naturales.
6. Los partidos políticos progresistas facilitan el aumento de la inmigración.
7. Una importante proporción de los niños de hoy pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.).
8. El PP aumentará las zonas de construcción o ‘edificables’.
9. El parlamento español ha aprobado la ley del aborto.
10. La ley del aborto producirá una disminución del uso de preservativos.
11. En nuestra sociedad está aumentando la competitividad y el sentimiento de soledad.
12. En los países ricos está disminuyendo la tasa de natalidad.
13. En la sociedad española está aumentando la proporción de madres primerizas a edades cercanas a los 40 años.
14. En China está aumentando la proporción de hombres que no encuentran pareja debido a que no nacen niñas.
15. En la sociedad española claramente está aumentando la obesidad infantil.
16. China es un país en el que la población está envejeciendo debido a que mantienen un estricto control de la natalidad.
17. En nuestra sociedad se ha extendido el uso de mochilas escolares con ruedas.
18. Los niños y adolescentes de nuestra sociedad emplean el tiempo de ocio en la consola, el ordenador y los chats.
19. El Partido Popular pretende realizar el trasvase del Ebro.
20. La ley del aborto aumenta la probabilidad de que sólo nazcan niños deseados.
21. Los supermercados Carrefour emplean bolsas reutilizables.
22. Los medios de transporte emiten gases invernadero.
23. Con el nuevo sistema universitario de grados ha disminuido el número de créditos académicos respecto del plan anterior.
24. En verano mucha gente enciende el aire acondicionado muy fuerte.
25. Muchas mujeres de esta sociedad de consumo comen alimentos dietéticos.
26. Los políticos cuidan mucho su apariencia física.
27. En verano a la gente le pican los mosquitos.
28. El coral es un ser vivo.
29. Muchos niños meten (y meterán) los dedos en los enchufes.
30. Valentino Rossi desayuna con Nutella.
31. Marcas de ropa deportiva como Adidas o Nike siempre emplean deportistas famosos para anunciar sus productos.

Filler Items of Low Credibility Group 1 (G1) and Group 2 (G2)

1. Los movimientos antisistema reivindican la extensión del capitalismo para los países pobres.
2. La UE pretende que llegue un día en el que sólo se hable inglés en toda Europa.
3. Los estudiantes cada vez se esfuerzan más, están más motivados y sacan mejores notas.
4. La ley del aborto libre ha sido aprobada en la mayoría de los países de Europa.
5. Hablar mucho por el móvil produce cáncer en 2 de cada 10 personas.
6. La ciudad de Tarragona está cerca de convertirse en capital cultural europea.
7. El Imperio Romano se extendía desde la India hasta Escocia.
8. En nuestra sociedad está disminuyendo la violencia de género.
9. Fueron los propios estadounidenses quienes tiraron las Torres Gemelas.
10. Estados Unidos nunca llegó a la Luna, fue todo un montaje.
11. En Italia, a la población le cuesta encontrar pareja.
12. El Fútbol Club Barcelona perderá la liga esta temporada.
13. El Nàstic de Tarragona será campeón de la Champions algún día.
14. Llegará un día en el que el idioma oficial de todas las poblaciones de la costa española sea el inglés.
15. La Tierra dejará de girar sobre sí misma de aquí a tres siglos.
16. El político Carod-Rovira es de origen chino.
17. El tabaquismo es un movimiento cultural destinado a salvar vidas humanas.
18. Los elefantes son animales que se caracterizan por tener grandes alas.
19. La dieta mediterránea es sana porque consiste en la ingesta de choricillos, morcillas, frituras variadas y mucho aceite de oliva.
20. En la época de Navidad los canales reducen la emisión de anuncios comerciales.

Appendix A-IV – Pretest Phase Instructions for Experiment 1

SPANISH VERSION - *Evaluation of Isolated Conclusions Instructions for Experimental Groups*

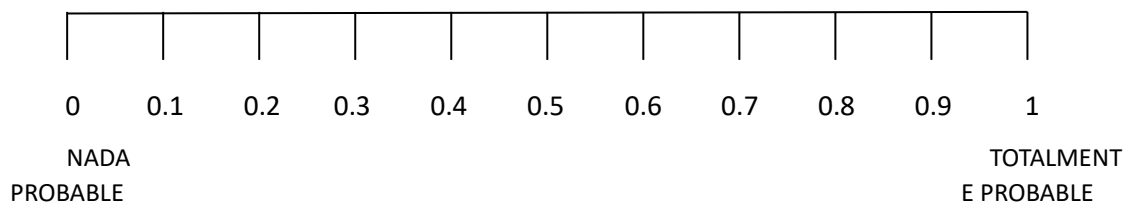
Instrucciones:

A continuación, verás una serie de afirmaciones. Tu tarea consiste en otorgarles una puntuación en una escala de probabilidad de 11 puntos que va de 0 (nada probable) a 1 (totalmente probable). Esta puntuación debe reflejar el grado en el que crees que las situaciones descritas en dichas afirmaciones pueden ser (o son) ciertas.

Ejemplo:

Por favor, estima la probabilidad de ocurrencia de la siguiente situación:

‘El planeta Tierra gira al rededor del Sol.’



Si, por ejemplo, consideras que la probabilidad de que ‘*el planeta Tierra gira al rededor del Sol*’ es totalmente probable, le otorgarás una puntuación de **1**. Para ello, **en el teclado**, deberás pulsar la tecla señalada con un 1.

Después de esto, te aparecerá una pantalla en la que se te indica que para pasar a la siguiente afirmación deberás pulsar la barra espaciadora. Cuando la pulses, te aparecerá otra afirmación diferente para la cual deberás estimar su probabilidad, y así sucesivamente.

Primero realizarás 6 ensayos de prueba para acostumbrarte a la tarea; cuando finalices estos 6 ensayos debes **parar** hasta que el experimentador te lo indique.

Después de haber hecho la mitad de la tarea, verás una pantalla indicándote que debes **parar** y realizar la tarea de lápiz y papel que tienes en tu mesa. Cuando la acabes, descansa en silencio hasta que el experimentador te dé la instrucción de empezar la segunda mitad de la tarea de probabilidad.

Cuando hayas finalizado el experimento, aparecerá una pantalla en negro: es muy importante que no toques nada. El experimentador os dirá en qué momento podéis abandonar la sala.

Muchas gracias por tu colaboración.

ENGLISH VERSION - *Evaluation of Isolated Conclusions Instructions for Experimental Groups*

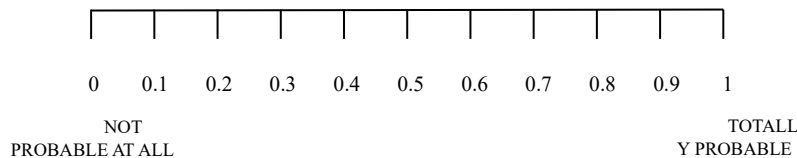
Instructions:

Below you will see a series of statements. Your task is to give them a score on an 11-point probability scale ranging from 0 (not at all likely) to 1 (extremely likely). This score should reflect the degree to which you believe that the situations described in these statements can be (or are) true.

Example:

Please estimate the probability of occurrence of the following situation:

'The planet Earth revolves around the Sun.'



If, for instance, you consider that the probability that *'the planet Earth revolves around the Sun'* is totally probable, you will give it a score of 1. To do this, on the keyboard, you must press the key marked with a 1.

After this, a screen will appear telling you that to move on to the next statement you must press the space bar. When you press it, another different statement will appear for which you must estimate its probability, and so on.

You will first do 6 practice trials to get used to the task; when you finish these 6 trials you must stop until the experimenter tells you to.

After you've done half of the task, you'll see a screen telling you to stop and do the paper-and-pencil task on your desk. When you are done, rest quietly until the experimenter instructs you to start the second half of the probability task.

When you have finished the experiment, a black screen will appear; it is very important that you do not touch anything. The experimenter will tell you when you can leave the room.

Thank you very much for your cooperation.

Appendix A-V – Sample questionnaires for credibility evaluation of isolated conclusions in the pretest and posttest phases by Control Group C1.

Tres últimos números del DNI: __ __ __ Nombre de tu abuela materna: _____

Instrucciones:

A continuación, verás una serie de afirmaciones. Tu tarea consiste en otorgarles una puntuación en una escala de probabilidad de 11 puntos que va de 0 (nada probable) a 1 (totalmente probable). Esta puntuación debe reflejar el grado en el que crees que las situaciones descritas en dichas afirmaciones pueden ser (o son) ciertas.

Ejemplo:

Por favor, estima la probabilidad de ocurrencia de la siguiente situación:

	NADA PROBABLE							TOTALMENTE E PROBABLE				
‘El planeta Tierra gira al rededor del Sol.’	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	

Si, por ejemplo, consideras que la probabilidad de que *‘el planeta Tierra gira al rededor del Sol’* es totalmente probable, le otorgarás una puntuación de **1**. Para ello, deberás fijarte en la casilla del número **1** y rodearlo con un círculo.

Muchas gracias por tu colaboración.

Por favor, estima la probabilidad de ocurrencia de las siguientes situaciones:

NADA
PROBABLE

TOTALMENT
E PROBALBE

											
En un futuro próximo todas las películas que se expongan en los cines serán en 3-D.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Un cuadrado tiene cuatro lados iguales unidos por esquinas de 90 grados.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los niños y adolescentes de nuestra sociedad gozarán de relaciones sociales de menor calidad y más soledad.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En nuestra sociedad aumentará la edad de jubilación a 67 años de edad.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Al meter una cerveza en el congelador durante más de 4 horas, se congelará y se estropeará.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Las personas que fuman aspiran humo de sus cigarros.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En la sociedad española claramente aumentarán las campañas en favor de hábitos alimentarios saludables.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En nuestra sociedad aumentará el número de niños con la columna desviada.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En China bajará el índice de matrimonios.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En verano la gente se rasca más de lo normal.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Valentino Rossi afrontará su día a día con optimismo y energía.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El año que viene la Sagrada Familia de Barcelona seguirá estando en obras.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La ley del aborto producirá un aumento de enfermedades de transmisión sexual.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
De aquí a dos años muchos países del continente africano seguirán siendo pobres.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Las manzanas caen de los árboles debido al efecto de la gravedad.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los políticos son presumidos.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La ley del aborto aumentará la probabilidad de que los niños sean más felices.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

	NADA PROBABLE							TOTALMENTE PROBABLE			
La industria de la automoción fabricará coches que sean cada vez más seguros.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los partidos políticos progresistas facilitarán una mayor diversidad cultural en esta sociedad.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En la sociedad española puede aumentar el número de familias con un solo hijo.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En verano la mucha gente pasa frío.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Algún día el Sol dejará de existir.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Las empresas automovilísticas contribuirán a mejorar el medio ambiente.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los supermercados Carrefour contribuirán a la disminuir la contaminación a largo plazo.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Las reservas petrolíferas se agotarán en menos de 200 años.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Varios islotes del archipiélago de las Baleares terminarán siendo de propiedad inglesa o alemana.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Durante los próximos dos años Oriente Medio seguirá en conflictos armados.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Marcas de ropa deportiva como Adidas o Nike aumentarán sus ventas respecto a otras marcas.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La sociedad mediterránea se caracteriza por ser muy alegre y abierta.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En los países ricos la población envejecerá.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La mayoría de grandes mamíferos salvajes se habrán extinguido ya en el siglo XXII.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Una importante proporción de la juventud española de hoy sufrirá daños cerebrales y psicológicos mañana.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El coral dispone de algún mecanismo respiratorio.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
China es un país que se verá en graves problemas económicos.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

	NADA PROBABLE							TOTALMENTE PROBABLE			
Con el nuevo sistema universitario de grados los alumnos verán menos materia y sabrán menos que antes.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Muchas mujeres de esta sociedad de consumo perderán peso en un futuro próximo.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los medios de transporte contribuirán a la extinción de los osos polares.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El descubrimiento oficial de América fue el 12 de Octubre de 1492.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En nuestra sociedad aumentarán las muertes por alcoholismo, drogadicción y suicidios.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Aunque prohíban las páginas web de descargas ilegales (música, películas, etc.), el impuesto del canon digital seguirá existiendo.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El reciclaje contribuirá fuertemente a fomentar el turismo de montaña y rural.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El parlamento español producirá indirectamente una disminución de la natalidad.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En otoño del 2011 los árboles caducifolios perderán sus hojas.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Muchos niños tienen (y tendrán) calambrazos.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En algún momento de la historia de la humanidad, EE.UU. dejará de tener la hegemonía económica mundial.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Una importante proporción de los niños de hoy podrá tener ataques de epilepsia mañana.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los estereotipos sociales presentes en ciertos programas de televisión pueden contribuir al aumento de la homosexualidad.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El PSOE ganó las últimas elecciones al Gobierno de España.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El Partido Popular reducirá la producción de arroz en el conjunto del Estado.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El trabajo que se hace en la universidad logrará que muchas mujeres se sientan esperanzadas.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El PP destruirá muchos bosques y espacios de valor ecológico.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Sample questionnaires for credibility evaluation of isolated conclusions in the pretest and posttest phases by Control Group C2.

Tres últimos números del DNI: __ __ __ Nombre de tu abuela materna: _____

Instrucciones:

A continuación verás una serie de afirmaciones. Tu tarea consiste en otorgarles una puntuación en una escala de probabilidad de 11 puntos que va de 0 (nada probable) a 1 (totalmente probable). Esta puntuación debe reflejar el grado en el que crees que las situaciones descritas en dichas afirmaciones pueden ser (o son) ciertas.

Ejemplo:

Por favor, estima la probabilidad de ocurrencia de la siguiente situación:

	NADA PROBABLE										TOTALMENT E PROBABLE
‘El planeta Tierra gira al rededor del Sol.’	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Si, por ejemplo, consideras que la probabilidad de que ‘*el planeta Tierra gira al rededor del Sol*’ es totalmente probable, le otorgarás una puntuación de **1**. Para ello, deberás fijarte en la casilla del número **1** y rodearlo con un círculo.

Muchas gracias por tu colaboración.

Por favor, estima la probabilidad de ocurrencia de las siguientes situaciones:

NADA
PROBABLE

TOTALMENT
E PROBALBE

											
El parlamento español producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En la sociedad española puede aumentar el número de recién nacidos con deficiencias.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La ley del aborto disminuirá las rupturas familiares.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El PSOE ganó las últimas elecciones al Gobierno de España.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Valentino Rossi afrontará su día a día con optimismo y energía.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El Partido Popular contribuirá a la destrucción de las reservas naturales del país.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El trabajo que se hace en la universidad salvará la vida de muchas mujeres.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Muchas mujeres de esta sociedad de consumo perderán peso en un futuro próximo.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Las reservas petrolíferas se agotarán en menos de 200 años.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Durante los próximos dos años Oriente Medio seguirá en conflictos armados.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El coral dispone de algún mecanismo respiratorio.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los supermercados Carrefour dañarán al sector comercial del plástico.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En España aumentará la proporción de personas con enfermedades degenerativas.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La sociedad mediterránea se caracteriza por ser muy alegre y abierta.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Marcas de ropa deportiva como Adidas o Nike aumentarán sus ventas respecto a otras marcas.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
De aquí a dos años muchos países del continente africano seguirán siendo pobres.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los medios de transporte contribuirán a la destrucción de los polos.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

	NADA PROBABLE							TOTALMENTE PROBABLE			
En algún momento de la historia de la humanidad, EE.UU. dejará de tener la hegemonía económica mundial.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los partidos políticos progresistas facilitarán un aumento del número de delitos.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Una importante proporción de la juventud española de hoy tendrá un peor rendimiento académico o laboral mañana.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El descubrimiento oficial de América fue el 12 de Octubre de 1492.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En nuestra sociedad aumentará la edad de jubilación a 67 años de edad.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El año que viene la Sagrada Familia de Barcelona seguirá estando en obras.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Una importante proporción de los niños de hoy acabará con problemas visuales.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En los países ricos muchos colegios tendrán que cerrar.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Con el nuevo sistema universitario de grados los alumnos tendrán más tiempo de ocio y diversión que antes.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En verano la mucha gente pasa frío.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Las empresas automovilísticas contribuirán a la disminución del consumo de petróleo y carburante.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En nuestra sociedad aumentarán las depresiones.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La ley del aborto producirá un crecimiento del número de embarazos.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Las personas que fuman aspiran humo de sus cigarros.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En China aumentará la prostitución y los crímenes sexuales.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En otoño del 2011 los árboles caducifolios perderán sus hojas.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El reciclaje contribuirá fuertemente a salvar la vida de muchos animales salvajes.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

	NADA PROBABLE							TOTALMENTE PROBABLE			
Los políticos son presumidos.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los niños y adolescentes de nuestra sociedad perderán habilidades de creatividad e imaginación.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La mayoría de grandes mamíferos salvajes se habrán extinguido ya en el siglo XXII.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Muchos niños tienen (y tendrán) calambrazos.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Al meter una cerveza en el congelador durante más de 4 horas, se congelará y se estropeará.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Varios islotes del archipiélago de las Baleares terminarán siendo de propiedad inglesa o alemana.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
La industria de la automoción fabricará coches que sean cada vez más seguros.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
El PP disminuirá el precio de la vivienda.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En la sociedad española claramente aumentará el riesgo de padecer enfermedades cardiovasculares.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Un cuadrado tiene cuatro lados iguales unidos por esquinas de 90 grados.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Algún día el Sol dejará de existir.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En nuestra sociedad disminuirán los problemas infantiles de espalda.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Aunque prohíban las páginas web de descargas ilegales (música, películas, etc.), el impuesto del canon digital seguirá existiendo.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Las manzanas caen de los árboles debido al efecto de la gravedad.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Los estereotipos sociales de los medios de comunicación pueden contribuir al aumento de embarazos no deseados.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En verano la gente se rasca más de lo normal.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
En un futuro próximo todas las películas que se expongan en los cines serán en 3-D.	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Appendix A-VI – Post-test Phase Instructions for the First Task

Warrant Evaluation Sample Questionnaires

It consists of a direct scoring estimation task and a contingency table estimation task (also called the conditional truth table). The order of tasks and items was counterbalanced.

Version for experimental group G1 - Order-A1

Últimos 3 Números de tu DNI:

Nombre de tu abuela materna:

Instrucciones:

A continuación, se te presentarán 12 ítems que consisten en oraciones del tipo: ‘*si pasa X, entonces pasará Y*’. Por ejemplo:

‘Si alguien se droga frecuentemente, entonces sufrirá daños psicológicos y cerebrales’.

Tu tarea consistirá en puntuar cuán probable es que pase lo que se dice en las frases, en una escala que va del 0 al 100%. Por ejemplo:

‘Si alguien se droga frecuentemente, entonces sufrirá daños psicológicos y cerebrales’: 80

Esto quiere decir que *creo* que lo que pone en la frase sucederá con una probabilidad del 80%.

No hay respuestas correctas ni incorrectas, simplemente debes hacer estimaciones en función de lo que tú crees.

ITEMS:

1. Si se promueve el sexo entre adolescentes, se puede contribuir al aumento de la homosexualidad:

2. Si se preservan los hábitats naturales, entonces se fomentará el turismo de montaña y rural: ____
3. Si uno pasa demasiado tiempo delante de una pantalla (ordenador, televisión, etc.), podrá tener ataques de epilepsia: ____
4. Si se aprueba la ley del aborto, entonces se producirá indirectamente una disminución de la natalidad: ____
5. Si en un país la población envejece, entonces se verá en graves problemas económicos (pocos para trabajar y muchas pensiones por pagar): ____
6. Si se emplea el tiempo de ocio en la consola, el ordenador y en los chats, entonces se gozará de relaciones sociales de menor calidad y más soledad: ____
7. Si sólo nacen niños deseados, entonces aumentará la probabilidad de que los niños sean más felices: ____
8. Si los supermercados empiezan a usar bolsas reutilizables, entonces contribuirán a disminuir la contaminación a largo plazo: ____
9. Si disminuye el número de créditos académicos, entonces los alumnos verán menos materia y sabrán menos que antes: ____
10. Si disminuye la tasa de nacimientos de un país, entonces la población envejecerá: ____
11. Si aumenta la proporción de madres primerizas a edades cercanas a los 40 años, entonces puede aumentar el número de familias con un solo hijo: ____
12. Si se extiende el uso mochilas escolares con ruedas, entonces aumentará el número de niños con la columna desviada: ____

Instrucciones:

A continuación, se te presentarán 12 ítems compuestos por 4 posibilidades, por ejemplo:

ITEM 1:

- a) Alguien se droga frecuentemente y sufre daños psicológicos y cerebrales.
- b) Alguien se droga frecuentemente y **no** sufre daños psicológicos ni cerebrales.
- c) Alguien **no** se droga frecuentemente, pero sufre daños psicológicos y cerebrales.
- d) Alguien **no** se droga frecuentemente y **no** sufre daños psicológicos y ni cerebrales.

Estas son todas posibilidades que puede haber, y no más. Por tanto, cada ítem tendrá esta forma que has visto, es decir: a) que pase A y pase B; b) que pase A y no pase B; c) que no pase A y sí pase B; d) que no pase A y no pase B.

Tu tarea consistirá en otorgar a cada una de las 4 posibilidades de cada ítem un porcentaje de probabilidad, de tal manera, que sumadas las cuatro puntuaciones deben resultar en el 100% total de probabilidad.

Por ejemplo:

ITEM 1:

- e) Alguien se droga frecuentemente y sufre daños psicológicos y cerebrales: 55
- f) Alguien se droga frecuentemente y no sufre daños psicológicos ni cerebrales: 10
- g) Alguien no se droga frecuentemente y sufre daños psicológicos y cerebrales: 15
- h) Alguien no se droga frecuentemente y no sufre daños psicológicos y ni cerebrales: 20

100%

En cada página hay ítems por delante y por detrás.

Como sabes, no hay respuestas correctas ni incorrectas, simplemente debes hacer estimaciones en función de lo que tú crees. No mires las puntuaciones que has puesto en el ejercicio anterior.

MUCHAS GRACIAS POR TU COLABORACIÓN

1.

- a) Se promueve el sexo entre adolescentes y se contribuye al aumento de la homosexualidad: ____
- b) Se promueve el sexo entre adolescentes y no se contribuye al aumento de la homosexualidad: ____
- c) No se promueve el sexo entre adolescentes y se contribuye al aumento de la homosexualidad: ____
- d) No se promueve el sexo entre adolescentes y no se contribuye al aumento de la homosexualidad: ____

2.

- a) Se preservan los hábitats naturales y se fomenta el turismo de montaña y rural: ____
- b) Se preservan los hábitats naturales y no se fomenta el turismo de montaña y rural: ____
- c) No se preservan los hábitats naturales y se fomenta el turismo de montaña y rural: ____
- d) No se preservan los hábitats naturales y no se fomenta el turismo de montaña y rural: ____

3.

- a) Uno pasa demasiado tiempo delante de una pantalla (ordenador, televisión, etc.), y tiene ataques de epilepsia: ____
- b) Uno pasa demasiado tiempo delante de una pantalla (ordenador, televisión, etc.), y no tiene ataques de epilepsia: ____
- c) Uno no pasa demasiado tiempo delante de una pantalla (ordenador, televisión, etc.), y tiene ataques de epilepsia: ____
- d) Uno no pasa demasiado tiempo delante de una pantalla (ordenador, televisión, etc.), y no tiene ataques de epilepsia: ____

4.

- a) Se aprueba la ley del aborto, y se produce indirectamente una disminución de la natalidad: ____
- b) Se aprueba la ley del aborto, y no se produce indirectamente una disminución de la natalidad: ____
- c) No se aprueba la ley del aborto, y se produce indirectamente una disminución de la natalidad: ____
- d) No se aprueba la ley del aborto, y no se produce indirectamente una disminución de la natalidad: ____

5.

- a) En un país la población envejece, y entra en graves problemas económicos: ____
- b) En un país la población envejece, y no entra en graves problemas económicos: ____
- c) En un país la población no envejece, y entra en graves problemas económicos: ____
- d) En un país la población no envejece, y no entra en graves problemas económicos: ____

6.

- a) Se emplea el tiempo de ocio en la consola, el ordenador y en los chats, y se goza de relaciones sociales de menor calidad y más soledad: ____
- b) Se emplea el tiempo de ocio en la consola, el ordenador y en los chats, y no se goza de relaciones sociales de menor calidad y más soledad: ____
- c) No se emplea el tiempo de ocio en la consola, el ordenador y en los chats, y se goza de relaciones sociales de menor calidad y más soledad: ____
- d) No se emplea el tiempo de ocio en la consola, el ordenador y en los chats, y no se goza de relaciones sociales de menor calidad y más soledad: ____

7.

- a) Sólo nacen niños deseados, y aumenta la probabilidad de que los niños sean más felices: ____
- b) Sólo nacen niños deseados, y no aumenta la probabilidad de que los niños sean más felices: ____
- c) No sólo nacen niños deseados, y aumenta la probabilidad de que los niños sean más felices: ____
- d) No sólo nacen niños deseados, y no aumenta la probabilidad de que los niños sean más felices: ____

8.

- a) Los supermercados usan bolsas reutilizables, y contribuyen a disminuir la contaminación a largo plazo: ____
- b) Los supermercados usan bolsas reutilizables, y no contribuyen a disminuir la contaminación a largo plazo: ____
- c) Los supermercados no usan bolsas reutilizables, y contribuyen a disminuir la contaminación a largo plazo: ____
- d) Los supermercados no usan bolsas reutilizables, y no contribuyen a disminuir la contaminación a largo plazo: ____

9.

- a) Disminuye el número de créditos académicos, y los alumnos ven menos materia y saben menos que antes: ____

b) Disminuye el número de créditos académicos, y los alumnos no ven menos materia y no saben menos que antes: ____

c) No disminuye el número de créditos académicos, y los alumnos ven menos materia y saben menos que antes: ____

d) No disminuye el número de créditos académicos, y los alumnos no ven menos materia y no saben menos que antes: ____

10.

a) Disminuye la tasa de nacimientos de un país, y la población envejece: ____

b) Disminuye la tasa de nacimientos de un país, y la población no envejece: ____

c) No disminuye la tasa de nacimientos de un país, y la población envejece: ____

d) No disminuye la tasa de nacimientos de un país, y la población no envejece: ____

11.

a) Aumenta la proporción de madres primerizas a edades cercanas a los 40 años, y aumenta el número de de familias con un solo hijo: ____

b) Aumenta la proporción de madres primerizas a edades cercanas a los 40 años, y no aumenta el número de de familias con un solo hijo: ____

c) No aumenta la proporción de madres primerizas a edades cercanas a los 40 años, y aumenta el número de de familias con un solo hijo: ____

d) No aumenta la proporción de madres primerizas a edades cercanas a los 40 años, y no aumenta el número de de familias con un solo hijo: ____

12.

a) Se extiende el uso mochilas escolares con ruedas, y aumenta el número de niños con la columna desviada: ____

b) Se extiende el uso mochilas escolares con ruedas, y no aumenta el número de niños con la columna desviada: ____

c) No se extiende el uso mochilas escolares con ruedas, y aumenta el número de niños con la columna desviada: ____

d) No se extiende el uso mochilas escolares con ruedas, y no aumenta el número de niños con la columna desviada: ____

Version for experimental group G2 - Order -B1

Últimos 3 Números de tu DNI:

Nombre de tu abuela materna:

Instrucciones:

A continuación, se te presentarán 12 ítems compuestos por 4 posibilidades, por ejemplo:

ITEM 1:

- a) Alguien se droga frecuentemente y sufre daños psicológicos y cerebrales.
- b) Alguien se droga frecuentemente y **no** sufre daños psicológicos ni cerebrales.
- c) Alguien **no** se droga frecuentemente, pero sufre daños psicológicos y cerebrales.
- d) Alguien **no** se droga frecuentemente y **no** sufre daños psicológicos y ni cerebrales.

Estas son todas posibilidades que puede haber, y no más. Por tanto, cada ítem tendrá esta forma que has visto, es decir: a) que pase A y pase B; b) que pase A y no pase B; c) que no pase A y sí pase B; d) que no pase A y no pase B.

Tu tarea consistirá en otorgar a cada una de las 4 posibilidades de cada ítem un porcentaje de probabilidad, de tal manera, que sumadas las cuatro puntuaciones deben resultar en el 100% total de probabilidad.

Por ejemplo:

ITEM 1:

- e) Alguien se droga frecuentemente y sufre daños psicológicos y cerebrales: 55
- f) Alguien se droga frecuentemente y no sufre daños psicológicos ni cerebrales: 10
- g) Alguien no se droga frecuentemente y sufre daños psicológicos y cerebrales: 15
- h) Alguien no se droga frecuentemente y no sufre daños psicológicos y ni cerebrales: 20

100%

En cada página hay ítems por delante y por detrás.

Como sabes, no hay respuestas correctas ni incorrectas, simplemente debes hacer estimaciones en función de lo que tú crees.

1.

a) Se promueve el sexo entre adolescentes, y se contribuye al aumento de embarazos no deseados:

b) Se promueve el sexo entre adolescentes, y no se contribuye al aumento de embarazos no deseados: ___

c) No se promueve el sexo entre adolescentes, y se contribuye al aumento de embarazos no deseados: ___

d) No se promueve el sexo entre adolescentes, y no se contribuye al aumento de embarazos no deseados: ___

2.

a) Se preservan los hábitats naturales, y se salva la vida de muchos animales salvajes: ___

b) Se preservan los hábitats naturales, y no se salva la vida de muchos animales salvajes: ___

c) No se preservan los hábitats naturales, y se salva la vida de muchos animales salvajes: ___

d) No se preservan los hábitats naturales, y no se salva la vida de muchos animales salvajes: ___

3.

a) Uno pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.), y acaba con problemas visuales: ___

b) Uno pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.), y no acaba con problemas visuales: ___

c) Uno no pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.), y acaba con problemas visuales: ___

d) Uno no pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.), y no acaba con problemas visuales: ___

4.

a) Se aprueba la ley del aborto, y se produce indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes: ___

b) Se aprueba la ley del aborto, y no se produce indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes: ___

c) No se aprueba la ley del aborto, y se produce indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes: ___

d) No se aprueba la ley del aborto, y no se produce indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes: ___

5.

- a) La población de un país envejece, y aumenta la proporción de personas con enfermedades degenerativas: ____
- b) La población de un país envejece, y no aumenta la proporción de personas con enfermedades degenerativas: ____
- c) La población de un país no envejece, y aumenta la proporción de personas con enfermedades degenerativas: ____
- d) La población de un país no envejece, y no aumenta la proporción de personas con enfermedades degenerativas: ____

6.

- a) Se emplea el tiempo de ocio en la consola, el ordenador y los chats, y se pierden habilidades de creatividad e imaginación: ____
- b) Se emplea el tiempo de ocio en la consola, el ordenador y los chats, y no se pierden habilidades de creatividad e imaginación: ____
- c) No se emplea el tiempo de ocio en la consola, el ordenador y los chats, y se pierden habilidades de creatividad e imaginación: ____
- d) No se emplea el tiempo de ocio en la consola, el ordenador y los chats, y no se pierden habilidades de creatividad e imaginación: ____

7.

- a) Sólo nacen niños deseados, y disminuyen las rupturas familiares: ____
- b) Sólo nacen niños deseados, y no disminuyen las rupturas familiares: ____
- c) No sólo nacen niños deseados, y disminuyen las rupturas familiares: ____
- d) No sólo nacen niños deseados, y no disminuyen las rupturas familiares: ____

8.

- a) Los supermercados empiezan a usar bolsas reutilizables, y dañan al sector comercial del plástico: ____
- b) Los supermercados empiezan a usar bolsas reutilizables, y no dañan al sector comercial del plástico: ____
- c) Los supermercados no empiezan a usar bolsas reutilizables, y dañan al sector comercial del plástico: ____
- d) Los supermercados no empiezan a usar bolsas reutilizables, y no dañan al sector comercial del plástico: ____

9.

- a) Disminuye el número de créditos académicos, y los alumnos tienen más tiempo de ocio y diversión que antes: ____
- b) Disminuye el número de créditos académicos, y los alumnos no tienen más tiempo de ocio y diversión que antes: ____
- c) No disminuye el número de créditos académicos, y los alumnos tienen más tiempo de ocio y diversión que antes: ____
- d) No disminuye el número de créditos académicos, y los alumnos no tienen más tiempo de ocio y diversión que antes: ____

10.

- a) Disminuye la tasa de nacimientos de un país, y muchos colegios tienen que cerrar: ____
- b) Disminuye la tasa de nacimientos de un país, y no muchos colegios tienen que cerrar: ____
- c) No disminuye la tasa de nacimientos de un país, y muchos colegios tienen que cerrar: ____
- d) No disminuye la tasa de nacimientos de un país, y no muchos colegios tienen que cerrar: ____

11.

- a) Aumenta la proporción de madres primerizas a edades cercanas a los 40 años, y aumenta el número de recién nacidos con deficiencias: ____
- b) Aumenta la proporción de madres primerizas a edades cercanas a los 40 años, y no aumenta el número de recién nacidos con deficiencias: ____
- c) No aumenta la proporción de madres primerizas a edades cercanas a los 40 años, y aumenta el número de recién nacidos con deficiencias: ____
- d) No aumenta la proporción de madres primerizas a edades cercanas a los 40 años, y no aumenta el número de recién nacidos con deficiencias: ____

12.

- a) Se extiende el uso mochilas escolares con ruedas, y disminuyen los problemas infantiles de espalda: ____
- b) Se extiende el uso mochilas escolares con ruedas, y no disminuyen los problemas infantiles de espalda: ____
- c) No se extiende el uso mochilas escolares con ruedas, y disminuyen los problemas infantiles de espalda: ____
- d) No se extiende el uso mochilas escolares con ruedas, y no disminuyen los problemas infantiles de espalda: ____

Instrucciones:

A continuación, se te presentarán 16 ítems que consisten en oraciones del tipo: ‘*si pasa X, entonces pasará Y*’. Por ejemplo:

‘Si alguien se droga frecuentemente, entonces sufrirá daños psicológicos y cerebrales’.

Tu tarea consistirá en puntuar cuán probable es que pase lo que se dice en las frases, en una escala que va del 0 al 100%. Por ejemplo:

‘Si alguien se droga frecuentemente, entonces sufrirá daños psicológicos y cerebrales’: 80

Esto quiere decir que *creo* que lo que pone en la frase sucederá con una probabilidad del 80%.

Recuerda que no hay respuestas correctas ni incorrectas, simplemente debes hacer estimaciones en función de lo que tú crees. No mires las puntuaciones que has puesto en el ejercicio anterior.

MUCHAS GRACIAS POR TU COLABORACIÓN

ITEMS:

1. Si se promueve el sexo entre adolescentes, se puede contribuir al aumento de embarazos no deseados: ____

2. Si se preservan los hábitats naturales, entonces se salvará la vida de muchos animales salvajes: ____

3. Si uno pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.), acabará con problemas visuales: ____

4. Si se aprueba la ley del aborto, entonces se producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes: ____

5. Si la población de un país envejece, aumentará la proporción de personas con enfermedades degenerativas: ____

6. Si se emplea el tiempo de ocio en la consola, el ordenador y los chats, entonces se perderán habilidades de creatividad e imaginación: ____
7. Si sólo nacen niños deseados, entonces disminuirán las rupturas familiares: ____
8. Si los supermercados empiezan a usar bolsas reutilizables, entonces dañarán al sector comercial del plástico: ____
9. Si disminuye el número de créditos académicos, entonces los alumnos tendrán más tiempo de ocio y diversión que antes: ____
10. Si disminuye la tasa de nacimientos de un país, entonces muchos colegios tendrán que cerrar: ____
11. Si aumenta la proporción de madres primerizas a edades cercanas a los 40 años, entonces puede aumentar el número de recién nacidos con deficiencias: ____
12. Si se extiende el uso mochilas escolares con ruedas, entonces disminuirán los problemas infantiles de espalda: ____

Appendix A-VII – Post-test Phase Instructions for the Second Task

SPANISH VERSION – *Arguments’ Conclusions Evaluation Task for Experimental Groups G1 and G2*

Instrucciones:

A continuación, se te presentarán 12 ítems compuestos por tres frases: las dos primeras frases te informan de un contexto determinado y la última frase es una afirmación que podría derivarse de las dos frases anteriores. Por ejemplo:

‘Si aumenta la obesidad infantil, aumentará el riesgo de padecer enfermedades cardiovasculares.’

‘En la sociedad española está aumentando la obesidad infantil.’

Y la afirmación que podría derivarse de ellas sería:

‘En la sociedad española aumentará el riesgo de padecer enfermedades cardiovasculares.’

Tu tarea consistirá en evaluar **hasta qué punto crees que es probable la afirmación** que se deriva de las dos primeras frases, otorgándole una puntuación de probabilidad en una **escala de 11 puntos, que va de 0 (nada probable) a 1 (totalmente probable)**. Para ello deberás emplear la escala de probabilidad que tienes **en el teclado** de tu ordenador.

Esta puntuación debe reflejar el grado en el que *crees* que las situaciones descritas en dichas afirmaciones pueden ser (o son) ciertas, en función de tu grado de creencia en las dos oraciones que la preceden.

Debes emplear la tecla de ‘espacio’ tanto para que vaya apareciendo cada frase en la pantalla como para pasar al siguiente ítem; y emplear las teclas de la escala de probabilidad para puntuar las afirmaciones. Veamos el procedimiento:

1º: pulsa un ‘espacio’ y te aparecerá en la pantalla lo siguiente:

“Dado que:”

2º: ‘espacio’:

‘Si aumenta la obesidad infantil, aumentará el riesgo de padecer enfermedades cardiovasculares.’

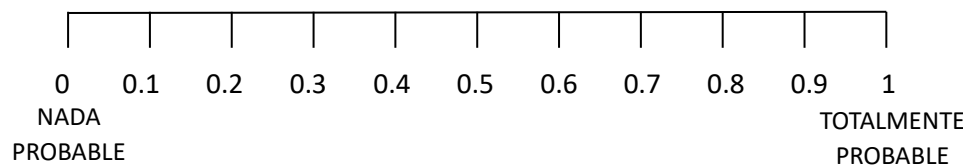
3º: ‘espacio’:

‘En la sociedad española está aumentando la obesidad infantil.’

4º: ‘espacio’:

Por favor, estima la probabilidad de ocurrencia de la siguiente situación:

‘En la sociedad española aumentará el riesgo de padecer enfermedades cardiovasculares.’



5º: pulsa la tecla de la escala de probabilidad (que hay en el teclado) que mejor se adecue a lo que tú crees. Por ejemplo, yo puedo creer que esto es probable en un 0.7 de un total de 1 y, por tanto, presionaré la tecla marcada con un 0.7.

6º: pulsa un ‘espacio’ para pasar al siguiente ítem.

Como sabes, no hay respuestas correctas ni incorrectas, simplemente debes hacer estimaciones en función de lo que tú crees.

¡Recuerda poner tu ID!

MUCHAS GRACIAS POR TU COLABORACIÓN

ENGLISH VERSION – *Arguments’ Conclusions Evaluation Task for Experimental Groups G1 and G2*

Instructions:

Next you will be presented with 12 items made up of three sentences: the first two sentences inform you about a certain context and the last sentence is a statement that could be derived from the previous two sentences. For example:

‘If childhood obesity increases, the risk of cardiovascular disease will increase.’

‘In Spanish society, childhood obesity is increasing.’

And the statement that could be derived from them would be:

‘In Spanish society, the risk of suffering from cardiovascular diseases will increase.’

Your task will be **to assess how likely you think the statement derived from the first two sentences is** by giving it a likelihood score on an **11-point scale, ranging from 0 (not at all likely) to 1 (totally likely)**. To do this you must use the probability scale that you have on your computer keyboard.

This score should reflect the degree to which you believe that the situations described in these statements can be (or are) true, based on your degree of belief in the two sentences that precede it.

You must use the 'space bar' key both for each sentence to appear on the screen and to move on to the next item; and use the keys of the probability scale to score the statements. Let us see the procedure:

1st: press the key space bar and you will see the next on the screen:

"Given that"

2nd: press the key space bar.

"If childhood obesity increases, the risk of cardiovascular disease will increase."

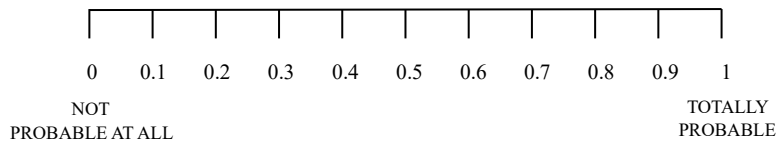
3rd: press the key bar.

"In Spanish society, childhood obesity is increasing."

4th: press the key space bar.

"Please, give an estimation of the probability of occurrence of the following situation:"

'In Spanish society, the risk of suffering from cardiovascular diseases will increase.'



5th: on the top of the keyboard, you will see a probabilistic scale that ranges from 0 (not probable at all) to 1 (totally probable). Press the key that best suits what you believe. For example, I can believe that this is likely to be 0.7 out of 1, and so I will press the key marked 0.7.

6th: press the 'space bar key' to pass to the next item.

As you know, there are no right or wrong answers, you just must make estimates based on what you believe.

Do not forget to insert your ID!

THANK YOU VERY MUCH FOR YOUR COOPERATION

Appendix A-VIII – Post-test Phase: Arguments used in the Second Task

Experimental Group G1

Arguments with Weakly Associated Warrants

(1)

Si se promueve el sexo entre adolescentes, se puede contribuir al aumento de la homosexualidad.

Los estereotipos sociales presentes en ciertos programas de televisión promueven el sexo entre adolescentes.

Los estereotipos sociales presentes en ciertos programas de televisión pueden contribuir al aumento de la homosexualidad.

(2)

Si se preservan los hábitats naturales, entonces se fomentará el turismo de montaña y rural.

El reciclaje contribuye fuertemente a preservar los hábitats naturales.

El reciclaje contribuirá fuertemente a fomentar el turismo de montaña y rural.

(3)

Si uno pasa demasiado tiempo delante de una pantalla (ordenador, televisión, etc.), podrá tener ataques de epilepsia.

Una importante proporción de los niños de hoy pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.).

Una importante proporción de los niños de hoy podrá tener ataques de epilepsia mañana.

(4)

Si se aprueba la ley del aborto, entonces se producirá indirectamente una disminución de la natalidad.

El parlamento español ha aprobado la ley del aborto.

El parlamento español producirá indirectamente una disminución de la natalidad.

(5)

Si aumenta la proporción de madres primerizas a edades cercanas a los 40 años, entonces puede aumentar el número de familias con un solo hijo.

En la sociedad española está aumentando la proporción de madres primerizas a edades cercanas a los 40 años.

En la sociedad española puede aumentar el número de familias con un solo hijo.

(6)

Si se extiende el uso mochilas escolares con ruedas, entonces aumentará el número de niños con la columna desviada.

En nuestra sociedad se ha extendido el uso de mochilas escolares con ruedas.

En nuestra sociedad aumentará el número de niños con la columna desviada.

Arguments with Strongly Associated Warrants

(1)

Si en un país su población envejece, entonces se verá en graves problemas económicos (pocos para trabajar y muchas pensiones por pagar).

La población china está envejeciendo debido a que mantienen un estricto control de la natalidad.

China se verá en graves problemas económicos.

(2)

Si se emplea el tiempo de ocio en la consola, el ordenador y en los chats, entonces se gozará de relaciones sociales de menor calidad y más soledad.

Los niños y adolescentes de nuestra sociedad emplean el tiempo en la consola, el ordenador y los chats.

Los niños y adolescentes de nuestra sociedad gozarán de relaciones sociales de menor calidad y más soledad.

(3)

Si sólo nacen niños deseados, entonces aumentará la probabilidad de que los niños sean más felices.

La ley del aborto aumenta la probabilidad de que sólo nazcan niños deseados.

La ley del aborto aumentará la probabilidad de que los niños sean más felices.

(4)

Si los supermercados empiezan a usar bolsas reutilizables, entonces contribuirán a disminuir la contaminación a largo plazo.

Los supermercados Carrefour emplean bolsas reutilizables.

Los supermercados Carrefour contribuirán a la disminuir la contaminación a largo plazo.

(5)

Si disminuye el número de créditos académicos, entonces los alumnos verán menos materia y sabrán menos que antes.

Con el nuevo sistema universitario de grados ha disminuido el número de créditos académicos respecto del plan anterior.

Con el nuevo sistema universitario de grados los alumnos verán menos materia y sabrán menos que antes.

(6)

Si disminuye la tasa de nacimientos de un país, entonces la población envejecerá.

En los países ricos está disminuyendo la tasa de natalidad.

En los países ricos la población envejecerá.

Experimental Group G2

Arguments with Weakly Associated Warrants

(1)

Si la población de un país envejece, aumentará la proporción de personas con enfermedades degenerativas.

En España la población está envejeciendo.

En España aumentará la proporción de personas con enfermedades degenerativas.

(2)

Si se emplea el tiempo de ocio en la consola, el ordenador y los chats, entonces se perderán habilidades de creatividad e imaginación.

Los niños y adolescentes de nuestra sociedad emplean el tiempo en la consola, el ordenador y los chats.

Los niños y adolescentes de nuestra sociedad perderán habilidades de creatividad e imaginación.

(3)

Si sólo nacen niños deseados, entonces disminuirán las rupturas familiares.

La ley del aborto aumenta la probabilidad de que solo nazcan niños deseados.

La ley del aborto disminuirá las rupturas familiares.

(4)

Si los supermercados empiezan a usar bolsas reutilizables, entonces dañarán al sector comercial del plástico.

Los supermercados Carrefour emplean bolsas reutilizables.

Los supermercados Carrefour dañarán al sector comercial del plástico.

(5)

Si disminuye el número de créditos académicos, entonces los alumnos tendrán más tiempo de ocio y diversión que antes.

Con el nuevo sistema universitario de grados ha disminuido el número de créditos académicos respecto del plan anterior.

Con el nuevo sistema universitario de grados los alumnos tendrán más tiempo de ocio y diversión que antes.

(6)

Si disminuye la tasa de nacimientos de un país, entonces muchos colegios tendrán que cerrar.

En los países ricos está disminuyendo la tasa de natalidad.

En los países ricos muchos colegios tendrán que cerrar.

Arguments with Strongly Associated Warrants

(1)

Si se promueve el sexo entre adolescentes, se puede contribuir al aumento de embarazos no deseados.

Los estereotipos sociales presentes en ciertos programas de televisión promueven el sexo entre adolescentes.

Los estereotipos sociales de los medios de comunicación pueden contribuir al aumento de embarazos no deseados.

(2)

Si se preservan los hábitats naturales, entonces se salvará la vida de muchos animales salvajes.

El reciclaje contribuye fuertemente a preservar los hábitats naturales.

El reciclaje contribuirá fuertemente a salvar la vida de muchos animales salvajes.

(3)

Si uno pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.), acabará con problemas visuales.

Una importante proporción de los niños de hoy pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.).

Una importante proporción de los niños de hoy acabará con problemas visuales.

(4)

Si se aprueba la ley del aborto, entonces se producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.

El parlamento español ha aprobado la ley del aborto.

El parlamento español producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.

(5)

Si aumenta la proporción de madres primerizas a edades cercanas a los 40 años, entonces puede aumentar el número de recién nacidos con deficiencias.

En la sociedad española está aumentando la proporción de madres primerizas a edades cercanas a los 40 años.

En la sociedad española puede aumentar el número de recién nacidos con deficiencias.

(6)

Si se extiende el uso mochilas escolares con ruedas, entonces disminuirán los problemas infantiles de espalda.

En nuestra sociedad se ha extendido el uso de mochilas escolares con ruedas.

En nuestra sociedad disminuirán los problemas infantiles de espalda.

Appendix A-IX – Analysis and Results Combining Groups G1 and G2, and C1 and C2 (text in Spanish)

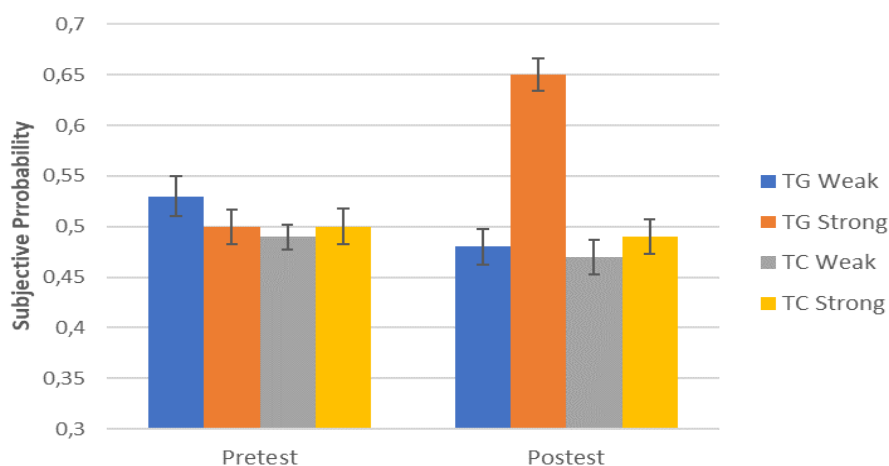
En este análisis, se agregan los grupos G1 y G2 para tratarlos como un solo grupo. De este modo, este grupo experimental tendrá un tamaño de $n = 55$. De modo similar, se agregan los grupos C1 y C2, de tal modo que el grupo control pasa a tener un tamaño de 38 participantes.

Tanto G1 como G2 han tenido el mismo diseño y procedimiento, y sus materiales fueron contruidos del mismo modo, siendo temáticamente complementarios. Pese a que este análisis conjunto no es metodológicamente correcto, puede aportar información valiosa al aumentar la potencia estadística y permite hacer análisis por ítems. Las medias se muestran en esta tabla.

Experimental	Big Group	
	D	F
Data Premise	.77 (.12)	.80 (.12)
Warrant		
Contingency	.60 (.12)	.74 (.09)
Direct Scoring	.52 (.13)	.66 (.13)
Critical Assertion		
Prior	.53 (.14)	.50 (.12)
Post	.48 (.12)	.65 (.12)
Control		
Critical Assertion		
Prior	.49 (.09)	.50 (.13)
Post	.47 (.12)	.49 (.12)

Este patrón de resultados se observa en el siguiente gráfico:

Subjective Probabilities Given to the Critical Assertions by the Total Group



Note: Mean subjective probabilities scores given by the total group (TG) as a function of argumentative treatment (prior – posterior) and the warrant association force (Strong or Weak). The error bars represent standard error.

En este análisis se observa un efecto principal de fuerza de asociación del garante ($F(1, 91) = 8.02$, $p = .006$, $\eta_p^2 = .081$) y del grupo ($F(1, 91) = 9.86$, $p = .002$, $\eta_p^2 = .098$).

También se observó una interacción del momento del contexto de argumentación con el grupo ($F(1, 91) = 8.55$, $p = .004$, $\eta_p^2 = .086$). Las comparaciones por pares corregidas por Bonferroni mostraron que no hubo diferencia de medias entre los grupos en el pretest ($t(91) = 1.04$, $p = .301$) pero sí que hubo una diferencia significativa en el posttest ($t(91) = 4.34$, $p < .001$; $d = .99$). En este caso, el grupo experimental obtuvo un promedio ($M = .57$) superior al del grupo control ($M = .48$). Por otro lado, también mostraron que mientras que en el grupo control no hubo diferencias entre el pretest y el posttest ($t(38) = 1.30$, $p = .201$), en el grupo experimental sí que las hubo ($t(55) = -3.01$, $p = .004$; $d = .60$), donde la media del posttest ($M = .57$) fue superior a la del pretest ($M = .52$).

El ANOVA también reveló una interacción entre el momento del contexto de argumentación y la fuerza de asociación en el garante del argumento ($F(1, 91) = 35.70$, $p < .001$, $\eta_p^2 = .282$). Las comparaciones post-hoc ajustadas por Bonferroni revelaron que no hubo diferencias en función de la fuerza de asociación en el pretest ($t(92) = .89$, $p = .372$). Sin embargo, revelaron una diferencia significativa en el posttest ($t(92) = -6.06$, $p < .001$; $d = .76$), donde la media de la condición de fuerte asociación ($M = .58$) fue superior a la media de credibilidad otorgada a las oraciones pertenecientes a la condición garantes de débil asociación ($M = .48$). Por otra parte, al comparar el pretest y el posttest en función de la fuerza de asociación, los análisis revelaron diferencias significativas tanto en la condición débil ($t(92) = 2.78$, $p = .006$, $d = .33$), como en la condición fuerte ($t(92) = -5.13$, $p < .001$, $d = .61$). En la condición débil, la media disminuyó al pasar del pretest ($M = .52$) al posttest ($M = .48$), mientras que en la condición fuerte aumentó (M Pretest = .50; M Posttest = .58).

Finalmente, también se observó una interacción triple significativa entre el grupo, el momento del contexto de argumentación y la fuerza de asociación ($F(1, 91) = 33.02$, $p < .001$, $\eta_p^2 = .266$). Este patrón de resultados se observa en la figura y en la tabla de más arriba. Las comparaciones por pares ajustadas por Bonferroni mostraron que, en el pretest, no hubo diferencias entre ambos grupos ni en la condición de asociación débil ($t(90.10) = 1.85$, $p = .067$), ni fuerte ($t(91) = -.11$, $p = .907$). En el posttest, no hubo diferencias entre los grupos en la condición de fuerza de asociación débil ($t(91) = .48$, $p = .907$), pero sí en la condición de asociación fuerte ($t(91) = 6.20$, $p < .001$, $d = 1.39$). La media del grupo experimental ($M = .65$) fue superior a la del grupo control ($M = .49$). Por otra parte, en el grupo experimental, al comparar las valoraciones obtenidas en el pretest con las del posttest en la condición débil no se observaron diferencias significativas ($t(54) = 2.54$, $p = .014$)⁷¹, o, si acaso, marginalmente significativas. Sin embargo, sí las hubo en la condición de fuerte asociación ($t(54) = -7.33$, $p < .001$, $d = 1.30$), donde la media del pretest ($M = .50$) fue inferior a la del posttest ($M = .65$). No hubo diferencias entre el pretest y el posttest en el grupo control, ni en la

⁷¹ Se considera no significativa debido al ajuste según el método de Bonferroni, que para cuatro comparaciones sitúa la significación en $p < .013$.

condición de débil asociación ($t(37) = 1.17, p = .247$), ni en la de fuerte asociación ($t(37) = .835, p = .409$).

Igual que en los análisis de G1 y G2, se realizó un ANOVA 2 x 2 para el grupo experimental. Este análisis reveló un efecto principal del momento del contexto de argumentación ($F(1, 54) = 9.10, p = .004, \eta_p^2 = .144$), donde la media del posttest ($M = .57$) fue superior a la del pretest ($M = .52$). También reveló un efecto principal de la fuerza de asociación ($F(1, 54) = 16.76, p < .001, \eta_p^2 = .237$), donde la media resultante de la condición de asociación fuerte ($M = .57$) fue superior a la de la condición de asociación débil ($M = .51$).

Este análisis también reveló un efecto de interacción entre el momento del contexto de argumentación y el de fuerza de asociación ($F(1, 54) = 65.99, p < .001, \eta_p^2 = .550$). Las comparaciones post-hoc ajustadas por Bonferroni revelaron que no hubo diferencias en el pretest en función de la fuerza de asociación ($t(54) = 1.54, p = .128$). Sin embargo, sí que hubo una diferencia significativa en el posttest ($t(54) = -9.79, p < .001, d = 1.47$), donde la media resultante de la condición de fuerte asociación ($M = .65$) resultó ser significativamente superior a la de débil asociación ($M = .48$).

Se realizó un análisis equivalente para el grupo control, y no se observaron efectos principales ni de interacción de ninguna de las variables independientes.

Appendix A-X – On the Two Methods of Measuring the Credibility of Warrant Premises

A continuación, se presentan los objetivos, el método y los resultados de la medición de la probabilidad subjetiva de las premisas garante y su relación con la investigación realizada aquí.

Objetivos Experimentales y Metodología

Este estudio tiene como objetivo secundario explorar la interpretación probabilística de los condicionales en el ámbito del razonamiento y la argumentación. Históricamente, se han empleado diversos métodos para medir la probabilidad subjetiva de los condicionales, contribuyendo significativamente a nuestra comprensión del razonamiento probabilístico (Cheng, 1997; George, 1995, 1997; Liu, Lo y Wu, 1996; Oaksford y Chater, 1994; Manktelow y Over, 1995). Un método destacado es el de la tabla de contingencia o “tabla de verdad del condicional” (Evans y Over, 2003; Mandel y Lehman, 1998; Manktelow y Over, 1995; Oaksford y Moussakowski, 2004; Oaksford y Wakefield, 2003; Over et al., 2007; Stevenson y Over, 1995, 2001). Esta técnica captura implícitamente los juicios de probabilidad subjetiva de las oraciones condicionales, que Over et al. (2007) utilizaron en su estudio centrado en los condicionales temáticoscausales. Sus hallazgos apoyaron la hipótesis de la probabilidad condicional, sugiriendo que la probabilidad subjetiva de un condicional cotidiano 'si p , entonces q ' coincide con la probabilidad condicional $P(q | p)$.

Aquí se pretende replicar y ampliar esta metodología para evaluar la credibilidad de los garantes en los argumentos. El experimento emplea dos medidas de probabilidad subjetiva: puntuación directa y tablas de contingencia (tablas de verdad condicional) para investigar cómo los participantes interpretan la credibilidad del garante en los argumentos. El estudio evaluará la equivalencia de estas medidas al evaluar no solo su correlación sino también su convergencia en las puntuaciones promedio. El objetivo es determinar si estos métodos de medición correlacionan y convergen lo suficiente, sugiriendo que estudios futuros podrían utilizar exclusivamente la puntuación directa debido a su aplicación natural y sencilla. Sin embargo, si surgen diferencias significativas entre estas medidas, será necesario realizar una evaluación detallada para determinar qué método se corresponde más estrechamente con la credibilidad final de la afirmación, capturando efectivamente la variabilidad en las respuestas de los participantes.

Al explorar estos aspectos, el estudio busca proporcionar información sobre la posible relación lineal, directa y sistemática entre las medidas de credibilidad del garante y la credibilidad de la conclusión. Este estudio exploratorio se abstiene de proponer hipótesis específicas sobre el ajuste superior de cualquiera de los métodos de medición, centrándose en cambio en la evaluación empírica y la refinación metodológica en el contexto del razonamiento y la argumentación probabilística.

Predicciones

En relación con el objetivo secundario, si la evaluación de la credibilidad de un condicional causal refleja un proceso probabilístico en el razonamiento cotidiano, se espera que las puntuaciones obtenidas mediante ambos métodos de evaluación del garante muestren correlaciones altas y significativas entre sí. Además, si calcular la credibilidad del condicional equivale a determinar su probabilidad condicionada, no se esperan diferencias en los promedios de los grados de credibilidad obtenidos con ambos métodos. Por último, se espera que las mediciones de probabilidad subjetiva del garante se relacionen de manera lineal, directa y sistemática con la probabilidad subjetiva de las conclusiones, reflejando así el proceso de razonamiento.

Método

En la primera tarea, se evaluó la credibilidad de las premisas garante mediante dos métodos. Uno de ellos fue la evaluación utilizando la tabla de contingencias, que consistía en presentar a los participantes las cuatro posibles combinaciones lógicas de los elementos del condicional (p y q , p y no- q , no- p y q , no- p y no- q), y luego asignar 100 puntos entre estas opciones según la percepción de probabilidad de cada una. Se les instruyó que la suma total de puntos debía ser igual a 100. Por ejemplo:

	%
Alguien se droga frecuentemente y acaba con daños cerebrales irreparables.	45
Alguien se droga frecuentemente y no acaba con daños cerebrales irreparables.	10
Alguien no se droga frecuentemente y acaba con daños cerebrales irreparables.	15
Alguien no se droga frecuentemente y no acaba con daños cerebrales irreparables.	30
	100

Después, con estos datos, se puede calcular la probabilidad condicionada para cada ítem mediante la fórmula $P(q|p) = \frac{p \wedge q}{(p \wedge q) + (p \wedge \neg q)} = \frac{45}{45+10} = 0.81$

El otro método consistió en evaluar los garantes mediante una puntuación directa sobre 100. En esta tarea, se solicitó a los participantes que asignaran directamente una probabilidad subjetiva al condicional, estimando el porcentaje de confianza en la futura ocurrencia real de las situaciones descritas en las oraciones. Por ejemplo:

Si alguien se droga frecuentemente, entonces acabará con daños cerebrales irreparables: 81%

El método de la tabla de contingencia obliga al participante a adoptar una interpretación probabilística, mientras que el método de puntuación directa no lo hace. De hecho, el método de puntuación directa es más natural y permite la influencia de sesgos y preferencias en la interpretación del condicional y en la evaluación de su probabilidad, reflejando así el pensamiento cotidiano. Al comparar ambos métodos, podemos observar la naturaleza probabilística del

pensamiento cotidiano y determinar si esto se refleja en las conclusiones resultantes del razonamiento completo.

Resultados sobre los Métodos de Medición del Garante

Correlaciones entre los dos Métodos

Se realizaron correlaciones de Pearson entre las valoraciones de probabilidad subjetiva del garante, obtenidas mediante el método de las tablas de contingencia y el de puntuaciones directas.

Para las correlaciones por ítems, es importante tener en cuenta que cada grupo (G1 y G2) evaluó 6 garantes de asociación débil y 6 garantes de asociación fuerte entre su causa (antecedente del condicional) y su efecto (consecuente). Debido a que los ítems de débil y fuerte asociación eran diferentes entre sí, se pudieron poner las 12 puntuaciones en una sola columna, sin tener que separar por fuerza de asociación. Por lo tanto, en cada grupo se llevó a cabo un análisis de correlación entre dos columnas de datos (una para cada método), cada una con 12 puntuaciones. La correlación por ítems obtenida en el grupo G1 fue de $r = .97$, $p < .001$, mientras que en el grupo G2 fue de $r = .90$, $p < .001$.

En las correlaciones por sujetos, cada participante genera cuatro puntuaciones, una para cada nivel de fuerza de asociación (fuerte y débil) en cada uno de los dos métodos (puntuación directa y tabla de contingencia). Dado que la fuerza de asociación se aplica a los ítems individuales, no es posible combinar las puntuaciones de cada nivel de fuerza de asociación sin duplicar los participantes. Por lo tanto, se requiere realizar correlaciones separadas para cada nivel de fuerza de asociación en cada grupo.

En el grupo G1 ($n = 30$), la correlación por sujetos entre los dos métodos en el nivel de asociación débil fue de $r = .60$, $p < .001$, y en el nivel de asociación fuerte fue de $r = .56$, $p = .001$. En el grupo G2 ($n = 25$), esta correlación en el nivel de asociación débil fue de $r = .64$, $p < .001$. Sin embargo, la correlación en el nivel de asociación fuerte no alcanzó significación con una $r = -.08$, $p = .351$.

Comparaciones de Medias entre los dos Métodos

Entonces, se realizaron comparaciones de las medias obtenidas por los dos métodos en cada grupo, por ítems y por sujetos. Las comparaciones por ítems revelaron que ambos métodos difieren significativamente en los promedios resultantes en G1 ($t(11) = 4.73$, $p = .001$) y en G2 ($t(11) = 9.10$, $p < .001$).

Por otra parte, en G1, las comparaciones por sujetos ajustadas por Bonferroni (i.e., $p < .025$) de ambos métodos revelaron diferencias significativas cuando la fuerza de asociación de los elementos de garante era débil ($t(29) = 3.78$, $p = .001$), así como cuando era fuerte ($t(29) = 2.45$, p

= .020). Del mismo modo, en G2 hubo diferencias significativas tanto en el nivel de asociación débil ($t(24) = 4.55, p < .001$) como fuerte ($t(24) = 3.74, p = .001$).

Correlaciones entre las Conclusiones y los dos Métodos

Dadas las diferencias de medias en los grados de probabilidad subjetiva obtenidos en cada método, es relevante determinar cuál de los dos métodos está más estrechamente relacionado con los resultados observados en las conclusiones de los argumentos. A continuación, se comentan las correlaciones de ambos métodos con las afirmaciones críticas en el posttest, por ítems y por sujetos para cada grupo.

En el grupo G1, el análisis de correlaciones por ítems no reveló relaciones sistemáticamente significativas entre ninguno de los dos métodos y la credibilidad de las conclusiones. En el grupo G2, este análisis reveló una correlación significativa entre las estimaciones de probabilidad de las conclusiones y las estimaciones de probabilidad del garante medidas mediante el método de la tabla de contingencias ($r = .63; p = .014$).

En el análisis de correlaciones por sujetos del grupo G1, se encontraron relaciones significativas entre las puntuaciones de las conclusiones y las puntuaciones del garante obtenidas mediante ambos métodos de medición, para el nivel de fuerza de asociación débil. Las correlaciones fueron $r = .44; p = .007$ (por tabla de contingencias) y $r = .65; p < .001$ (por puntuación directa). Sin embargo, para el nivel de fuerza de asociación fuerte, solo se encontró una relación significativa entre las puntuaciones de las conclusiones y las del garante según el método de puntuación directa ($r = .36, p = .025$).

En el grupo G2, las correlaciones por sujetos también revelaron relaciones significativas entre las puntuaciones otorgadas a las conclusiones y las puntuaciones del garante tanto en el nivel de fuerza de asociación débil como en el de fuerza de asociación fuerte. Específicamente, en el nivel de asociación débil, estas relaciones significativas se produjeron con ambos métodos: $r = .70; p < .001$ (tabla de contingencias) y $r = .39; p = .028$ (puntuación directa). Sin embargo, en el nivel de asociación fuerte, solo se produjeron con el método de puntuación directa ($r = .61; p = .001$).

Discusión de los Resultados sobre los dos Métodos de Medición

El propósito de comparar los dos métodos era explorar la interpretación probabilística de los condicionales por parte de los participantes, replicando parcialmente el estudio de Over et al. (2007). Se buscaba verificar si la verdad o falsedad de las oraciones condicionales de temática causal se determinaba mediante el cálculo de probabilidades subjetivas condicionadas (es decir, estimando la probabilidad de q dado p). Si los participantes realizaban una interpretación probabilística, se esperaban correlaciones significativas y altas entre el método de la tabla de contingencias y el método de estimación de puntuaciones directas. Si esto se confirmaba, se

pretendía investigar si los dos métodos eran realmente equivalentes, o si generaban grados de credibilidad diferentes. Finalmente, si se observaban diferencias en los grados de credibilidad, se buscaba determinar si alguno de los métodos registraba esos grados de una manera más congruente con las credibilidades de las conclusiones.

Se realizaron correlaciones tanto por ítems como por sujetos. La correlación por ítems permite identificar patrones sistemáticos en las respuestas de los participantes a través de los diferentes ítems y responde a la pregunta planteada. La elevada correlación encontrada es consistente con el supuesto de que los participantes interpretan los condicionales de manera probabilística, replicando los resultados encontrados en Over et al. (2007). Sin embargo, en la correlación por sujetos los resultados no son tan claros. Las correlaciones realizadas en el grupo G1 y en el nivel de asociación débil del grupo G2 sugieren que no hay grandes diferencias individuales en la interpretación de los ítems por parte de los sujetos. Sin embargo, la ausencia de significación en la correlación del nivel de asociación fuerte en G2 indica lo contrario. Las correlaciones significativas fueron del orden del $r = .6$, mientras que la no significativa fue de $r = .08$. Es una diferencia importante, más aún teniendo en cuenta que en G1 fue significativa. Sería necesario llevar a cabo un análisis más detallado para determinar la causa exacta de esta discrepancia en los resultados.

En su conjunto, los resultados respaldan la hipótesis de una interpretación probabilística de la credibilidad de los condicionales, mostrando una alta correlación entre ambos métodos en la evaluación de los ítems. Sin embargo, la falta de relación en el nivel de fuerte asociación del G2 sugiere que la consistencia observada entre los participantes en esa interpretación podría no ser tan sólida como cabría esperar, lo que requiere cautela en la interpretación de los resultados. Dados estos resultados, es pertinente comprobar si ambos métodos de medición derivan promedios equivalentes del grado de probabilidad subjetiva. Los resultados mostraron diferencias significativas.

Los resultados de los ANOVA indicaron que la fuerza de asociación del garante tiene efectos sobre la credibilidad de la conclusión de los argumentos. Entonces, dado que ambos métodos de medición muestran una sensibilidad distinta en el registro de las probabilidades subjetivas, es posible que uno de ellos se ajuste con mayor precisión a la variabilidad observada en las conclusiones de los argumentos. Pero los resultados de las correlaciones por ítems tan solo mostraron una correlación significativa en el grupo G2 con el método de la tabla de contingencias. Esto podría indicar que este método captura mejor el tipo de operaciones implicadas en el cómputo de la credibilidad, reflejando así una mayor congruencia con la credibilidad observada en las conclusiones de los argumentos. Pero la ausencia de significación en esta correlación en G1 sugiere que debería realizarse una evaluación más detallada para determinar qué método de medición se ajusta mejor a la variabilidad de los resultados observados en las conclusiones de los argumentos en cada grupo.

Los resultados de las correlaciones por sujetos indican que, cuando la fuerza de asociación del garante es débil, tanto el método de puntuación directa como el de la tabla de contingencias

muestran correlaciones significativas entre las puntuaciones de probabilidad atribuidas al garante y la credibilidad de la conclusión. Sin embargo, cuando la fuerza de asociación del garante es fuerte, solo la puntuación directa muestra una correlación significativa con la credibilidad de la conclusión. Estos hallazgos sugieren que, mientras las puntuaciones directas mantienen una relación consistente con la credibilidad de la conclusión tanto en condiciones de garante débil como fuerte, la tabla de contingencias solo exhibe esta relación en condiciones de garante fuerte. En consecuencia, se puede concluir que la puntuación directa muestra una correlación más persistente con la credibilidad de la conclusión que la tabla de contingencias.

Conclusiones Generales

Este estudio replica parcialmente investigaciones anteriores. Se buscaba verificar si los participantes calculaban probabilidades subjetivas condicionadas para determinar la verdad o falsedad de las oraciones condicionales. Los resultados respaldaron esta hipótesis, mostrando una alta correlación entre los métodos utilizados en la evaluación de los ítems. Sin embargo, se observaron diferencias entre grupos G1 y G2 y niveles de asociación del garante, indicando que la consistencia en la interpretación probabilística podría no ser sólida en todas las condiciones.

Por otro lado, las diferencias en los promedios de probabilidad subjetiva del garante derivados por ambos métodos llevaron a realizar correlaciones con la credibilidad de las conclusiones en el posttest. En el análisis por sujetos, se encontró una relación entre las puntuaciones de las conclusiones y de los garantes (por ambos métodos) en la condición de asociación débil. Sin embargo, en asociación fuerte, solo las puntuaciones de los garantes obtenidas mediante el método de puntuación directa mostraron una relación con las puntuaciones de las conclusiones. Estos resultados sugieren que las puntuaciones directas mantienen una relación consistente con la credibilidad de la conclusión tanto en condiciones de garante débil como fuerte, mientras que la tabla de contingencias solo exhibe esta relación en condiciones de garante fuerte.

Como cuestión aparte, volviendo a los resultados principales del experimento 1, se ha interpretado que el hecho de que la probabilidad subjetiva de las conclusiones en la condición débil del grupo G2 no disminuyera indica que los participantes ignoraron las premisas de sus argumentos al considerarlas no creíbles. Sin embargo, al menos las puntuaciones de la premisa garante, registradas *por ambos métodos*, estuvieron significativamente relacionadas con las conclusiones en el momento posterior al tratamiento argumentativo. Esto sugiere que los participantes sí prestaron atención a la probabilidad de las premisas. Es posible entonces que la magnitud de credibilidad de las premisas, al integrarse, haya derivado en esa misma credibilidad para la conclusión. Es decir, que la magnitud resultante de la integración de la credibilidad de las premisas fuera equivalente a la magnitud de credibilidad a priori.

Acerca de la Naturaleza Probabilística del Cómputo de Credibilidad del Garante

El estudio también buscaba determinar si los participantes realizan cálculos probabilísticos en la interpretación y evaluación de los garantes. Parcialmente replicando el trabajo de Over et al. (2007), se esperaban altas correlaciones entre las probabilidades subjetivas de los garantes obtenidas mediante el método de puntuación directa y las calculadas mediante la fórmula de probabilidad condicionada a partir de los datos obtenidos con el método de la tabla de contingencia.

Los resultados mostraron altas correlaciones entre las probabilidades subjetivas de los garantes obtenidas por ambos métodos, como se esperaba. Esto sugiere que, al asignar una probabilidad subjetiva a una oración condicional, los participantes están estableciendo su grado de creencia en el consecuente asumiendo que el antecedente es cierto. Esta observación apunta a la posibilidad de que la credibilidad en general pueda establecerse de manera probabilística, no solo en el contexto de oraciones condicionales.

Si las probabilidades subjetivas de ambas premisas se integran en el argumento y la de la premisa de datos se mantiene constante en niveles altos, las variaciones en la probabilidad del garante deberían reflejarse en la probabilidad atribuida a la conclusión. Los promedios de las probabilidades subjetivas medidas por cada método fueron distintos. Por tanto, es posible que uno de los métodos sea más sensible para capturar las características de la probabilidad subjetiva que el sistema de creencias computa en el razonamiento.

Los resultados indicaron claramente que los grados de probabilidad subjetiva de las premisas están relacionados con los grados de probabilidad atribuidos a la conclusión. Sin embargo, no permitieron observar si alguno de los métodos funcionaba mejor que el otro. Las correlaciones por ítems indicaron que lo hacía el método de la tabla de contingencia, pero solo en el grupo G2. Las correlaciones por sujetos mostraron que, cuando la fuerza de asociación del garante era débil, la probabilidad registrada mediante ambos métodos estaba relacionada con la de la conclusión en los dos grupos. Sin embargo, cuando la fuerza de asociación era fuerte, solo lo hacía el método de puntuación directa (también en ambos grupos). Estos resultados apoyan la idea de que los cálculos de credibilidad del argumento son de orden probabilístico y que, posiblemente, el método de puntuaciones directas esté funcionando mejor para medir la probabilidad subjetiva.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

APÉNDICE B: EXPERIMENTO 2

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

Appendix B-I – Instructions for Experiment 2

SPANISH VERSION – *First Task Instructions: Premises Evaluation Task*

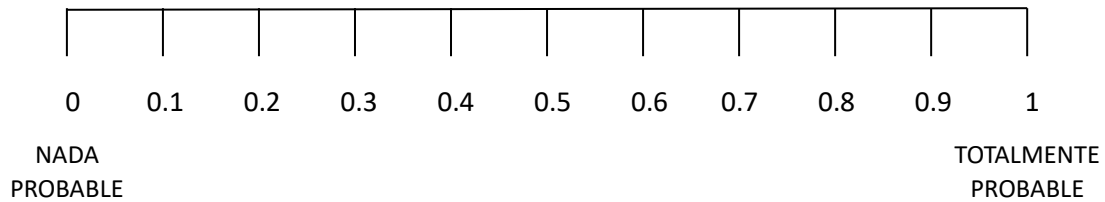
Instrucciones:

A continuación, verás una serie de afirmaciones. Tu tarea consiste en otorgarles una puntuación en una escala de probabilidad de 11 puntos que va de 0 (nada probable) a 1 (totalmente probable) pasando por 0, 0.1, 0.2... 0.9, hasta 1. Esta puntuación debe reflejar el grado en el que crees que las situaciones descritas en dichas afirmaciones pueden ser (o son) ciertas.

Ejemplo:

Por favor, estima la probabilidad de ocurrencia de la siguiente situación:

'El planeta Tierra gira al rededor del Sol.'



Si, por ejemplo, consideras que la probabilidad de que ‘*el planeta Tierra gira al redor del Sol*’ es totalmente probable, le otorgarás una puntuación de **1**. Para ello, **en el teclado**, deberás pulsar la tecla señalada con un 1 (*fíjate en las etiquetas*).

Después de esto, te aparecerá una pantalla en la que se te indica que para pasar a la siguiente afirmación deberás pulsar la barra espaciadora. Cuando la pulses, te aparecerá otra afirmación diferente para la cual deberás estimar su probabilidad, y así sucesivamente.

Primero realizarás 4 ensayos de prueba para acostumbrarte a la tarea; cuando finalices estos 4 ensayos debes **parar** hasta que el experimentador te lo indique.

Después de haber hecho el primer bloque de la tarea (más de la mitad del total), verás una pantalla indicándote que debes **parar** y realizar la tarea de lápiz y papel que tienes en tu mesa. Cuando la acabes, descansa en silencio hasta que el experimentador te dé las instrucciones del segundo bloque del experimento.

Al finalizar el experimento, es muy importante que no toques nada. El experimentador os dirá en qué momento podéis abandonar la sala.

Muchas gracias por tu colaboración.

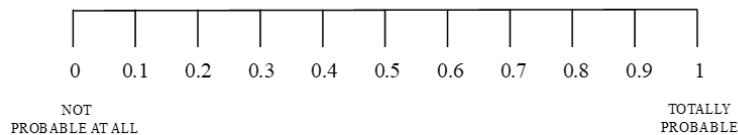
ENGLISH VERSION – *First Task Instructions: Premises Evaluation Task*

Next you will see a series of statements. Your task is to give them a score on an 11-point probability scale ranging from 0 (not probable at all) to 1 (totally probable) through 0, 0.1, 0.2... 0.9, up to 1. This score should reflect the degree in which you believe that the situations described in said statements can be (or are) true.

Example:

Please estimate the probability of occurrence of the following situation:

‘The planet Earth revolves around the Sun.’



If, for instance, you consider that the probability that *'the planet Earth revolves around the Sun'* is totally probable, you will give it a score of **1**. To do this, **on the keyboard**, you must press the key marked with a 1 (*look at the labels*).

After this, a screen will appear telling you that, to move on to the next statement, you must press the space bar. When you press it, another different statement will appear for which you must estimate its probability, and so on.

You will first do 4 practice trials to get used to the task; when you finish these 4 trials you must **stop** until the experimenter tells you to start with the first task.

After you've finished this first evaluation task, you'll see a screen telling you to **stop** and do the paper-and-pencil task on your desk. When you are done, rest quietly until the experimenter instructs you to start the next task of the experiment. Importantly, do not touch anything related with the computer.

Thank you very much for your cooperation.

SPANISH VERSION – *Second Task Instructions: Argument's Conclusion Evaluation Task*

Instrucciones:

A continuación, se te presentarán una serie de ítems compuestos por tres frases: las dos primeras frases te informan de un contexto determinado y la última frase podría derivarse de las dos anteriores. Por ejemplo, las oraciones que te informan de un contexto podrían ser como las siguientes:

‘Si aumenta la obesidad infantil, aumentará el riesgo de padecer enfermedades cardiovasculares.’

‘En la sociedad española está aumentando la obesidad infantil.’

Y la afirmación que podría derivarse de ellas sería:

‘En la sociedad española aumentará el riesgo de padecer enfermedades cardiovasculares.’

ENGLISH VERSION – *Second Task Instructions: Argument's Conclusion Evaluation Task*

Next you will be presented with a series of items made up of three sentences: the first two sentences inform you about a certain context and the last sentence is a statement that could be derived from the previous two sentences. For example, sentences informing you about a context could be as follows:

'If childhood obesity increases, the risk of cardiovascular disease will increase.'

'In Spanish society, childhood obesity is increasing.'

And the statement that could be derived from them would be:

'In Spanish society, the risk of suffering from cardiovascular diseases will increase.'

Obviously, for us, some or many of these statements will be more or less likely (either not at all, slightly, very or totally likely) and not so much a question of black and white.

Your task, for each item, is to read the two context sentences and **from there, assess how likely you think the statement derived from them is**, scoring them on the same 11-point scale as before, which ranges from 0 (not at all likely) to 1 (totally likely). Thus, this score should reflect the degree to which *you believe* that the situations described in those statements can be (or are) true, based on your degree of belief for the context sentences.

You must use the 'space bar' key both for each sentence to appear on the screen and to move on to the next item; and use the keys of the probability scale to score the statements. Let us see the procedure:

1st: press the key space bar and you will see the next on the screen:

"Given the probability that"

2nd: press the key space bar.

"If childhood obesity increases, the risk of cardiovascular disease will increase."

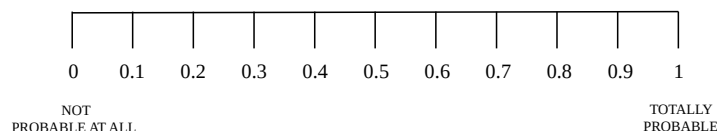
3rd: press the key bar.

"In Spanish society, childhood obesity is increasing."

4th: press the key space bar.

"Please, estimate the probability of occurrence of the following situation:"

'In Spanish society, the risk of suffering from cardiovascular diseases will increase.'



5th: press the key of the probability scale (on the keyboard) that best suits what you believe. For example, since the context sentences seem quite likely to me, I can believe that the latter is 0.7 out of 1 likely, and so I will press the key marked 0.7.

6th: press the 'space bar key' to pass to the next item.

As you know, there are no right or wrong answers, you just must make estimates based on what you believe.

When you have finished the whole experiment, it is very important that you do not touch anything. The experimenter will tell you when you can leave the room.

THANK YOU VERY MUCH FOR YOUR COOPERATION

Appendix B-II –Experiment 2 Materials

Initial Credibility Establishment Phase

Experimental Items (i.e., isolated conclusions)

1. La industria de la automoción contribuirá a mejorar el medioambiente.
2. La ley del aborto aumentará la probabilidad de que los niños sean más felices.
3. El parlamento español producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.
4. Una importante proporción de los niños de hoy acabará con problemas visuales.
5. En los países ricos muchos colegios tendrán que cerrar.
6. Con el nuevo sistema universitario de grados los alumnos verán menos materia y sabrán menos que antes.
7. China se verá en graves problemas económicos.
8. En nuestra sociedad disminuirán los problemas infantiles de espalda.
9. En España aumentará la proporción de personas con enfermedades degenerativas.
10. Los estereotipos sociales de los medios de comunicación pueden contribuir al aumento de embarazos no deseados.
11. Los niños y adolescentes de nuestra sociedad gozarán de relaciones sociales de menor calidad y más soledad.
12. Los medios de transporte contribuirán a la extinción de los osos polares.
13. El Partido Popular reducirá la producción de arroz en el conjunto del Estado.
14. Los partidos políticos progresistas facilitarán un aumento del número de delitos.
15. Una importante proporción de la juventud española de hoy sufrirá daños cerebrales y psicológicos mañana.
16. En la sociedad española puede aumentar el número de recién nacidos con deficiencias.
17. Tarragona tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud.
18. Para salir de la crisis será conveniente que el tráfico en las calles y carreteras empeore.
19. El empleo de los medios de transporte contribuirá, a la larga, a la inundación de las ciudades de la costa mediterránea.
20. Las empresas de informática contribuirán al aumento del desempleo.
21. En la población femenina española aumentará el número de casos con riesgo de cáncer.
22. La salida de la crisis económica conllevará esquilmar los recursos naturales y aumentar la cantidad de basura.
23. El PSOE facilitará el aumento de la desigualdad social.
24. Las universidades y empresas fomentarán la tala de bosques.

Filler items

1. El PSOE ganó las últimas elecciones al Gobierno de España.
2. En nuestra sociedad aumentará la edad de jubilación a 67 años de edad.
3. El descubrimiento oficial de América fue el 12 de octubre de 1492.

4. Las manzanas caen de los árboles debido al efecto de la gravedad.
5. Las personas que fuman aspiran humo de sus cigarros.
6. Un cuadrado tiene cuatro lados iguales unidos por esquinas de 90 grados.
7. La sociedad mediterránea se caracteriza por ser muy alegre y abierta.
8. Al meter una cerveza en el congelador durante más de 4 horas, se congelará y se estropeará.
9. De aquí a dos años muchos países del continente africano seguirán siendo pobres.
10. El año que viene la Sagrada Familia de Barcelona seguirá estando en obras.
11. La industria de la automoción fabricará coches que sean cada vez más seguros.
12. Varios islotes del archipiélago de las Baleares terminarán siendo de propiedad inglesa o alemana.
13. En algún momento de la historia de la humanidad, EE.UU. dejará de tener la hegemonía económica mundial.
14. La mayoría de grandes mamíferos salvajes se habrán extinguido ya en el siglo XXII.
15. Las reservas petrolíferas se agotarán en menos de 200 años.
16. Durante los próximos dos años Oriente Medio seguirá en conflictos armados.
17. En un futuro próximo todas las películas que se expongan en los cines serán en 3-D.
18. Aunque prohíban las páginas web de descargas ilegales (música, películas, etc.), el impuesto del canon digital seguirá existiendo.

Experimental Phase

Filler items used in the *Premises Evaluation Task*

Fillers for the warrant premise's block

High credibility

1. En un futuro cercano, los servicios que hoy son públicos pasarán a ser privados.
2. El próximo mes de julio la temperatura media será superior a 10 grados centígrados.
3. El año que viene la Sagrada Familia de Barcelona seguirá estando en obras.
4. Mañana, igual que hoy, las colchonetas de playa seguirán flotando en el agua.
5. La mayor parte de la carne de producción industrial proviene de animales que viven hacinados en pequeñas jaulas.

Low Credibility

1. Llegará un día en el que el idioma oficial de todas las poblaciones de la costa española sea el inglés.
2. En la época de Navidad los canales reducen la emisión de anuncios comerciales.
3. Los movimientos antisistema reivindican la extensión del capitalismo para los países pobres.
4. El tabaquismo es un movimiento cultural destinado a salvar vidas humanas.
5. La ley del aborto libre ha sido aprobada en todos los países de la Europa católica.

Fillers for the data premise's block

High credibility

1. Si se enciende el aire acondicionado muy fuerte, entonces se pasa frío
2. Si presiono el pedal del freno, el coche se detiene.
3. Si se decapita a un mamífero, entonces muere.
4. Si un ser vivo vuela, entonces tiene alas.
5. Si una persona come alimentos dietéticos, entonces perderá peso en un futuro próximo.

Low Credibility

1. Si todos los gobiernos del mundo prohibieran los combustibles fósiles mañana, el cambio climático se revertiría en una semana.
2. Si todas las ciudades de España se convirtieran en zonas peatonales, se resolverían todos los problemas de contaminación.
3. Si todos los políticos españoles fueran honestos, no habría corrupción en el país.
4. Si todos los ciudadanos dejaran de usar plásticos de un solo, no habría contaminación plástica en los mares.
5. Si se eliminara el IVA mañana, la economía florecería instantáneamente.

Arguments used in the *Argument's Conclusion Evaluation Task*

Condition HH: high credibility warrant premise and high credibility data premise.

1.
 - a) Si la mayoría de la gente usa el coche eléctrico, entonces se contribuirá a mejorar el medio ambiente.
 - b) La industria de la automoción contribuirá a que la mayoría de la gente use el coche eléctrico.
 - c) La industria de la automoción contribuirá a mejorar el medioambiente.
2.
 - a) Si sólo nacen niños deseados, entonces aumentará la probabilidad de que los niños sean más felices.
 - b) La ley del aborto aumenta la probabilidad de que sólo nazcan niños deseados.
 - c) La ley del aborto aumentará la probabilidad de que los niños sean más felices.
3.
 - a) Si se aprueba la ley del aborto, entonces se producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.
 - b) El parlamento español ha aprobado la ley del aborto.
 - c) El parlamento español producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.
- 4.

- a) Si uno pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.), acabará con problemas visuales.
 - b) Una importante proporción de los niños de hoy pasa demasiado tiempo al día delante de una pantalla (ordenador, televisión, etc.).
 - c) Una importante proporción de los niños de hoy acabará con problemas visuales.
- 5.
- a) Si disminuye la tasa de nacimientos de un país, entonces muchos colegios tendrán que cerrar.
 - b) En los países ricos está disminuyendo la tasa de natalidad.
 - c) En los países ricos muchos colegios tendrán que cerrar.
- 6.
- a) Si disminuye el número de créditos académicos, entonces los alumnos verán menos materia y sabrán menos que antes.
 - b) Con el nuevo sistema universitario de grados ha disminuido el número de créditos académicos respecto del plan anterior.
 - c) Con el nuevo sistema universitario de grados los alumnos verán menos materia y sabrán menos que antes.
- 7.
- a) Si en un país su población envejece, entonces se verá en graves problemas económicos.
 - b) En China la población está envejeciendo debido a que mantienen un estricto control de la natalidad.
 - c) China se verá en graves problemas económicos.
- 8.
- a) Si se extiende el uso mochilas escolares con ruedas, entonces disminuirán los problemas infantiles de espalda.
 - b) En nuestra sociedad se ha extendido el uso de mochilas escolares con ruedas.
 - c) En nuestra sociedad disminuirán los problemas infantiles de espalda.
- 9.
- a) Si la población de un país envejece, aumentará la proporción de personas con enfermedades degenerativas.
 - b) En España la población está envejeciendo.
 - c) En España aumentará la proporción de personas con enfermedades degenerativas.
- 10.

- a) Si se promueve el sexo entre adolescentes, se puede contribuir al aumento de embarazos no deseados.
- b) Los estereotipos sociales presentes en ciertos medios de comunicación promueven el sexo entre adolescentes.
- c) Los estereotipos sociales de los medios de comunicación pueden contribuir al aumento de embarazos no deseados.

11.

- a) Si se emplea el tiempo de ocio en la consola, el ordenador y en los chats, entonces se gozará de relaciones sociales de menor calidad y más soledad.
- b) Los niños y adolescentes de nuestra sociedad emplean el tiempo en la consola, el ordenador y los chats.
- c) Los niños y adolescentes de nuestra sociedad gozarán de relaciones sociales de menor calidad y más soledad.

12.

- a) Si se siguen emitiendo gases invernadero, entonces se contribuirá a la extinción de los osos polares.
- b) Los medios de transporte emiten gases invernadero.
- c) Los medios de transporte contribuirán a la extinción de los osos polares.

13.

- a) Si se produce el trasvase del Ebro, se reducirá la producción de arroz en el conjunto del Estado.
- b) El Partido Popular pretende realizar el trasvase del Ebro.
- c) El Partido Popular reducirá la producción de arroz en el conjunto del Estado.

14.

- a) Si aumenta la inmigración, entonces aumentará el número de delitos.
- b) Los partidos políticos progresistas facilitan un aumento de la inmigración.
- c) Los partidos políticos progresistas facilitarán un aumento del número de delitos.

15.

- a) Si alguien se droga frecuentemente hoy, entonces sufrirá daños cerebrales y psicológicos mañana.
- b) Una importante proporción de la juventud española de hoy se droga pronto y frecuentemente.
- c) Una importante proporción de la juventud española de hoy sufrirá daños cerebrales y psicológicos mañana.

16.

- a) Si aumenta el número de madres primerizas a edades cercanas a los 40 años, entonces puede aumentar el número de recién nacidos con deficiencias.
- b) En la sociedad española está aumentando el número de madres primerizas a edades cercanas a los 40 años.
- c) En la sociedad española puede aumentar el número de recién nacidos con deficiencias.

17.

- a) Si una ciudad está rodeada de petroquímicas, entonces tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud.
- b) Tarragona está rodeada de petroquímicas.
- c) Tarragona tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud.

18.

- a) Si la venta de automóviles aumenta, entonces el tráfico en las calles y carreteras empeorará.
- b) Para salir de la crisis es conveniente que aumente la venta de automóviles.
- c) Para salir de la crisis será conveniente que el tráfico en las calles y carreteras empeore.

19.

- a) Si continúa el calentamiento global, entonces, a la larga, se inundarán las ciudades de la costa mediterránea.
- b) El empleo de los medios de transporte contribuye a que continúe el calentamiento global.
- c) El empleo de los medios de transporte contribuirá, a la larga, a la inundación de las ciudades de la costa mediterránea.

20.

- a) Si continúa el desarrollo de las nuevas tecnologías y nueva maquinaria, entonces se producirá un aumento del desempleo.
- b) Las empresas de informática contribuyen al desarrollo de las nuevas tecnologías y nueva maquinaria.
- c) Las empresas de informática contribuirán al aumento del desempleo.

21.

- a) Si en un grupo aumenta el número de personas que fuma, entonces aumentará el número de casos con riesgo de cáncer.
- b) En la población femenina española está aumentando el número de personas que fuma.
- c) En la población femenina española aumentará el número de casos con riesgo de cáncer.

22.

- a) Si aumenta el consumo de masas, entonces se esquilmarán los recursos naturales y aumentará la cantidad de basura.
- b) La salida de la crisis económica conlleva aumentar el consumo de masas.
- c) La salida de la crisis económica conllevará esquilmar los recursos naturales y aumentar la cantidad de basura.

23.

- a) Si aumenta la inmigración, entonces aumentará la desigualdad social.
- b) El PSOE facilita el aumento de la inmigración.
- c) El PSOE facilitará el aumento de la desigualdad social.

24.

- a) Si se abusa del empleo de papel, se fomentará la tala de bosques.
- b) Las universidades y empresas abusan del empleo de papel.
- c) Las universidades y empresas fomentarán la tala de bosques.

Condition HL: high credibility warrant premise and low credibility data premise.

1.

- a) Si se preservan los espacios naturales, entonces se contribuirá a mejorar el medioambiente.
- b) La industria de la automoción contribuye a preservar los espacios naturales.
- c) La industria de la automoción contribuirá a mejorar el medioambiente.

2.

- a) Si se educa a los niños con afecto, entonces aumentará la probabilidad de que los niños sean más felices.
- b) La ley del aborto contribuye a educar a los niños con afecto.
- c) La ley del aborto aumentará la probabilidad de que los niños sean más felices.

3.

- a) Si cierran todas las empresas que fabrican anticonceptivos, entonces se producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.
- b) El parlamento español ha aprobado cerrar todas las empresas que fabrican anticonceptivos.
- c) El parlamento español producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.

4.

- a) Si uno pasa demasiado tiempo al día mirando directamente al Sol, entonces acabará con problemas visuales.

- b) Una importante proporción de los niños de hoy pasa demasiado tiempo al día mirando directamente al Sol.
 - c) Una importante proporción de los niños de hoy acabará con problemas visuales.
- 5.
- a) Si toda la población de un país emigra a otro lado, entonces muchos colegios tendrán que cerrar.
 - b) En los países ricos toda la población está emigrando a otro lado.
 - c) En los países ricos muchos colegios tendrán que cerrar.
- 6.
- a) Si se permite que se armen alborotos en clase, entonces los alumnos verán menos materia y sabrán menos que antes.
 - b) Con el nuevo sistema universitario de grados se permite que se armen alborotos en clase.
 - c) Con el nuevo sistema universitario de grados los alumnos verán menos materia y sabrán menos que antes.
- 7.
- a) Si un país no tiene infraestructura industrial, entonces se verá en graves problemas económicos.
 - b) China es un país que no tiene infraestructura industrial.
 - c) China se verá en graves problemas económicos.
- 8.
- a) Si se logra que los niños hagan mucho ejercicio físico, entonces disminuirán los problemas infantiles de espalda.
 - b) En nuestra sociedad se ha logrado que los niños hagan mucho ejercicio físico.
 - c) En nuestra sociedad disminuirán los problemas infantiles de espalda.
- 9.
- a) Si aumenta mucho la contaminación ambiental por aluminio y otros tóxicos, aumentará la proporción de personas con enfermedades degenerativas.
 - b) En España ha aumentado mucho la contaminación ambiental por aluminio y otros tóxicos.
 - c) En España aumentará la proporción de personas con enfermedades degenerativas.
- 10.
- a) Si se baja el nivel de calidad de las píldoras anticonceptivas, se puede contribuir al aumento de embarazos no deseados.
 - b) Los estereotipos sociales presentes en ciertos medios de comunicación promueven bajar el nivel de calidad de las píldoras anticonceptivas.

- c) Los estereotipos sociales de los medios de comunicación pueden contribuir al aumento de embarazos no deseados.

11.

- a) Si uno pasa el tiempo de ocio únicamente leyendo libros, entonces se gozará de relaciones sociales de menor calidad y más soledad.
- b) Los niños y adolescentes de nuestra sociedad pasan el tiempo de ocio únicamente leyendo libros.
- c) Los niños y adolescentes de nuestra sociedad gozarán de relaciones sociales de menor calidad y más soledad.

12.

- a) Si se favorece la caza furtiva de grandes mamíferos, entonces se contribuirá a la extinción de los osos polares.
- b) Los medios de transporte favorecen la caza furtiva de grandes mamíferos.
- c) Los medios de transporte contribuirán a la extinción de los osos polares.

13.

- a) Si en España se produce una gran sequía, se reducirá la producción de arroz en el conjunto del Estado.
- b) El Partido Popular pretende producir una gran sequía en España.
- c) El Partido Popular reducirá la producción de arroz en el conjunto del Estado.

14.

- a) Si aumenta la pobreza del país, entonces aumentará el número de delitos.
- b) Los partidos políticos progresistas facilitan el aumento de la pobreza del país.
- c) Los partidos políticos progresistas facilitarán un aumento del número de delitos.

15.

- a) Si alguien sufre un infarto cerebral, entonces sufrirá daños cerebrales y psicológicos mañana.
- b) Una importante proporción de la juventud española de hoy sufre infartos cerebrales.
- c) Una importante proporción de la juventud española de hoy sufrirá daños cerebrales y psicológicos mañana.

16.

- a) Si aumenta el número de mujeres que fuman y beben alcohol durante el embarazo, puede aumentar el número de recién nacidos con deficiencias.
- b) En la sociedad española está aumentando el número de mujeres que fuman y beben alcohol durante el embarazo.

- c) En la sociedad española puede aumentar el número de recién nacidos con deficiencias.

17.

- a) Si una ciudad se encuentra en una zona de alta actividad sísmica o volcánica, entonces tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud.
- b) Tarragona se encuentra en una zona de alta actividad sísmica o volcánica.
- c) Tarragona tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud.

18.

- a) Si llueve y nieva mucho, entonces el tráfico en las calles y carreteras empeorará.
- b) Para salir de la crisis es conveniente que llueva y nieve mucho.
- c) Para salir de la crisis será conveniente que el tráfico en las calles y carreteras empeore.

19.

- a) Si se produce un diluvio de meses de duración, entonces, a la larga, se inundarán las ciudades de la costa mediterránea.
- b) El empleo de los medios de transporte contribuye a que se produzcan diluvios de meses de duración.
- c) El empleo de los medios de transporte contribuirá, a la larga, a la inundación de las ciudades de la costa mediterránea.

20.

- a) Si aumenta mucho la población de un país, entonces se producirá un aumento del desempleo.
- b) Las empresas de informática contribuyen a que aumente mucho la población de un país.
- c) Las empresas de informática contribuirán al aumento del desempleo.

21.

- a) Si en un colectivo aumenta el número de personas que se expone diariamente a altas dosis de radiación, entonces aumentará el número de casos con riesgo de cáncer.
- b) En la población femenina española está aumentando el número de personas que se expone diariamente a altas dosis de radiación.
- c) En la población femenina española aumentará el número de casos con riesgo de cáncer.

22.

- a) Si disminuye el reciclaje, entonces se esquilmarán los recursos naturales y aumentará la cantidad de basura.
- b) La salida de la crisis económica conlleva la disminución del reciclaje.

- c) La salida de la crisis económica conllevará esquilmar los recursos naturales y aumentar la cantidad de basura.

23.

- a) Si se capitaliza y se privatizan los servicios que hoy son públicos, entonces aumentará la desigualdad social.
- b) El PSOE facilita la capitalización y privatización de los servicios que hoy son públicos.
- c) El PSOE facilitará el aumento de la desigualdad social.

24.

- a) Si se pretende aumentar las tierras de cultivo, se fomentará la tala de bosques.
- b) Las universidades y empresas pretenden aumentar las tierras de cultivo.
- c) Las universidades y empresas fomentarán la tala de bosques.

Condition LH: low credibility warrant premise and high credibility data premise.

1.

- a) Si aumenta la contaminación, entonces se contribuirá a mejorar el medioambiente.
- b) La industria de la automoción contribuye a que aumente la contaminación.
- c) La industria de la automoción contribuirá a mejorar el medioambiente.

2.

- a) Si se permite la destrucción de embriones humanos, entonces aumentará la probabilidad de que los niños sean más felices.
- b) La ley del aborto permite la destrucción de embriones humanos.
- c) La ley del aborto aumentará la probabilidad de que los niños sean más felices.

3.

- a) Si se retrasa la edad de jubilación a los 67 años, entonces se producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.
- b) El parlamento español ha aprobado retrasar la edad de jubilación a los 67 años.
- c) El parlamento español producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.

4.

- a) Si uno suele comer bollería industrial, entonces acabará con problemas visuales.
- b) Una importante proporción de los niños de hoy suele comer bollería industrial.
- c) Una importante proporción de los niños de hoy acabará con problemas visuales.

5.

- a) Si en un país la gente pasa de ser pobre a vivir bien, entonces muchos colegios tendrán que cerrar.
- b) En los países ricos la gente ha pasado de ser pobre a vivir bien.
- c) En los países ricos muchos colegios tendrán que cerrar.

6.

- a) Si se aplican programas de evaluación continuada en las aulas universitarias, entonces los alumnos verán menos materia y sabrán menos que antes.
- b) Con el nuevo sistema universitario de grados se aplican programas de evaluación continuada en las aulas universitarias.
- c) Con el nuevo sistema universitario de grados los alumnos verán menos materia y sabrán menos que antes.

7.

- a) Si aumentan las exportaciones de un país a nivel internacional, entonces se verá en graves problemas económicos.
- b) China está realizando cada vez más exportaciones a nivel internacional.
- c) China se verá en graves problemas económicos.

8.

- a) Si los niños pasan la mayor parte del tiempo sentados (colegio, ordenador, televisor), entonces disminuirán los problemas infantiles de espalda.
- b) En nuestra sociedad los niños pasan la mayor parte del tiempo sentados (colegio, ordenador, televisor).
- c) En nuestra sociedad disminuirán los problemas infantiles de espalda.

9.

- a) Si se consume bastante aceite de oliva, vino y pan, entonces aumentará la proporción de personas con enfermedades degenerativas.
- b) En España se consume bastante aceite de oliva, vino y pan.
- c) En España aumentará la proporción de personas con enfermedades degenerativas.

10.

- a) Si se promueve la discriminación de otras razas y etnias, se puede contribuir al aumento de embarazos no deseados.
- b) Los estereotipos sociales presentes en ciertos medios de comunicación promueven la discriminación de otras razas y etnias.

- c) Los estereotipos sociales de los medios de comunicación pueden contribuir al aumento de embarazos no deseados.

11.

- a) Si se está matriculado en algún colegio o instituto, entonces se gozará de relaciones sociales de menor calidad y más soledad.
- b) Los niños y adolescentes de nuestra sociedad están matriculados en algún colegio o instituto.
- c) Los niños y adolescentes de nuestra sociedad gozarán de relaciones sociales de menor calidad y más soledad.

12.

- a) Si se facilita la exportación e importación de productos, entonces se contribuirá a la extinción de los osos polares.
- b) Los medios de transporte facilitan la exportación e importación de productos.
- c) Los medios de transporte contribuirán a la extinción de los osos polares.

13.

- a) Si se privatiza gran parte de los servicios que ahora son públicos, se reducirá la producción de arroz en el conjunto del Estado.
- b) El Partido Popular pretende privatizar gran parte de los servicios que ahora son públicos.
- c) El Partido Popular reducirá la producción de arroz en el conjunto del Estado.

14.

- a) Si se facilitan ayudas para la dependencia o para el alquiler de la vivienda, entonces aumentará el número de delitos.
- b) Los partidos políticos progresistas facilitan ayudas para la dependencia o para el alquiler de la vivienda.
- c) Los partidos políticos progresistas facilitarán un aumento del número de delitos.

15.

- a) Si alguien vive con sus padres hoy, entonces sufrirá daños cerebrales y psicológicos mañana.
- b) Una importante proporción de la juventud española de hoy vive con sus padres.
- c) Una importante proporción de la juventud española de hoy sufrirá daños cerebrales y psicológicos mañana.

16.

- a) Si aumenta el número de personas que se someten a tratamientos de cirugía estética, entonces puede aumentar el número de recién nacidos con deficiencias.

- b) En la sociedad española está aumentando el número de personas que se someten a tratamientos de cirugía estética.
- c) En la sociedad española puede aumentar el número de recién nacidos con deficiencias.

17.

- a) Si una ciudad conserva muchas ruinas del Imperio Romano, entonces tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud.
- b) Tarragona conserva muchas ruinas del Imperio Romano.
- c) Tarragona tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud

18.

- a) Si aumenta el consumo, entonces el tráfico en las calles y carreteras empeorará.
- b) Para salir de la crisis es conveniente que aumente el consumo.
- c) Para salir de la crisis será conveniente que el tráfico en las calles y carreteras empeore.

19.

- a) Si la gente viaja con facilidad y rapidez, entonces, a la larga, se inundarán las ciudades de la costa mediterránea.
- b) El empleo de los medios de transporte contribuye a que la gente pueda viajar con facilidad y rapidez.
- c) El empleo de los medios de transporte contribuirá, a la larga, a la inundación de las ciudades de la costa mediterránea.

20.

- a) Si se desarrollan programas de software que permiten llevar a cabo cálculos complejos, entonces se producirá un aumento del desempleo.
- b) Las empresas de informática contribuyen al desarrollo de programas de software que permiten llevar a cabo cálculos complejos.
- c) Las empresas de informática contribuirán al aumento del desempleo.

21.

- a) Si en un colectivo aumenta el número de personas que estudian carreras relacionadas con el cuidado de los demás (educación, enfermería, psicología, etc.), entonces aumentará el número de casos con riesgo de cáncer.
- b) En la población femenina española hay muchas personas que estudian carreras relacionadas con el cuidado de los demás (educación, enfermería, psicología, etc.).

- c) En la población femenina española aumentará el número de casos con riesgo de cáncer.

22.

- a) Si se genera empleo y mejoran las condiciones laborales, entonces se esquilmarán los recursos naturales y aumentará la cantidad de basura.
- b) La salida de la crisis económica conlleva generar empleo y mejorar las condiciones laborales.
- c) La salida de la crisis económica conllevará esquilmar los recursos naturales y aumentar la cantidad de basura.

23.

- a) Si se facilitan ayudas para la dependencia o para el alquiler de la vivienda, entonces aumentará la desigualdad social.
- b) El PSOE facilita ayudas para la dependencia o para el alquiler de la vivienda.
- c) El PSOE facilitará el aumento de la desigualdad social.

24.

- a) Si se contribuye al desarrollo humano y económico de la sociedad, se fomentará la tala de bosques.
- b) Las universidades y empresas contribuyen al desarrollo humano y económico de la sociedad.
- c) Las universidades y empresas fomentarán la tala de bosques.

Condition LL: low credibility warrant premise and low credibility data premise.

1.

- a) Si se talan los bosques, entonces se contribuirá a mejorar el medioambiente.
- b) La industria de la automoción contribuye a la tala de bosques.
- c) La industria de la automoción contribuirá a mejorar el medioambiente.

2.

- a) Si aumentan los casos de cáncer de mama, entonces aumentará la probabilidad de que los niños sean más felices.
- b) La ley del aborto contribuye al aumento de los casos de cáncer de mama.
- c) La ley del aborto aumentará la probabilidad de que los niños sean más felices.

3.

- a) Si las farmacias dispensan anticonceptivos de forma gratuita, entonces se producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.
 - b) El parlamento español ha aprobado que las farmacias dispensen anticonceptivos de forma gratuita.
 - c) El parlamento español producirá indirectamente una disminución del uso de anticonceptivos por parte de muchos jóvenes.
- 4.
- a) Si uno come muchas zanahorias, entonces acabará con problemas visuales.
 - b) Una importante proporción de los niños de hoy come muchas zanahorias.
 - c) Una importante proporción de los niños de hoy acabará con problemas visuales.
- 5.
- a) Si aumenta la caza de elefantes, entonces muchos colegios tendrán que cerrar.
 - b) En los países ricos está aumentando la caza de elefantes.
 - c) En los países ricos muchos colegios tendrán que cerrar.
- 6.
- a) Sin aumenta el número de tizas y pizarras, entonces los alumnos verán menos materia y sabrán menos que antes.
 - b) Con el nuevo sistema universitario de grados ha aumentado el número de tizas y pizarras.
 - c) Con el nuevo sistema universitario de grados los alumnos verán menos materia y sabrán menos que antes.
- 7.
- a) Si un país dispone de grandes pozos petrolíferos, entonces se verá en graves problemas económicos.
 - b) China dispone de grandes pozos petrolíferos.
 - c) China se verá en graves problemas económicos.
- 8.
- a) Si se extiende el trabajo y la explotación infantil, entonces disminuirán los problemas infantiles de espalda.
 - b) En nuestra sociedad se ha extendido el trabajo y la explotación infantil.
 - c) En nuestra sociedad disminuirán los problemas infantiles de espalda.
- 9.
- a) Si en un país la mayoría de la gente practica yoga, aumentará la proporción de personas con enfermedades degenerativas.

- b) En España la mayoría de la gente practica yoga.
- c) En España aumentará la proporción de personas con enfermedades degenerativas.

10.

- a) Si se promueve el terrorismo armado, se puede contribuir al aumento de embarazos no deseados.
- b) Los estereotipos sociales presentes en los medios de comunicación promueven el terrorismo armado.
- c) Los estereotipos sociales de los medios de comunicación pueden contribuir al aumento de embarazos no deseados.

11.

- a) Si uno juega mucho al ajedrez con sus amigos, entonces se gozará de relaciones sociales de menor calidad y más soledad.
- b) Los niños y adolescentes de nuestra sociedad juegan mucho al ajedrez con sus amigos.
- c) Los niños y adolescentes de nuestra sociedad gozarán de relaciones sociales de menor calidad y más soledad.

12.

- a) Si se incrementan las ventas de telefonía móvil, entonces se contribuirá a la extinción de los osos polares.
- b) Los medios de transporte incrementan las ventas de telefonía móvil.
- c) Los medios de transporte contribuirán a la extinción de los osos polares.

13.

- a) Si un gobierno aumenta la inversión en ayudas y obras sociales, se reducirá la producción de arroz en el conjunto del Estado.
- b) El Partido Popular pretende aumentar la inversión en ayudas y obras sociales.
- c) El Partido Popular reducirá la producción de arroz en el conjunto del Estado.

14.

- a) Si se expulsa a los inmigrantes, entonces aumentará el número de delitos.
- b) Los partidos políticos progresistas facilitan la expulsión de los inmigrantes.
- c) Los partidos políticos progresistas facilitarán un aumento del número de delitos.

15.

- a) Si alguien lee muchos libros hoy, entonces sufrirá daños cerebrales y psicológicos mañana.
- b) Una importante proporción de la juventud española de hoy lee muchos libros.

- c) Una importante proporción de la juventud española de hoy sufrirá daños cerebrales y psicológicos mañana.

16.

- a) Si aumenta el número de puestos de trabajo, entonces puede aumentar el número de recién nacidos con deficiencias.
- b) En la sociedad española está aumentando el número puestos de trabajo.
- c) En la sociedad española puede aumentar el número de recién nacidos con deficiencias.

17.

- a) Si una ciudad recibe muy poco turismo playero, entonces tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud.
- b) Tarragona recibe muy poco turismo playero.
- c) Tarragona tiene más probabilidades que otras ciudades de que se produzca un accidente con graves consecuencias para la salud.

18.

- a) Si aumenta la deuda pública del Estado, entonces el tráfico en las calles y carreteras empeorará.
- b) Para salir de la crisis es conveniente que aumente la deuda pública del Estado.
- c) Para salir de la crisis será conveniente que el tráfico en las calles y carreteras empeore.

19.

- a) Si se produce una gran sequía, entonces, a la larga, se inundarán las ciudades de la costa mediterránea.
- b) El empleo de los medios de transporte contribuye a que se produzcan grandes sequías.
- c) El empleo de los medios de transporte contribuirá, a la larga, a la inundación de las ciudades de la costa mediterránea.

20.

- a) Si aumenta la demanda en de bienes y productos, entonces se producirá un aumento del desempleo.
- b) Las empresas de informática contribuyen al aumento de la demanda de bienes y productos.
- c) Las empresas de informática contribuirán al aumento del desempleo.

21.

- a) Si en un sector poblacional aumenta drásticamente el número de ingenieros informáticos, entonces aumentará el número de casos con riesgo de cáncer.

- b) En la población femenina española ha aumentado drásticamente el número de ingenieros informáticos.
- c) En la población femenina española aumentará el número de casos con riesgo de cáncer.

22.

- a) Si se prohíbe el toreo, entonces se esquilmarán los recursos naturales y aumentará la cantidad de basura.
- b) La salida de la crisis económica conlleva prohibir el toreo.
- c) La salida de la crisis económica conllevará esquilmar los recursos naturales y aumentar la cantidad de basura.

23.

- a) Si todo el mundo fuma, entonces aumentará la desigualdad social.
- b) El PSOE facilita que todo el mundo fume.
- c) El PSOE facilitará el aumento de la desigualdad social.

24.

- a) Si se deja de emplear el papel, se fomentará la tala de bosques.
- b) Las universidades y empresas han dejado de emplear el papel.
- c) Las universidades y empresas fomentarán la tala de bosques.

Appendix B-III - Unexpected Results and Alternative Interpretations about the Difference HL > LH

We will thoroughly investigate the nature of the discrepancies between these conditions to determine if specific factors may explain the observed differences in results. If the values of H were identical in both conditions (HL and LH), as well as those of L, and the multiplication hypothesis fully accounted for the results, we should not observe differences between the HL and LH conditions.

As the credibility values of the premises were empirically obtained, we explored the possibility that differences in the conclusions between both conditions might be attributed to differences already present in the credibility of their respective original premises. To investigate this, *post hoc* analyses were conducted. We compared the means of H premises between HL and LH ($t_1(63) = -2$, $p = .05$; $t_2(23) = -.8$, $p = .434$) and the means of L premises ($t_1(63) = 3$, $p = .004$; $t_2(23) = -1.44$, $p = .162$). We observed differences among subjects but not among items. In any case, the observed difference in H cannot account for the finding HL > LH, as it goes in the opposite direction (i.e., mean credibility of the H premise in HL is lower than that of the H premise in LH). The difference found in L does go in the expected direction (i.e., L in HL is greater than L in LH). However, it is not clear whether this difference in L, found only in the subject-level analysis, can explain the HL > LH finding obtained by subjects and items. If this difference in L resulted in an increase in the mean credibility of H and L premises in HL compared to LH, it could potentially explain the difference between HL and LH. However, no significant differences were observed or, at most, marginally significant at the subject level ($t_1(63) = 1.75$, $p = .085$; $t_2(23) = .6$, $p = .556$).

Finally, the possibility of differences between the scores *predicted* by HL and those *predicted* by LH was explored. Indeed, if the finding HL > LH was exclusively due to the credibility of the premises, then their multiplication should result in distinct credibilities, replicating the finding. The mean of the multiplication in HL is 0.2, and in LH, it is 0.15, where the direction of the difference is consistent with the HL > LH finding. However, the results were not clear as differences were found by subjects but not by items analyses ($t_1(63) = 2.84$, $p = .006$; $t_2(23) = 1.25$, $p = .221$).

Thus, the results of these *post hoc* analyses have not provided clear outcomes, and it cannot be asserted that the cause of the difference between HL and LH is simply due to circumstantial empirical issues. Subject-level analyses are consistent with the possibility that the HL > LH difference is due to higher initial values in HL (e.g., a higher value of L, a marginally higher average of the premises, and a higher average of the scores predicted by the multiplication). However, none of these was observed in item-level analyses. Therefore, these analyses do not allow us to distinguish whether these differences are due to empirical issues or to other theoretical issues, such as the different functional nature of the premises. Thus, further research would be needed to elucidate this.

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

APÉNDICE C: FUNDAMENTOS TEÓRICOS - entimemas, tópicos y esquemas de argumentación

UNIVERSITAT ROVIRA I VIRGILI

DEDUCTIVE BELIEF DYNAMICS: THEORETICAL FOUNDATIONS AND EXPERIMENTAL STUDY OF AN ARGUMENTATIVE MODEL
OF THE BELIEF SYSTEM

Daniel Rivera Guajardo

ENTIMEMAS Y ESQUEMAS DE ARGUMENTACIÓN

A lo largo de la primera parte de este apéndice realizaré una presentación del concepto de entimema describiendo su origen, sus raíces retóricas, epistemológicas y lógicas, así como sus partes, las funciones de cada parte, las precondiciones para su efectividad y su modo de funcionamiento global.

La segunda parte de este apéndice tendrá por objeto la naturaleza y estructura de la premisa mayor y sus diversos niveles de abstracción, centrándonos en los esquemas de argumentación y las relaciones tópicas. En tanto que la relación tópica con la que trabajaré en todos los experimentos de esta tesis es la denominada “de la causa al efecto”, dedicaré un apartado especial en este apéndice para explicarla brevemente y mostrar la conexión directa que guarda con los estudios sobre razonamiento causal realizados en psicología expuestos a partir de la sec. 14, cap. 4, pt. I. Finalmente, cerraré el apéndice con un apartado que pretende establecer una serie de relaciones entre los conceptos cognitivos de esquemas y relaciones de coherencia con los conceptos retórico-argumentativos de esquemas de argumentación y relaciones tópicas además del papel que juega la probabilidad subjetiva como cuantificación de la magnitud de la justificación transferida mediante esas relaciones.

Introducción

Como es palmario que el método propio del arte es el que se refiere a las pruebas por persuasión y que la persuasión es una especie de demostración (puesto que nos persuadimos sobre todo cuando pensamos que algo ha quedado demostrado); como, por otra parte, la demostración retórica es el entimema y éste es, hablando en absoluto, la más firme de las pruebas por persuasión; y como el entimema, en fin, es un silogismo [...], resulta evidente que el que mejor pueda teorizar a partir de qué y cómo se produce el silogismo, ése será también el más experto en entimemas, con tal que llegue a comprender sobre qué materias versa el entimema y qué diferencias tiene respecto de los silogismos lógicos. Porque corresponde a una misma facultad reconocer lo *verdadero* y lo *verosímil* [...], estar en disposición de discernir sobre lo *plausible* es propio de quien está en la misma disposición con respecto a la *verdad*.

Aristóteles, *Rhet.* I, 1.4

Esta cita condensa en pocas palabras lo que va a desarrollarse en este apéndice. Tal y como expresaron Van Eemeren y Grootendorst (2003; sec. 3.1., cap. 1, pt. I), el conocimiento de la Retórica de Aristóteles es una condición necesaria en la teoría de la argumentación actual en tanto que algunas de sus concepciones teóricas son una parte integral de la argumentación y la evaluación del discurso argumentativo. Este comentario cobra especial relevancia en este apéndice, que desarrolla el concepto de entimema como herramienta retórica para el cambio de creencias. Herramienta, por otro lado, esencial en el desarrollo de la presente investigación en sus planos teórico y empírico.

Este apéndice dedicado a los entimemas recoge y complementa las cuestiones ya tratadas en las sec. 1.2., 3.1., 5., cap.1 y sec. 3, cap.2, ambos de pt. I. Antes de adentrarnos en los aspectos estructurales más relevantes para la investigación sobre la credibilidad que se aborda en esta tesis, se presentará el contexto en el que surge su teorización. Por otro lado, los entimemas genuinos tienen algunas características propias que no serán atendidas en este trabajo. Es por esto por lo que parece apropiado mencionarlas aquí para obtener una idea cabal de la dimensión del concepto original.

Raíces retóricas y dialécticas de los entimemas

Aristóteles desarrolla la conceptualización del entimema en el seno de sus estudios sobre retórica. Desde el momento en que los seres humanos empiezan a intercambiar razones para influir sobre las creencias y las acciones de los demás están empleando mecanismos de argumentación. En sociedades más avanzadas la dependencia de las habilidades argumentativas es mayor, pues más cuestiones pasan a dirimirse argumentando y persuadiendo a los interlocutores. Tanto es así, que en la antigua sociedad ateniense encontramos ya lugares destinados al debate y discurso públicos, en la colina del Pnyx, en el Areópago o en el Ágora. En esos lugares se deliberaba acerca de los beneficios o perjuicios futuros que pudiera comportar la adopción de nuevas leyes o acciones públicas. Las personas reunidas en el consejo participaban de la discusión aconsejando o disuadiendo en la toma de decisiones. Por otra parte, también eran lugares para despachar asuntos forenses o judiciales, discutiendo sobre lo justo o lo injusto de ciertas acciones pasadas, defendiendo o atacando al supuesto agente en consecuencia y decidiendo sobre su inocencia o culpabilidad y el grado de la pena. Finalmente, eran lugares para la exposición de panegíricos o diatribas, elogiando o censurando a alguna figura pública en función de que sus acciones pudieran ser presentadas como bellas y apreciables o vergonzosas y despreciables ante una audiencia que evalúa y juzga. En una sociedad así, era menester el desarrollo de un cuerpo de conocimiento teórico que sistematizara y analizara todo aquel conocimiento práctico que venía ejercitándose desde muy antiguo. Aristóteles viene a cubrir este vacío, estructurando los elementos de la retórica de tal modo que formasen una unidad inteligible. No solo muestra la adecuación de las distintas técnicas a las distintas situaciones, sino que dota a la retórica de unas bases teóricas filosóficas que permiten entender por qué funciona y cuáles son sus fundamentos más profundos.

La retórica se constituye como “la facultad de teorizar lo que es adecuado en cada caso para convencer” (*Ret.* 1355, b, 25), preferentemente, a través del discurso. Así pues, en principio, la retórica no versa sobre ningún género definido, sino que se aplica a “cada caso”. Saber retórica no le hace a uno especialista en ninguna materia particular. Sus principios se pueden aplicar a cualquier ámbito, ya sean cuestiones de física o de justicia. Tales principios derivan de procedimientos generales al margen de las temáticas particulares. Esos procedimientos sirven para conducir el razonamiento acerca de cualquier contenido particular en orden a lograr la persuasión. En cualquier caso, dado que el terreno tradicional en el que se mueve es el discurso público, los géneros a los

que se aplica han solido quedar constituidos por el deliberativo, el forense y el epidíctico. Pero no hay un límite potencial.

La función de la retórica es la de proporcionar razones a una audiencia determinada que justifiquen *por qué* debería sostener ciertas opiniones (Oaksford y Chater, 2007). Una representación apropiada de la audiencia por parte del hablante resulta indispensable para adecuar la forma y el contenido del discurso. El tipo de audiencia puede variar en función de la implicación por la temática a tratar o del nivel cultural, o ambas cosas. En este sentido, es relevante tener en cuenta cuáles son sus conocimientos o creencias previas, así como sus intereses, objetivos, valores y actitudes en orden a adaptar así el discurso. Las personas se persuaden mediante vías tanto emocionales como racionales. Y ambas se pueden articular a través del discurso.

Así, el lenguaje puede funcionar como un signo que indique a la audiencia el talante del orador (i.e. *ethos*). Mediante una adecuada representación de la audiencia, el hablante sabe que, si dice esto y aquello, y si emplea determinados términos, entonces logrará que la audiencia infiera que es una persona honesta que sabe de lo que habla. Una persona digna de crédito. Por otro lado, el contenido y la forma del discurso también pueden agitar las emociones de la audiencia (i.e. *pathos*). El papel de la audiencia es el de aceptar o rechazar; y no se acepta ni se rechaza lo mismo estando en un estado emocional positivo que negativo. Además, las finalidades de todos los géneros de discurso retórico tienen un polo positivo (lo beneficioso, lo justo, lo bello) y otro negativo (lo perjudicial, lo injusto y lo vergonzoso). Es relevante conocer las distintas combinaciones de las polaridades emocionales y de finalidad en relación con la aceptación y el rechazo de una aserción. Y todo esto en función de los distintos contextos situacionales de emisión del discurso. Los recursos afectivos no pueden ser considerados como elementos meramente auxiliares de la persuasión. Nuevamente, para manejarlos adecuadamente es necesaria una representación precisa de la audiencia. En este sentido, es propio de la retórica emplear enunciados afectivos como premisas de un razonamiento, que operan en la formación del juicio de un modo no demostrativo (Racionero, 1990). Estas serían pues las vías emocionales a la persuasión.

Finalmente, las personas también se persuaden por el discurso mismo (i.e. *logos*) en virtud de lo que éste demuestra o parece demostrar. El logos posee una doble vertiente. En primer lugar, el logos se puede entender como un cuerpo de enunciados no-afectivos que denota aquello que puede resultar más convincente respecto a cada uno de los asuntos tratados en los diferentes géneros discursivos. Son enunciados que apelan a los fines de cada género del discurso retórico (i.e. lo beneficioso y perjudicial; lo justo y lo injusto; lo bello y lo vergonzoso) y se organizan según una cierta lógica de preferencias. En segundo lugar, el logos contiene una vertiente fundamental de razonamiento cuasi demostrativo. La persuasión se concibe como una especie de demostración en tanto que uno se ve persuadido acerca de algo cuando considera que ha quedado demostrado. En este sentido, el discurso convence en función de que ofrezca cierto tipo de pruebas respecto de la tesis que pretende defender. También para esto será menester partir de una adecuada representación de la audiencia. El hablante debe saber cuáles son las cuestiones ya aceptadas por

la audiencia para emplearlas como premisas plausibles en un razonamiento. En todo caso, esto lo veremos más adelante, pues es este aspecto cuasi demostrativo el que se desarrolla en esta tesis. El logos constituiría la vía racional a la persuasión.

Además de estas características fundamentales, la retórica tiene una vertiente expresiva. Su función es la de prescribir el mejor modo lingüístico de presentar los contenidos. No basta con saber lo que hay que decir, sino que también es propio de la retórica especificar cómo debe decirse teniendo en cuenta las características de la audiencia. Así, se tratan cuestiones como la selección léxica, el uso de metáforas y otras figuras retóricas, la longitud de las oraciones, la prosodia, lo que se deja implícito y lo que se hace explícito u ostensivo y la organización estructural de las partes del discurso. Todo ello al servicio de la producción de un discurso elegante. Un discurso elegante es bello, proporcionado, ordenado, escueto, claro y preciso de un modo tal que facilita la comprensión. Frente al aburrimiento que causa lo obvio y a la confusión que causa algo difícil de entender, lo que se aprende con facilidad resulta agradable.

Todas estas cosas se montan sobre unas estructuras silogísticas que posibilitan que el razonamiento y las emociones fluyan a través del discurso. Estas estructuras son los entimemas, considerados por Aristóteles como “el cuerpo de la persuasión”. En efecto, según Racionero (1990), los entimemas o silogismos retóricos son las estructuras sobre las que discurren los enunciados al servicio del ethos, del pathos y del logos relativos a la materia particular que sea. En los entimemas, estos enunciados se expresan siguiendo las prescripciones lingüísticas de proporción, brevedad y claridad expositiva. Su estructura silogística le otorga la capacidad de funcionar como una especie de demostración a partir de razonamientos plausibles, siempre basados en las creencias previas de la audiencia.

Raíces epistemológicas: los hechos, lo posible, lo probable y lo plausible

En los capítulos 2 y 3 de la parte I vimos cómo el marco en el que se mueve el razonamiento retórico es el de la posibilidad. Sólo se persuade de lo posible. Dentro de lo posible, hay cosas que suceden siempre y otras la mayoría de las veces, pero pueden no suceder. Lo que se sabe que ocurre la mayoría de las veces es lo probable. Y cuando lo probable encarna una opinión común aceptada por la comunidad epistémica, se le llama plausible. Así, algo puede resultar plausible sólo si es posible y probable. Lo plausible constituye el contenido de las premisas de los entimemas. Por otra parte, hay cosas que, a primera vista, pueden parecer improbables a determinados individuos o grupos. La retórica pretende establecer conclusiones sobre sucesos desconocidos considerados improbables mediante la remisión a los hechos conocidos o aceptados. Para ello, aplica ciertas reglas generales de relación plausibles entre el hecho establecido y el hecho incierto pero posible, de tal modo que se vuelva probable. Las estructuras del discurso que permiten estos procesos en el razonamiento son los entimemas.

Pese al abordaje profundamente filosófico sobre los fundamentos del fenómeno persuasivo que encontramos en la Retórica, esta obra es, por lo demás, de un marcado carácter práctico. Su objeto de estudio no es meramente el cambio de creencias, sino la persuasión entendida como acción social a través del discurso público. No se persuade sobre cualquier cosa, sino sobre cuestiones que requieren la implicación activa de la audiencia: ir o no a la guerra, subir o no los impuestos, declarar a alguien culpable o inocente, desterrarle o encerrarle, declarar a alguien ciudadano ejemplar o persona non grata, digno o indigno de ser elegido cargo público, etc. Se persuade sobre aquello en lo que a la audiencia le es *posible* participar para producir un cambio efectivo en la sociedad. La retórica es una herramienta relativa a los razonamientos que se fundamentan en el terreno de lo posible en general, y en el terreno de lo posible entendido como objeto de la acción humana en particular. A través del discurso, se produce una interacción entre la mente y el mundo que rodea a los participantes de la situación comunicativa.

Como se ha visto, los fines de los géneros discursivos tradicionales conllevan siempre una evaluación sobre la bondad de lo tratado. En este sentido, el proceso persuasivo no se detiene en establecer que algo sucedió, o que sucede o que sucederá. En tanto que el suceso no es neutro, sino que es susceptible de ser valorado como deseable o repudiable, habrá que actuar en consecuencia. Esta descripción del fenómeno persuasivo consta entonces de tres vertientes: una, cuasi demostrativa, cuya función es la de hacer probable una *conjetura fáctica* ante la audiencia. Otra, afectiva, cuya función es la *evaluación* de la valencia hedónica del posible suceso de tal modo que pueda provocar determinadas emociones en la audiencia. La última, conativa o conductual, cuya función es la de inducir una *intención* o instigar una *acción*. Y, desde luego, la faceta cuasi demostrativa se puede extender más allá de la conjetura fáctica. Si es necesario, puede dar razones para mostrar que el posible suceso es deseable o repudiable. Asimismo, puede dar razones para mostrar que, del abanico de posibles conductas a su respecto, una es mejor que otra. Así pues, se pueden dar razones para creer en el supuesto suceso, para evaluarlo de un modo u otro, y para emprender unas acciones u otras.

De los entimemas Genuinos a Nuestros Entimemas

Como dije al inicio de este apéndice, varios de los rasgos comentados hasta aquí no serán atendidos en esta tesis. Algunas de las características de los entimemas nos servirán como herramienta para investigar sobre el cambio de creencias, pero no todas ellas. Así pues, pese a su relevancia en la práctica ecológica de la persuasión, no se atenderá a nada relativo al ethos, al pathos, ni al logos entendido como asunto de algún género discursivo concreto. Tampoco se atenderá a ninguna de las prescripciones lingüísticas mencionadas más arriba relativas al “arte del bien decir”. Por último, en tanto que el objeto de la presente investigación se centra en el cambio de creencias y no en el proceso persuasivo del cual podría llegar a formar parte, se deja fuera de esta tesis la vertiente afectivo-evaluativa relativa a la polaridad de los fines, así como la vertiente conativa dirigida a la toma de decisiones de cara a la acción.

Nos centraremos, fundamentalmente, en los aspectos funcionales probativos de su estructura argumental, pero no tratándolos como silogismos lógicos analíticos aislados, sino como silogismos probabilísticos basados en las consideraciones subjetivas de la audiencia. Así pues, del cuerpo de la persuasión, nos quedaremos con su esqueleto.

Raíces lógicas: los silogismos

En tanto que la base de los entimemas son los silogismos, se expondrá brevemente en qué consisten los silogismos, y de qué maneras se pueden presentar. Un silogismo designa, a nivel general, un razonamiento, inferencia o argumento; a un nivel más específico, se refiere a un razonamiento lógico deductivamente válido que establece la unión o la separación de dos términos en un juicio expresado por un enunciado aseverativo. Tal y como quedó presentado en el capítulo 1 de la parte I, el silogismo es un argumento que se compone de *tres proposiciones*: dos premisas y una conclusión, la última de las cuales se deriva de las otras dos mediante una relación de consecuencia lógica. Por ejemplo:

Proposición 1: Todos los hombres son mortales.

Proposición 2: Sócrates es un hombre.

Proposición 3: Sócrates es mortal.

Cada proposición se descompone en dos términos diferentes, habiendo un total de *tres términos* en el silogismo (i.e. mortales, hombres, Sócrates). Esos términos se suelen representar mediante letras como, por ejemplo, A, B y C. Tradicionalmente, se ha considerado que estos términos pueden variar en generalidad de tal modo que $A \geq B \geq C$. Las dos premisas del silogismo tienen un término en común, al que se ha llamado *término medio* (p. ej. hombre). Los otros dos términos reciben el nombre de *extremos*. Así, cada premisa relaciona un término extremo con el término medio, con el que se *compara*. Como resultado, en la conclusión desaparece el término medio, y aparecen relacionados los términos extremos de un modo que no lo estaban en las premisas. Los extremos se conectan en la conclusión en función de su conexión con el término medio en las premisas. Cuando resulta en un enunciado afirmativo, los términos se presentan unidos, y cuando es negativo, separados. De este modo, el razonamiento silogístico funciona como un *método de establecer relaciones entre términos*.

Por otra parte, los dos términos de cada proposición hacen la función de sujeto y predicado. En este sentido, el término extremo que conforma el *predicado de la conclusión* recibe el nombre de *término mayor* (p. ej. mortal) y el que conforma el *sujeto de la conclusión*, el de *término menor* (p. ej. Sócrates). Así, la premisa que contiene el predicado de la conclusión recibe el nombre de *premisa mayor*, y la que contiene el sujeto de la conclusión recibe el nombre de *premisa menor*. Por su parte, el término medio puede jugar el papel tanto de sujeto como de predicado de las premisas. Esto puede ocurrir de tres modos:

- a) Puede ser el sujeto de una premisa y el predicado de otra.
- b) Puede ser el predicado de las dos premisas.
- c) Puede ser el sujeto de las dos premisas.

De estas tres posibilidades surgen tres esquemas silogísticos o *figuras*:

	Primera Figura		Segunda Figura		Tercera Figura	
	Sujeto	Predicado	Sujeto	Predicado	Sujeto	Predicado
Premisa mayor	B	A	B	A	C	A
Premisa menor	C	B	C	A	C	B
Conclusión	C	A	C	B	B	A

La lógica silogística es una lógica de enunciados, y los enunciados están compuestos de un predicado y del sujeto de la predicación. El nombre que reciben las premisas está en función del nivel de generalidad del papel gramatical de los términos de la conclusión. Es frecuente encontrar representados esos términos, no ya por las letras A, B y C (atendiendo a su grado de generalidad), sino por las de P, M y S, donde P y S se corresponden respectivamente con el predicado y el sujeto de la conclusión, y M con el término medio de las premisas (atendiendo ahora a su rol funcional). Así pues, la tabla anterior también se podría representar sustituyendo esas letras por las correspondientes P, M, S pertinentemente. En tanto que el razonamiento silogístico establece relaciones entre términos, y esos términos tienen los roles de sujeto y predicado, *las conclusiones de los silogismos consisten en afirmar o negar que P sea predicable o atribuible al sujeto S*. Consisten en una demostración de la verdad de esa aseveración.

Cada uno de los tres enunciados del silogismo puede darse de seis modos distintos en función de que se afirme o se niegue de forma universal, particular o indefinida. Esto se muestra en la siguiente tabla (donde A y B representan, en este caso, cualquier término):

	Universal	Particular	Indefinida
Afirmación	Todo A es B	Algún A es B	A es B
Negación	Ningún A es B	Algún A no es B	A no es B

Así lo universal se predica afirmativa o negativamente del todo, lo particular, de la parte y lo indefinido simplemente se predica afirmativa o negativamente, pero sin indicar lo universal ni lo particular; esto es, carece de cuantificador. Esta característica lo hace asimilable tanto a las particulares como a las universales. Esta ambigüedad se adecua bien para su aplicación al marco de la posibilidad en razonamientos no concluyentes, pero lo convierte en inadecuado para el razonamiento demostrativo concluyente. No permite hacer aseveraciones concluyentes sobre su verdad o la falsedad. Es por esto por lo que no ha sido muy atendido por la tradición más formal de la silogística.

Validez deductiva

A partir de aquí se pueden analizar sistemáticamente todas las combinaciones posibles de las dos premisas en cada una de las tres figuras en cada uno de sus distintos modos definidos. En virtud de una serie de reglas complementarias, se puede demostrar entonces que la conclusión se sigue necesariamente de las premisas, o que no se sigue ninguna conclusión de ellas. Esto, a su vez, permite distinguir conclusiones correctas obtenidas deductivamente de conclusiones falaces. Y es que una deducción es “un discurso en el que, sentadas ciertas cosas, se sigue necesariamente algo distinto de lo ya establecido por el simple hecho de darse esas cosas” (*Anal. Pr. I.1, 24b 18-20*). Esta breve definición introduce dos elementos importantes: el concepto de “seguirse necesariamente”, y ese “por el simple hecho de darse esas cosas”.

Esto último significa que ese “algo distinto” se debe seguir mecánicamente de esas “ciertas cosas”; no es necesario precisar de ningún término ajeno para que se dé necesariamente la conclusión (Candel-Sanmartín, 2008). Las premisas son fuertemente pertinentes: cada una es necesaria y conjuntamente son suficientes para establecer la conclusión (Vega, 2012). La conclusión está, en cierto modo, contenida en las premisas. Así, este elemento remite a las propiedades analíticas y explicativas de la deducción. Por otra parte, ese “seguirse necesariamente” hace referencia al criterio de validez. La *deducción lógicamente válida* conlleva una *relación de consecuencia lógica*. El que la conclusión se siga necesariamente significa que la conclusión no puede ser falsa si sus premisas son verdaderas. Dicho de otro modo, el consecuente no puede ser falso si el antecedente es verdadero. De serlo, se incurriría en una contradicción. En este sentido, uno de los métodos más empleados para demostrar la validez de un argumento es el conocido como la reducción al absurdo. Brevemente, este método consiste en demostrar que una conclusión dada es verdadera, probando mediante una cadena de inferencias lógicas que, si no lo fuera, conduciría a una contradicción. Así, en un argumento válido, la verdad de las premisas garantiza la verdad de la conclusión.

La teoría de la deducción silogística busca dar con las leyes que garanticen el paso de la verdad de unas premisas *dadas* a un *nuevo* enunciado que se establece en la conclusión. Por tanto, una característica general de los argumentos deductivos lógicamente válidos es la de ser *preservadores de la verdad*.

Por otra parte, lo opuesto no es cierto. La dirección de la relación de consecuencia va del antecedente al consecuente, pero no a la inversa. El hecho de que, si las premisas son verdaderas, entonces la conclusión es necesariamente verdadera no implica el hecho de que, si la conclusión es verdadera, entonces unas premisas dadas en el argumento también tengan que serlo. Es perfectamente posible que se den argumentos válidos con premisas falsas y conclusión verdadera. La conclusión puede ser verdadera por razones distintas a las expuestas en el argumento. La validez lógica no preserva cualquier valor de verdad, sino únicamente el valor de verdadero.

Reglas complementarias

Finalmente, enumeraré una serie de reglas generales a las que se atienen los silogismos en aras a su validez. En primer lugar, el silogismo no puede tener menos de tres términos, pues conecta dos de ellos en virtud de un tercero, ni más de tres, pues se volvería ambiguo dando lugar a falacias. Segundo, los términos de la conclusión no pueden tener mayor extensión que los términos de las premisas. De lo contrario el argumento se vuelve ampliativo y, por tanto, inductivo. Y los argumentos inductivos no garantizan la verdad de la conclusión pese a que las premisas sean verdaderas (i.e. no tienen validez deductiva). Tercero, el término medio tiene la función de concatenar los términos extremos. Es la herramienta mediante la que opera el razonamiento para enlazar términos, por tanto, debe quedar excluido de la conclusión. En la conclusión se presentan los términos ya relacionados, es el producto final de tales operaciones. En cuarto lugar, un argumento que se compone de dos premisas particulares no puede dar lugar a conclusión alguna. Por ejemplo, supongamos una proposición como “algunos españoles son altos” y otra como “algunos escritores son españoles”. Es claro que no se puede deducir de ahí que “algunos escritores son altos”. Tampoco lo logramos si invertimos la segunda premisa por “algunos españoles son escritores”. Para que pueda darse una conclusión es necesario que el término medio sea tomado como universal (esto es, es necesario que esté *distribuido*) al menos una vez en las premisas. Pero eso no basta. Si las dos premisas son negativas, entonces tampoco puede deducirse ninguna conclusión. Por lo menos una debe ser afirmativa. Quinto, si las dos premisas son afirmativas, entonces la conclusión también tendrá que serlo. *La conclusión siempre sigue la parte más débil*. Se entiende por débil la negativa frente a la afirmativa, y la particular frente a la universal. Así pues, siempre que haya una premisa negativa y/o particular, la conclusión deberá ser negativa y/o particular. Esta última regla tiene ciertos paralelismos con las propiedades de la credibilidad.

Concepto general de entimema: ¿qué es?

Con los silogismos, Aristóteles intenta crear un *mecanismo de convicción* que haga patente la concatenación entre los extremos por su misma estructura, por su figura (Candel-Sanmartín, 2008). Los silogismos aristotélicos buscan establecer la relación entre dos términos: un sujeto y un predicado, que se unen o se separan en los juicios expresados en los enunciados de la conclusión. Esta función de mecanismo de convicción modificador de la credibilidad alcanza su máximo nivel de especialización en los entimemas o silogismos retóricos.

Los entimemas son estructuras argumentativas silogísticas que un hablante puede emplear en un contexto comunicativo para influir sobre la credibilidad que los oyentes adscriben a una idea o contenido dados. Los entimemas tienen la función de guiar el razonamiento de la audiencia por una serie de pasos para que establezca la relación de unión o separación de un sujeto y un predicado. De este modo, son los oyentes los que establecen que, efectivamente, P es predicable (o no) de S. Es la audiencia la que deduce esa conclusión mediante un razonamiento aparentemente lógico. Según Racionero (1990), la reducción del argumento a su forma lógica produce una sensación de

seguridad y confianza en la conclusión, en tanto que esta se presenta como una especie de demostración sustentada en un razonamiento, y “nos persuadimos cuando creemos que algo ha quedado demostrado” (*Rhet.* 1355a6).

Pero, para poder encauzar el razonamiento de la audiencia, se deben tener en cuenta ciertas características de esa audiencia. Aquí, interesa saber cuáles son sus creencias y conocimientos previos. Qué ideas o contenidos tiene ya aceptados como verdades. De este modo, podrán emplearse esos contenidos como premisas en los entimemas. Esta orientación a la audiencia refleja uno de los significados de *enthymêma* como compuesto de *en* + *thymôi* que significa *en la mente*, y designa *conclusiones basadas en consideraciones subjetivas* (Candel-Sanmartín, 2008). Por tanto, la marca de los entimemas consiste en emplear las creencias ya presentes en la mente del auditorio al que se dirige como premisas de cara a la demostración de la conclusión crítica. A decir de Perelman y Olbrecht-Tyteca (1958), el acuerdo sobre las premisas constituye el punto de partida de la argumentación. En este sentido, un entimema concreto se crea para que una audiencia concreta acepte una idea concreta en un momento concreto. El entimema se desarrolla en un medio social particular, para una audiencia que cambia y evoluciona. Para que los entimemas alcancen su objetivo deben diseñarse *ad-hoc*, pudiendo perder su efectividad de un día para otro, o de una audiencia a otra.

Las premisas de los entimemas no suelen contener verdades universales, necesarias, primeras, axiomáticas, ni tampoco teoremas, aunque pueden. Por lo general, los entimemas no suelen estar constituidos por premisas propiamente demostrativas. Sólo lo hacen cuando tales verdades coinciden con la opinión general. Y es que el objetivo de los entimemas no es tanto lograr mostrar la verdad como lograr convencer al auditorio sobre la posible verdad de algo. La verdad objetiva no garantiza necesariamente la aceptación por parte de la audiencia. Nada impide que las personas tomen por falsas cosas verdaderas y por verdaderas cosas que no lo son. Lo relevante es mostrar cómo aquello que se pretende establecer se deriva de lo que la audiencia *considera* verdadero o tentativamente verdadero. Son, por tanto, premisas dialécticas que parten signos y probabilidades, pero solo en tanto que constituyan *opiniones generalmente admitidas por la audiencia*; es decir, parten de opiniones plausibles (i.e. *éndoxxa*). Es, precisamente, este carácter de plausibilidad doxástica el que dota de generalidad y estabilidad a los juicios individuales de probabilidad, de tal modo que puedan emplearse con fines retóricos. En este sentido, Aristóteles dirá que “si alguien lleva a los demás a admitir su punto de vista partiendo de opiniones que son tan generalmente admitidas como el caso en cuestión requiere, este ha argumentado correctamente” (*Tóp.* VIII 11, 161b35).

Así pues, no hay un razonamiento sobre lo probable al margen de la opinión. Las premisas de los entimemas se toman de lo que, en general, *se sabe* que la mayoría de las veces ocurre o no ocurre de determinada manera, o que es o no es. Esto, a su vez, se complementa tomando los hechos ya ocurridos y conocidos por la audiencia como signos que señalan hacia la posible ocurrencia de aquello que se pretende establecer. En efecto, “si al existir algo, existe una cosa, o al

producirse algo, antes o después se ha producido otra cosa, aquello es signo de que existe dicha cosa” (*Anal. Pr.* II 27, 70 a 6-10), valga esto todas o la mayoría de las veces conocidas por la audiencia. De este modo, el signo puede funcionar como el antecedente o el consecuente de un condicional al que puede adscribirse cierto grado de probabilidad. Una probabilidad subjetiva condicionada. El núcleo del hecho que conforma el signo funcionará como término medio M del entimema en cualquiera de las tres figuras, pues se pretende demostrar algo a partir de él. Ilustraré este punto tomando, directamente, los ejemplos de Aristóteles:

Demostrar que una mujer ha dado a luz porque tiene leche se hace a partir de la primera figura: pues el medio es *tener leche* [...] En cambio, demostrar que los sabios son honrados porque Pítaco era honrado se hace a través de la última figura (i.e. tercera), pues el medio es *Pítaco* (ser Pítaco es el núcleo) [...] En cuanto a probar que una mujer ha dado a luz porque está pálida quiere tener lugar a través de la figura intermedia (i.e. segunda: pues el medio es *estar pálida*).

Aristóteles, *Anal. Pr.* II 27 (énfasis y paréntesis añadidos)

El sujeto S y el predicado P de la conclusión quedan conectados porque ambos se relacionan creíblemente con el término medio. Dice el autor que, cuando los entimemas concluyen una afirmación indefinida o universal, sólo la primera de las figuras puede establecer el hecho de forma válida y con certeza. Las otras dos, por su propia figura, siempre derivarán en proposiciones con mero valor probable pese a considerar las premisas totalmente creíbles. Esto es, la capacidad persuasiva del entimema no solo depende de la credibilidad de las premisas, sino también de la forma del argumento (esto se detallará más abajo). De hecho, en estos casos, guardarían la estructura de falacias formales que derivan en conclusiones ampliativas, asemejándolas más a nuestro concepto actual de inducción que al de deducción. Por otro lado, el hecho de que, pese a permitir la inclusión de cuantificadores, los entimemas suelen expresarse de modo indefinido contribuye a la vaguedad de su interpretación, dejando que el oyente haga su propia estimación. El citado fragmento también indica que la figura a emplear en la demostración retórica está en función del grado de generalidad del hecho empleado como término medio en relación con los extremos S y P.

Los condicionales formados por M y P pueden funcionar como reglas generales para la construcción de argumentos en la forma de silogismos hipotéticos⁷². A diferencia de los valores de verdad, la probabilidad atribuible a las ideas no es dicotómica, sino que se distribuye en un espacio graduado. Dicho espacio tiene como extremos la *certeza* sobre la imposibilidad y sobre la necesidad de la ocurrencia de un evento. Entre esos dos extremos quedan comprendidos los distintos grados de *incertidumbre*. Así, el evento puede ser considerado desde nada probable hasta totalmente probable. La probabilidad es una magnitud que se da en mayor o menor grado. Esto es de la mayor

⁷² Un silogismo hipotético es aquél que tiene en su premisa mayor un condicional, y en su premisa menor una aserción categórica, o bien, ambas premisas son condicionales.

relevancia para esta tesis, pues asumimos que los entimemas computan credibilidades que se pueden medir como probabilidades subjetivas.

Validez deductiva de los entimemas de la primera figura

En esta tesis se ha trabajado exclusivamente con entimemas de conclusión afirmativa y proposiciones indefinidas contruidos mediante la primera figura. Estos entimemas parecen tener un estatus de validez distinto al de las otras dos figuras cuando las conclusiones tienen estas características. Por tanto, todo lo que se diga a continuación está referido especialmente a este tipo de entimemas de la primera figura. En todo caso, es posible que también pudiera ser aplicable a los formados mediante las otras figuras cuando no concluyen en afirmativo o en universal.

Los entimemas son silogismos con premisas tomadas de las opiniones y creencias de la audiencia con mero valor probable. Al ser silogismos, producen un razonamiento deductivo. En tanto que el valor de sus premisas no es dicotómico sino probable, el razonamiento consiste en una deducción que conlleva una relación de consecuencia *laxa*, no estrictamente concluyente (Burnyeat, 1996; Dyck, 2002; Vega, 2004). Es, pues, un razonamiento deductivo probabilístico. En tanto que deductivo, lo que aparece en la conclusión debe estar, de algún modo, contenido en las premisas y no ir más allá de éstas (Vega, 2004). Del mismo modo, si las premisas son tenidas por verdaderas, necesariamente la conclusión también será considerada verdadera en la misma medida. Y en la medida en que una o ambas se consideren probables, así también se considerará la conclusión. De lo contrario, se incurriría en una contradicción. Los entimemas son *preservadores de la credibilidad* de las premisas a la conclusión. El cómo se produce esa preservación es una de las cuestiones se han investigado en esta tesis.

Ante esta última afirmación, uno podría argüir que, igual que era posible encontrar conclusiones verdaderas precedidas de premisas falsas en argumentos lógicamente válidos, aquí también sería posible encontrar conclusiones creíbles o probables precedidas de premisas consideradas poco probables. En efecto, igual que antes, la dirección de la relación de consecuencia (*laxa*) va de las premisas (i.e. antecedente) a la conclusión (i.e. consecuente), y no al revés. Y ésta sólo nos dice que, si las premisas son creídas (totalmente o hasta cierto punto), entonces la conclusión también deberá ser creída (totalmente o hasta cierto punto), y no a la inversa. Podría decirse que los entimemas preservan necesariamente la credibilidad de las premisas a la conclusión, pero no preservan *necesariamente* la “incredibilidad”. Uno podría creer *previamente* en el contenido de la conclusión por otras razones que no fueran las que se exponen en las premisas. En este caso, del mismo modo en que la falsedad de las premisas no afectaba a la verdad de la conclusión en el silogismo tradicional, la incredibilidad de las premisas podría no afectar a la credibilidad de la conclusión en los entimemas.

Ahora bien, pese a que este caso sea posible, no será lo que generalmente ocurra en contextos argumentativos. En tales contextos, se pretende convencer a la audiencia acerca de la corrección o

verdad de algo. Se pretende que la audiencia acabe creyendo algo; a saber, el contenido de lo expresado en la conclusión. Esto implica que, en un primer momento, ese contenido *no* es creído por la audiencia. No tendría mucho sentido tratar de convencer al oyente de lo que ya esté convencido.

Ahora, para simplificar, supongamos que nos hallamos en unas condiciones argumentativas *ideales* en las que la audiencia no cree en absoluto en el contenido en cuestión. La probabilidad que le asigna a tal contenido en un primer momento es nula, inexistente. Supongamos también que un hablante pretende que aumente la credibilidad que los oyentes adscriben a ese contenido, y que, para hacerlo, construye un argumento. Las premisas de ese argumento sólo podrán ser consideradas igual o más probables que el contenido en cuestión, que ahora se presentará como su conclusión. Si el hablante ha sido hábil, habrá seleccionado las premisas a partir de las creencias generalmente compartidas por el auditorio. Esto es, habrá construido un entimema. De ser así, las premisas sólo podrán ser consideradas como más probables que la conclusión que se pretende establecer. Incluso siendo consideradas como poco probables, lo serán más que la conclusión. Esa probabilidad atribuida a las premisas, aunque fuera poca, se transferirá a la conclusión. Así pues, en un contexto argumentativo donde se pretende convencer de algo que no se cree al principio, la acción de los entimemas consistirá siempre en la transmisión de la credibilidad que hay en el antecedente al consecuente. Una transmisión laxa, no dicotómica, una transmisión de cierta magnitud de credibilidad. Si la credibilidad puede medirse como probabilidades subjetivas, entonces es una magnitud que discurre entre unos límites: 0 y 1, totalmente y nada probable, imposible y necesario, etc. Debe situarse en algún punto de este continuo. Por otro lado, la parte no creída (i.e. la incredibilidad) es el resto complementario a la credibilidad. En estas condiciones, al preservarse la credibilidad, automáticamente se preserva también la incredibilidad. Por lo que es perfectamente adecuado decir que, en condiciones argumentativas, la incredibilidad de las premisas afecta a la conclusión en la misma medida que la credibilidad.

Así pues, en lo relativo a la validez de los entimemas, no se debe perder de vista su función de producir la atribución de credibilidad a una idea que no la tenía. No se debe perder de vista, por tanto, el contexto argumentativo al que pertenecen de forma natural. Esto es lo que hace que no se comporten como los silogismos lógicos tradicionales en los aspectos comentados. De este modo, otra marca fundamental propia de los entimemas es la de que *las premisas deben ser más plausibles que la conclusión* que se pretende defender. De otro modo, no es solo que el argumento pierda todo su sentido, sino que, directamente, no es un entimema.

En conclusión, las premisas se toman de las opiniones y creencias que ya posee la audiencia para conectarlas en un silogismo de tal modo que la tesis a defender quede como resultado y conclusión de éstas; en este sentido, es evidente que, de buena entrada, las premisas serán consideradas probables y plausibles, mientras que la tesis deberá ser inicialmente controvertida.

A raíz de todo esto pueden surgir varias preguntas. Por ejemplo, uno puede considerar que esas condiciones argumentativas ideales son “vacas esféricas” que nunca se dan en la realidad.

Siempre se le puede atribuir un mínimo de credibilidad a todo. Se podría preguntar entonces de qué modo afecta la in/credibilidad previa de la conclusión a su dinámica en función de la transmisión de credibilidad que recibe desde las premisas. Una primera respuesta es que el estándar según el cual las premisas deben ser más plausibles que la conclusión sigue valiendo y es de necesario cumplimiento para que el entimema funcione. Pero ahora vendría la siguiente pregunta: ¿cuánto más plausibles deben ser las premisas respecto de la conclusión? Tal vez exista un umbral de distancia de plausibilidad antes del cual no se producen cambios en la credibilidad de la conclusión. Además, ¿cómo se puede cuantificar la plausibilidad de las premisas, en plural, en su conjunto? Estas y otras cuestiones han sido tratadas con detalle en esta tesis.

Otra forma de concebir los entimemas

Tal y como ponen de manifiesto Ferrater Mora (1994/2005) y Luís Vega (2012), en la mayoría de los diccionarios y libros de texto de lógica, se considera que el entimema es un silogismo incompleto, truncado, abreviado. Se considera, en este sentido, que su característica esencial consiste en que una de sus premisas no está expresada por considerarse tan conocida que el propio oyente la restituye en el proceso de comprensión. Normalmente, la premisa implícita es la mayor. Candel Sanmartín (1988/2008, n. 455; p. 294) dice del entimema que podría tratarse como “un silogismo al que le falta una premisa, cuya conclusión, por tanto, es más fruto de la intuición del que lo expone o lo escucha que de la necesidad objetiva de la propia deducción”. Asimismo, otros también añaden la posibilidad de que el elemento no expresado sea la conclusión. En este caso, tal omisión no se debería al hecho de que la conclusión fuese muy conocida, pues el argumento sería entonces irrelevante. La omisión busca, más bien, hacer partícipe al oyente en la derivación de conclusiones. Finalmente, como señala Vega (2004, 2012), también se encuentra relativamente extendida la concepción del entimema como un tipo de argumento que, tomado literalmente, resulta inválido, pero que deviene válido al añadirse como premisa implícita una proposición demasiado obvia o conocida para ser declarada expresamente. Según Vega, esta última versión resulta trivial e irrelevante a nivel argumentativo por tratar de fijar la definición en función de su supuesto encaje con una noción formal de validez lógica, quedando completamente despojada de su orientación retórica. Además, resulta ser claramente inaplicable para los entimemas de la segunda y tercera figuras cuando resultan en una afirmación indefinida o universal.

Como se puede observar, poco tienen que ver estas definiciones con el objetivo retórico de convencer o de hacer variar las credibilidades asignadas a un contenido. No hay ninguna razón para suponer que el mero hecho de omitir algún elemento del silogismo vaya a convertirlo en más convincente o persuasivo. Es evidente que, si los entimemas constituyen “el cuerpo de la persuasión”, no será por esta característica aparentemente trivial, sino por todo lo comentado antes. Cabe preguntarse entonces de dónde ha salido esta idea tan extendida y por qué ha sido la predominante hasta hace muy poco.

La cuestión es que, tal y como apunta Candel Sanmartín (1988/2008), la palabra entimema tenía ya desde su origen cierta ambigüedad en su significado: no sólo se refiere a tomar las opiniones del oyente sobre la probabilidad de los eventos, sino también a que éste puede suplir posibles elementos omitidos recurriendo a los conocimientos que alberga en su mente. Así, la mayoría de estas definiciones conciben al entimema como un silogismo de carácter comunicativo, basado en la interacción participativa entre un hablante que lo emite, un oyente que lo acepta e infiere la información relevante y un conocimiento compartido entre ambos que permite al oyente realizar la interpretación pretendida por el hablante.

En la *Retórica* también se encuentra esta referencia a la omisión de alguna premisa cuando dice que los entimemas deberán expresarse “a partir de pocas premisas, incluso menos de las que consta el silogismo de la primera figura. Porque si alguna de estas premisas es bien conocida, no hace falta enunciarla: el propio oyente la suple” (*Ret.* I, 1357a 17-20). Se considera, en todo caso, una característica accidental del entimema y no sustancial. El hecho es que, en el habla ordinario, los razonamientos suelen expresarse de este modo; y en particular, en el habla retórico, puede ser conveniente hacerlo por una serie de razones orientadas a la adaptación del discurso a las capacidades e intereses de la audiencia. Para maximizar la posibilidad de que las claves persuasivas del mensaje no pierdan su efecto es importante lograr mantener la atención de la audiencia. En este sentido, será adecuado presentar los razonamientos de forma breve, simple y precisa. Estos razonamientos deben condensar lo esencial del mensaje en unos pocos pasos lógicos, dejando implícito lo que los oyentes ya saben a fin de no cansarles con una verborrea redundante. El discurso debe ser ágil. Además, es recomendable que la argumentación no arranque desde muy lejos ni recorra todos los pasos pues la extensión oscurece el mensaje dificultando su adecuado seguimiento. Así pues, la incompletitud es, básicamente, una cuestión del estilo expresivo que deben adoptar los razonamientos retóricos en el discurso a fin de optimizar su procesamiento. Pero, tal y como indica Racionero (1990), todos los entimemas pueden ser desplegados en la forma de un silogismo, puesto que contienen implícitamente sus premisas o su conclusión. El silogismo es, digamos, la estructura profunda que está en la base de todo entimema (pero, para una interpretación alternativa, véase Bermejo-Luque, 2009). Según Racionero, no debería hablarse, pues, de silogismos incompletos, sino de silogismos disfrazados o encubiertos.

Así pues, en tanto que cuestión de estilo sin poderes suasorios sobre la credibilidad, la incompletitud no es necesaria para que un silogismo sea considerado un entimema. De hecho, Aristóteles distingue entre “verdaderos entimemas” para sus versiones completas, y “partes del entimema” cuando contenga una premisa implícita. A diferencia de estos últimos, los primeros “aclaran la causa de lo dicho” (*Ret.* II 21, 94b20). Si bien dicha causa suele encontrarse en la premisa menor, su aclaración se encuentra en la mayor. Consiste en mostrar en qué sentido lo dicho en la menor es *pertinente* para establecer la conclusión. El elemento omitido suele ser la premisa mayor, pero pueden omitirse también los otros. La escolástica medieval propuso una clasificación de los entimemas incompletos que sigue estando actualmente aceptada (e.g. Bermejo-Luque, 2006; Copi, Cohen y Rodych, 2018; Parry y Hacker, 1991):

- a) Entimemas de primer orden: cuando se omite la premisa mayor.
- b) Entimemas de segundo orden: cuando se omite la premisa menor.
- c) Entimemas de tercer orden: cuando se omite la conclusión.

Cuando se trata del procesamiento de argumentos inmersos en el contexto de un discurso más amplio, posiblemente esta omisión tenga sentido. Pero nada parece indicar que optimice los efectos de los entimemas cuando estos se emiten de forma aislada. Aun así, algunos autores apuntan la posibilidad de que omitiendo premisas poco creíbles se obtienen mejores efectos que con sus respectivas versiones explícitas. La razón de ello sería que, al dar por supuesto su contenido, los oyentes no lo cuestionan. Pero esto me parece discutible. Si el contenido implícito se infiere, entonces se procesa y se valida igual que si fuera uno explícito. Tal vez, se podría argüir que, si es poco creíble, entonces no se infiere pues no forma parte de los conocimientos previos del oyente. Pero esto, otra vez, me parece discutible en varios sentidos. Por un lado, es una cuestión empírica y plausible que se puedan inferir contenidos mecánicamente a partir de una estructura argumental dada, independientemente de los conocimientos; estos sólo intervendrían después en un proceso de validación de la información inferida. Y, por otro lado, si no se infiere la premisa implícita poco creída, entonces la conclusión no surge de ningún razonamiento, y no podría heredar ningún tipo de credibilidad de las premisas. Nada de esto se ha investigado empíricamente. De modo similar, Vega (2004) considera que dejar la conclusión implícita podría tener efectos persuasivos positivos por la complicitad que se establece con el oyente, induciéndole a que sea él mismo quien saque las conclusiones pertinentes. Igual que antes, no veo la necesidad de que eso sea así; en todo caso, este último aspecto no recibirá ningún tratamiento en la presente tesis.

Partes del entimema

Ha quedado claro ya que el entimema es un silogismo y que, como tal, se compone de tres proposiciones y tres términos. Dos de esas proposiciones son las premisas y la restante es la conclusión. La premisa menor contiene el sujeto S y la mayor el predicado P de la conclusión. De este modo, ambas están conectadas con la conclusión. Asimismo, ambas premisas quedan conectadas entre sí al contener un término repetido. Dicho término recibe el nombre de término medio M. En las premisas, S y P se comparan con M.

En tanto que portadora del predicado, la premisa mayor es más general que la menor. De hecho, la premisa mayor de los entimemas incorpora o expresa una regla general de relación plausible acerca del funcionamiento regular de las cosas. Estas relaciones están regidas por algún tipo de criterio que permite vincular dos elementos. Normalmente, vincula el aspecto nuclear de un hecho con un supuesto futuro, pasado o presente. Ese aspecto del hecho es M, y el supuesto es el P de la conclusión. Por otro lado, en la premisa menor se encuentra S de la conclusión, que suele tener un alcance general menor a P. En este sentido, lo expresado en esta premisa suele ser más concreto y particular. Por otra parte, es crucial que S tenga un referente identificable y conocido para la audiencia. Al presentarse S relacionado con M, la premisa menor establece un hecho

concreto reconocible por los oyentes. La remisión al hecho asumido como cierto por parte de la audiencia tiene la función de servir de fuente original de la credibilidad. Por otro lado, la regla general de la premisa mayor tiene la función de servir de llave de paso que permita verter esa credibilidad sobre la conclusión en la que se presentan relacionados el sujeto con el predicado. A continuación, voy a presentar las características principales de la conclusión, la premisa menor y la premisa mayor.

La tesis a establecer: futura conclusión del entimema

Los entimemas pueden emplearse tanto para establecer como para refutar una idea. La presente investigación se limitará al ámbito del establecimiento positivo de contenidos; por tanto, ni me ocuparé de la refutación ni de los vínculos que hay entre el establecimiento de ideas y la refutación de sus opuestas.

Sabemos ya que la credibilidad sobre algo es una magnitud que se da en mayor o menor grado oscilando entre los extremos de certeza en la imposibilidad o en la necesidad de ese algo. La certeza es un estado subjetivo de creencia que se da en la forma: *es cierto que* algo es o no es. Consiste en el estado de máxima magnitud de credibilidad acerca de que S es P y mínima acerca de que S no es P o viceversa, para lo cual es necesario que S tenga algún tipo de referente identificable y conocido para la audiencia. Así pues, la certeza positiva y negativa demarca los límites de los estados subjetivos de la convicción.

Para que un entimema pueda lograr el aumento en la credibilidad de una tesis es necesario que ésta no sea totalmente creída de entrada. De hecho, cuanto menos creída sea, más probable es que el entimema pueda mostrar sus efectos. En este sentido, lo deseable es que se considere que P es poco o nada predicable de S. Predicable en un sentido pragmático o epistémico, no sintáctico o semántico. Traducido a términos probabilísticos, es deseable que la probabilidad de P dado S (i.e. $P(P | S)$) sea baja o nula. En tanto que el aumento de credibilidad se produce sobre ese mismo contenido SP, al presentarse como conclusión del argumento se espera que P pase a ser probable dado S.

A continuación, se muestran tres ejemplos de posibles conclusiones, de los cuales, sólo uno podría cumplir con los criterios necesarios:

1. María está en el sofá.
2. Los medios de transporte contribuirán a la extinción de los osos polares.
3. Las hojas de los árboles caducifolios caen al suelo en otoño.

A la hora de plantearnos la credibilidad de las situaciones presentadas en estos enunciados, vemos que en los dos últimos casos nos es más sencillo que en el primero. Y es que en el primer caso desconocemos totalmente los referentes de los nombres mencionados. No sabemos de qué María ni de qué sofá nos están hablando. El principal problema es que no disponemos de un conocimiento

compartido con el hablante acerca del referente del sujeto gramatical. Podemos *suponer* que se trata de un ser humano, femenino y singular, pero nada más. La mención de un nombre propio cuyo referente singular nos es desconocido bloquea la posibilidad de adscribirle credibilidad a lo que se diga de él. Como consecuencia, tampoco podremos inferir el referente de *ese* sofá. No se nos está hablando de los sofás en general, sino de uno en particular que se predica de María. Por tanto, es imposible saber si “estar en el sofá” es un predicado más o menos probable de “María”. El enunciado describe una situación irrelevante a efectos de la credibilidad. Gran parte de los argumentos que se emplean en psicología del razonamiento y del discurso son de este tipo, lo que les hace ser inadecuados para el estudio de la credibilidad.

Uno podría decir que el “estar en el sofá” es predicable de algo como “María” si asumimos que se trata de una persona cualquiera. Es predicable a nivel gramatical y semántico. Pero la credibilidad no versa sobre el lenguaje, sino sobre la realidad o la representación mental de la realidad. Sólo si suponemos que “María” pertenece al conjunto “humano” y a su subconjunto “femenino” podríamos inferir que lo que no se predica del conjunto, tampoco de sus miembros, y que lo que se predica de todos ellos, también de María. Así, podríamos otorgar credibilidad baja o alta a determinados predicados en función de nuestros conocimientos sobre esos conjuntos de elementos. Pero para ello ya habría que hacer esas suposiciones, que siempre pueden ser erróneas. Por otra parte, todo aquello que se predique de María comprendido en el rango de la posibilidad contingente resultará totalmente incierto.

Por otro lado, el tercer ejemplo es, tal vez, demasiado creíble. Por lo menos, para los individuos escolarizados de las regiones templadas de las sociedades europeas. Por tanto, tampoco sería un contenido candidato a conformar la conclusión de un entimema que pretende establecer su credibilidad.

Finalmente, otorgarle algún grado de credibilidad al segundo ejemplo será relativamente sencillo, puesto que podemos realizar una identificación suficientemente aproximada de los referentes. Éstos pueden variar en su grado de singularidad, desde individuos únicos pero conocidos a conjuntos más extensos de entidades. Habría mucho que decir acerca de esto, pero sobrepasa los límites del presente trabajo. Pues bien, en la medida en que uno *no* crea que la extinción de los osos polares sea (pragmática o epistémicamente) predicable de los medios de transporte, le dará una credibilidad baja a la posibilidad de ocurrencia de la situación representada por ese enunciado.

La premisa menor o hecho establecido

La premisa menor del entimema contiene la relación entre el sujeto S o término de menor generalidad de la conclusión y el término medio M. Más arriba ya se ha dicho, por un lado, que S debe tener un referente reconocible para los oyentes y, por el otro, que M consiste en el aspecto nuclear de un hecho. Ese aspecto nuclear es impersonal y descontextualizado. Valgan como ejemplos tener leche, pisar el freno, estar en el salón, la emisión de gases invernadero, etc. Así, S y

M entran en relación en la proposición conformada por la premisa menor, ya sea actuando como sujeto o como predicado de ésta. En tanto que S tiene un referente conocido por el oyente y M es la partícula nuclear de un hecho, la relación de S con M contextualiza ese aspecto del hecho en una situación reconocible. En virtud del referente de S, el contenido de la premisa menor queda enlazado a un espacio y/o un tiempo determinado para el oyente. De este modo, la premisa menor en su conjunto expresa un *hecho propiamente dicho*, presente, pasado o posible, susceptible de ser evaluado como verdadero o falso por el oyente. Esto es, una situación susceptible de ser evaluada como más o menos creíble o probable.

Así pues, en la práctica retórica, la premisa menor *remite a los hechos* (ver sec. 3.2., cap. 2, pt. 1). En los entimemas, los hechos remitidos deben ser conocidos, aceptados o, de entrada, *creíbles* para los destinatarios del argumento. La remisión a unos hechos plausibles tiene la función de constituir la *fuentes primitiva de la credibilidad* del argumento. Según van Eemeren et al. (2009), la premisa menor alberga la “*assumptio*” de Cicerón, entendida como un punto de partida aceptado por los oyentes. En argumentación jurídica, la premisa menor expresaría hechos que se consideran probados y que, a su vez, sirven de base para el establecimiento de la prueba (Atienza, 2011). En la práctica, cuanto más concretos, perceptibles, ostensibles o mostrables sean esos hechos, mejor. Por ejemplo, todos estaríamos más o menos de acuerdo en que:

- En nuestro país, los grandes supermercados han disminuido su emisión de bolsas de plástico.

Esto podría servir de base a partir de la cual argumentar otra cosa que se derive de la ocurrencia de ese hecho ampliamente aceptado como cierto. Pero el recurso a contrafactuales o a hechos meramente posibles e imaginables también puede funcionar bien siempre y cuando resulten prácticamente innegables para una audiencia dada. Por ejemplo, la mayoría de los miembros competentes de nuestra cultura estaría de acuerdo con el siguiente enunciado, pese a no haber pensado nunca en ello por lo estrambótico de su contenido (i.e. pese a no residir en la memoria):

- Un elefante cabe en una catedral gótica.

No es, hasta donde yo sé, un hecho real ocurrido al que pueda señalar en la construcción de mi argumento sobre otra cosa. Pero sí que es un hecho que podría darse o haberse dado. Dado mi conocimiento ingenuo sobre sus referentes y sus tamaños típicos, es un hecho posible que resulta creíble. Por tanto, puede funcionar como premisa para algún argumento. En todo caso, para contar como un hecho con un referente real aceptado, debería enunciarse como “tal o cual elefante cabe (o cupo) en tal o cual catedral gótica”. Y ahora, dependiendo de *qué* elefante y de *qué* catedral gótica estemos hablando y de los conocimientos de que disponga la audiencia acerca de tales referentes podría ser más o menos aceptado.

En la práctica argumentativa, la premisa menor conforma un *dato* (Toulmin, 1958) o evidencia en la que basarse para afirmar el contenido de la conclusión (Braet, 1999; Cortés Gabaudán, 1994; Racionero, 1990). Será la *razón* que justifica la conclusión una vez se muestre su pertinencia para

esta. En un esquema causal, expresa la causa de lo que se dice, razón por la cual Hood (1984) llamó a la premisa menor “cláusula porque” (*because clause*).

Observemos ahora los siguientes ejemplos:

1. María está en el salón.
2. Los medios de transporte contribuyen a la emisión de gases invernadero.
3. Las hojas de los árboles caducifolios brotan en otoño.

Pese a que el primero de estos ejemplos describe una situación posible, hasta cierto punto fácil de imaginar, gramatical y semánticamente correcta, no hay forma de plantearse adecuadamente su credibilidad. Los motivos son los mismos que en el primer ejemplo visto más arriba para la conclusión. Es necesario conocer los referentes de los nombres que ahí aparecen. Sólo así podrá plantearse la credibilidad y aceptabilidad de la situación contingente descrita. Asimismo, sólo si resulta aceptable y creíble podrá pasar a formar parte del entimema.

Igual que sucedía con los ejemplos del apartado anterior, atribuirles credibilidad a las otras dos situaciones descritas es más sencillo por la facilidad de determinar los referentes. Sin embargo, el tercer ejemplo tampoco constituiría una buena premisa menor de un entimema en nuestra cultura. La mayoría de los miembros competentes de nuestra comunidad epistémica no creería en la realidad de lo que dice.

Finalmente, el segundo ejemplo sí que podría funcionar adecuadamente como premisa menor de un posible entimema. Estimar su credibilidad no presenta grandes dificultades. Pese a no especificar exactamente qué medios de transporte, rápidamente nos vendrán a la mente aquéllos que se dan con mayor frecuencia y/o saliencia en nuestra experiencia cotidiana. De modo que, fácilmente, podemos identificar sus referentes prototípicos de manera suficiente como para que la frase tenga sentido y podamos estar más o menos de acuerdo. En la medida en que se considere que “la emisión de gases invernadero” es predicable de “los medios de transporte” se le atribuirá alta credibilidad.

Por último, cabe señalar que el hecho enunciado por la premisa menor constituye un caso específico dentro del campo que cubre la regla enunciada en la premisa mayor. Esto se entenderá mejor habiendo visto el siguiente apartado.

La premisa mayor o regla general de relación

La premisa mayor del entimema contiene el predicado P o término de mayor generalidad de la conclusión, y el término medio M. Al estar constituida por el aspecto nuclear de un hecho descontextualizado y un predicado general, en esta premisa no necesariamente se podrán

encontrar referentes concretos identificables⁷³. Esta premisa refiere reglas generales no contextualizadas sobre las cosas que suelen ocurrir la mayoría de las veces. O, por lo menos, sobre lo que los oyentes saben o creen que suele ocurrir. Por ejemplo, que cuando hay nubarrones, entonces puede llover; que si un ser vivo vuela, entonces probablemente tenga alas; que cuando dos personas se llevan bien, entonces suelen ayudarse mutuamente; o que, si una persona es violenta, entonces puede tener problemas; etc. O, con un nivel mayor de abstracción, que, si se produce una causa, entonces normalmente se produce su efecto; que, si ocurre lo que es menos probable, entonces seguramente también lo que es más probable; que, si ha ocurrido lo posterior, entonces también lo anterior; que, si se quería y se podía, entonces se hizo; etc. Es decir, relaciones normales o esperables entre categorías o eventos que invocan *conocimientos genéricos y esquemáticos* de los participantes en el proceso de razonamiento. Crucialmente, los distintos niveles de abstracción aquí presentados no son incompatibles, sino subsumibles, donde el más abstracto subyace al que lo es menos. Desarrollaré esta idea con detalle a lo largo del siguiente gran apartado dedicado a los esquemas tópicos de argumentación.

Más arriba se mencionó que la remisión a un hecho plausible puede funcionar como signo que señala hacia la hipotética existencia de *otra cosa*. La premisa menor, en su conjunto, pretende hacer ostensible que el hecho se ha producido en la realidad o que es incuestionablemente factible o realizable. Pero aquello que permite inferir esa otra cosa no es el hecho *per se*, sino el conocimiento genérico que los individuos *tengan vinculado o puedan vincular* con la partícula nuclear del hecho. Esta partícula es el término medio M, por lo que esa otra cosa será el predicado P.

Ambos elementos se vincularán en la memoria de tal modo que al obtener M podamos viajar hacia P. M se convierte así en la base del desplazamiento mental hacia espacios y tiempos hipotéticos posibles P en la realización de suposiciones. En la premisa mayor, M podrá servir como antecedente o consecuente de un condicional al cual se le pueda adscribir cierta probabilidad. Dado que esta tesis se circunscribe a los entimemas de la primera figura, M funcionará como antecedente y, por lo tanto, P será el consecuente: si M, entonces P.

En todo caso, esto no es más que una primera aproximación. Antes de entrar en más detalles, se expondrán algunos ejemplos que podrían o no funcionar como premisa mayor de un posible entimema de la primera figura:

1. Si una persona está en el salón, entonces estará en el sofá.
2. Si la emisión de gases invernadero continúa, entonces se extinguirán los osos polares.

⁷³ Salvo en los entimemas contruidos mediante la tercera figura, donde el término medio es el más particular y ejerce de sujeto en las dos premisas: "Pítaco es honrado; Pítaco es sabio: los sabios son honrados". A mi juicio, lo que aquí encontramos es más un caso de ejemplificación inductiva que de entimema deductivo. Las premisas consisten en dos hechos singulares, sin diferencias formales ni funcionales entre ellas. De tales hechos se deriva una conclusión ampliativa que tiene aspecto de regla hipotética. De este modo, parece estar incurriendo en un proceso inductivo de confirmación de hipótesis por casos singulares o ejemplos. Por deducción, incluso en un entimema, la conclusión no debería ir más allá del particular "algún sabio es honrado". En cualquier caso, esta discusión es tangencial a los objetivos de esta tesis, además de exceder holgadamente sus límites.

3. Si algo brota en otoño, entonces cae al suelo en otoño.

En este caso, en nuestra cultura, los dos primeros ejemplos podrían resultar creíbles y, por tanto, podrían servir como premisa mayor de un posible entimema. Seguramente haya bastante acuerdo en que, cuando una persona está en el salón, la mayoría de las veces está en el sofá. Aunque no sea así necesariamente. De este modo, si yo supiera quién es esa tal María, y me quisieran convencer de que está en el sofá dado que está en el salón, seguramente lo creería. Uno hace este tipo de inferencias constantemente en la vida cotidiana. Por otro lado, también en nuestra cultura y en nuestro tiempo, la mayoría de las personas estarían de acuerdo con el segundo enunciado. Tal vez no por experiencia directa, pero sí vicaria, por el discurso y la comunicación recibida. Así, cuando me digan que los medios de transporte van a contribuir a la extinción de los osos polares dado que emiten gases invernadero, ahora me lo creeré; por lo menos, hasta cierto punto.

En todo caso, quisiera señalar una diferencia aparentemente relevante respecto del condicional del primer enunciado. En este segundo caso, parecería poco adecuado decir que el condicional esté expresando algo que pasa la mayoría de las veces. Sería raro decir que la mayoría de las veces que se emiten gases invernadero de forma continuada, se suelen extinguir los osos polares. Entonces cabe preguntarse en qué sentido lo considero probable. Podría resultar una premisa plausible por frecuencia de exposición a ese mensaje por los medios de comunicación, o porque sea la opinión de la mayoría o de los expertos. Y puedo considerar que, la mayoría de las veces, la opinión de la mayoría o de los expertos es la correcta. Por otro lado, podría considerarlo probable por la activación de un dominio de conocimiento mayor según el cual, la mayoría de las veces que se destruye el hábitat de una especie, la especie se extingue⁷⁴. Aunque no necesariamente. Tal y como Hastings (1962) y Hahn, Bluhm y Zenker (2017) ponen de manifiesto, en los argumentos cotidianos pocas veces se hacen explícitas absolutamente todas las premisas en las que se basan. Más bien, la aceptabilidad de cada aserción depende de muchos elementos con probabilidades variables. En un sentido similar, Vega (2004) señala que todo argumento constituye la parte expresa o manifiesta de un iceberg discursivo. La mayor parte del conocimiento requerido para su interpretación se da de forma implícita (Hahn et al., 2017; Van Dijk, 2012)⁷⁵.

Finalmente, pocos miembros competentes de nuestro alrededor estarán de acuerdo con el contenido del tercer condicional. Pese a la posibilidad de que esta regla se cumpla en algún lugar, en ciertas especies vegetales, o para determinadas comunidades epistémicas especializadas, no es algo que forme parte del conocimiento común de las personas que nos rodea. La gente podría guardar fuertes reservas acerca de la verdad de este enunciado. Por tanto, no sería un buen candidato para ser la premisa mayor de un posible entimema para la mayoría de las personas de nuestro entorno. Llegados a este punto, observemos el siguiente argumento con premisa implícita: “dado que lo que brota en otoño cae al suelo en otoño, las hojas de los árboles caducifolios caen al

⁷⁴ Ver Khemlani, Lotstein, y Johnson-Laird, (2015) para una solución similar a las estimaciones de probabilidad de eventos únicos.

⁷⁵ En la fig. 8 de la sec. 1.3.5.3., cap. 2, pt. II, se desarrolló un ejemplo del entramado de un dominio causal y del acceso a sus refutadores.

suelo en otoño”. Todo él es raro. La cuestión es que no solo podemos estar en un estado de completa incertidumbre o de relativo desacuerdo con la premisa mayor, sino que también estamos en desacuerdo con la menor (que se halla implícita). Sin embargo, seguramente estaremos de acuerdo con la conclusión. Esto es, podemos estar de acuerdo con la conclusión a pesar de estar en desacuerdo con el argumento. Uno puede creer en la conclusión por otras razones distintas a las presentadas en el argumento. Son casos en los que no es necesario que nadie trate de convencernos de lo que ya estamos convencidos (ver, más arriba, acerca de la validez de los entimemas).

Más abajo, se presenta la Tabla 16, con los tres argumentos de la primera figura empleados en estos ejemplos en su versión completa. De estos, sólo el segundo podría conformar un buen candidato de entimema para convencer de la conclusión a una comunidad epistémica particular.

Como se puede observar a en dichos ejemplos, P podrá ser considerado como un consecuente de M en función de que los conocimientos genéricos del individuo permitan establecer esa vinculación. Ahora bien, hay muchísimas formas en las que M y P se pueden vincular. Es posible que, en ocasiones, M y P estén previamente vinculados en la memoria semántica, como, por ejemplo, “nubarrones” y “lluvia” en “si hay nubarrones, entonces lloverá”, o con “camarero” y “restaurante” en “si es camarero, entonces trabaja en un restaurante”. Esto podría darse en casos de conocimientos muy cristalizados por la alta frecuencia de coocurrencia de los dos eventos en la experiencia, o por la relevancia que tenga el hecho de disponer de estos conocimientos para la preservación del bienestar del individuo.

Tabla 16

	Ejemplo 1	Ejemplo 2	Ejemplo 3
Premisa Mayor	Si una persona está en el salón, entonces estará en el sofá.	Si la emisión de gases invernadero continúa, entonces se extinguirán los osos polares.	Si algo brota en otoño, entonces cae al suelo en otoño.
Premisa Menor	María está en el salón.	Los medios de transporte contribuyen a la emisión de gases invernadero.	Las hojas de los árboles caducifolios brotan en otoño.
Conclusión	María está en el sofá.	Los medios de transporte contribuirán a la extinción de los osos polares.	Las hojas de los árboles caducifolios caen al suelo en otoño.

Pero en muchos otros casos esta vinculación previa no existirá. M y P pueden hacerse *vinculables* pese a que no hayan coocurrido nunca ni en la experiencia directa, ni en la experiencia vicaria que se da a partir del testimonio y la comunicación, ni en la conciencia del individuo. Este podría ser el caso del condicional relativo al papel de la emisión de gases invernadero sobre la extinción de los osos polares. Es posible que “gases invernadero” y “osos polares” no hayan estado previamente vinculados en la mente del individuo. Sin embargo, en el condicional se hacen vinculables en virtud de cierto tipo de relación. En este caso, en virtud de una relación causal. De este modo, ese condicional quedaría subsumido por una regla general de relación abstracta, del tipo: si una causa está presente, entonces también lo estarán sus efectos. Así, según van Eemeren et al. (1996), la premisa mayor expresa unos principios de justificación denominados “propositio” por Cicerón, o “garante” por Toulmin (1958).

Poniendo un caso más complejo, imaginemos que alguien quiere convencer a su interlocutor de que destruir el medioambiente (ozono, flora, fauna, etc.) no tiene consecuencias negativas para los humanos porque antes habremos logrado producir vida en Marte, disponiendo así de un “planeta B”. Su interlocutor podría contestarle diciendo que “si hemos destruido la vida donde ya la había, no la produciremos donde no la hay”. Sería muy arriesgado suponer que los participantes del intercambio comunicativo dispusieran de esta creencia condicional como tal ya formada en la memoria. Sin embargo, resulta convincente por el tipo de relación que se establece entre el antecedente y el consecuente. En este caso se da una relación compleja que entremezcla relaciones de contrariedad con relaciones de mayor o menor probabilidad. Si eliminamos la de contrariedad, quedaría como “si donde había vida no hemos sabido preservarla, donde no la hay no sabremos crearla”. La *regla general abstracta* que subyace a esta *relación* condicional es la de que, si lo más probable no ha pasado, entonces lo menos probable tampoco ocurrirá. A estos tipos de reglas se les llama *esquemas de argumentación*, y están regidos por unas *relaciones tópicas*.

Relaciones tópicas y esquemas de argumentación

Los conceptos de garante, esquema de argumentación, tópico y lugar común son altamente complejos, y han dado lugar a una extensa producción bibliográfica en filosofía lógica y epistemológica, en estudios del discurso, de argumentación y de comunicación, así como en inteligencia artificial. Una de las claves teóricas que propongo en esta tesis es que las relaciones tópicas constituyen la vía principal por la cual discurre la credibilidad en la red de creencias. Esta idea ha sido ya desarrollada en el modelo teórico propuesto en esta tesis.

Muchos autores consideran que los conceptos previamente mencionados son una misma cosa, y que tal cosa se encuentra implícita en la premisa mayor, la cual, a su vez, también puede estar implícita en la expresión del entimema. Como se mencionó más arriba, debido a su función en el argumento, la premisa mayor de un entimema integra a la vez varios niveles de abstracción. Empezaré hablando de su *función justificativa* para luego poder abordar sus niveles de abstracción, el concepto de esquema tópico de argumentación, su estructura y las relaciones del contenido y la

forma con la credibilidad. Su función adicional en la *construcción* de argumentos (i.e. en la *inventio*) no será abordada en esta tesis.

Función justificativa de regla de paso o puente de transición

La mera presentación de un hecho, por innegable que sea, no basta para conformar una buena razón que justifique la tesis o hipótesis bajo cuestión, por lo que tampoco bastará para generar convicción. Ese hecho deberá conformar una evidencia *adecuada* para una tesis particular.

Aducir que en el cielo hay nubarrones para defender la predicción de que va a llover puede ser adecuado si el oyente sabe que, de hecho, los hay y si conoce algún tipo de *relación* mediante la que conectar los nubarrones con la lluvia. En este ejemplo, la mayoría de la gente sabe que hay una relación de causalidad necesaria, aunque no suficiente o, como mínimo, de coocurrencia, donde las nubes siempre preceden a la lluvia. Si alguien no lo supiera, esa aducción no le serviría de razón para creer en la predicción; sería irrelevante. Del mismo modo, aducir que los nubarrones están encima de mi casa no parece ser una buena razón para establecer que va a llover. Pese a que el hecho aducido fuera real y lo sepamos, no vemos una conexión necesaria ni suficiente entre la situación de los nubarrones respecto de mi casa y la lluvia. Pero si alguien creyese que la hay, creer que las nubes están encima de mi casa le podría servir para creer que va a llover. Así pues, no basta con que el individuo acepte totalmente el hecho aducido, sino que debe ver que el hecho es pertinente y/o relevante para la conclusión. Para ello, deberá atender al contenido del mensaje y contrastarlo con algún dominio de conocimiento del mundo que albergue en su memoria semántica. De los distintos niveles de abstracción de los que hablaré, este sería el primero, el menos abstracto de los abstractos.

La pertinencia del hecho para la conclusión hipotética requiere del establecimiento de una regla o principio que permita realizar la inferencia, mostrando la legitimidad del paso de los datos aducidos a la afirmación de la conclusión, actuando como un puente entre ambos elementos (Toulmin, 1958). Esa regla de relación licencia y valida el *paso de la credibilidad* de la premisa menor a la conclusión. Tales reglas quedan instanciadas por un condicional hipotético de carácter general expresado en la premisa mayor, que Toulmin (1958 / 2005; p. 134) formula como “si <datos> D, entonces <conclusión> C”, o como “datos tales como D permiten extraer conclusiones o realizar afirmaciones tales como C”, o alternativamente: “dados los datos D, puede asegurarse que C”. Toulmin llamó *garantes* a tal tipo de reglas, y en su modelo (ver sec. 4.2., cap. 1, pt. 1) queda simbolizado con una flecha de D a C, que denota su función relacional. En el ejemplo anterior, ese condicional quedaría formulado con un contenido como “si hay nubarrones, entonces llueve”. Ese condicional general estaría ejemplificando un *esquema de argumentación* situado en un segundo nivel de abstracción. Un esquema cuasi-formal, regido por un criterio de relación de carácter semántico, ontológico o epistémico (Macagno y Walton, 2014, 2015). Tal criterio es el tipo de *relación tópica*, que en el caso del ejemplo consistiría en una relación de causalidad. Así pues, su

función principal es la de servir de conducto o puente de conexión entre el hecho aducido en la premisa menor y la conclusión controvertida, garantizando la transmisión de la credibilidad.

Niveles de abstracción: situación de los esquemas de argumentación en el razonamiento discursivo

Actualmente, se considera que los esquemas de argumentación son la evolución moderna de los antiguos tópicos que surgen con las obras de *Tópicos* y *Retórica* de Aristóteles, y constituye un tema de creciente interés en la teoría de la argumentación y en la rama argumentativa de la inteligencia artificial (e.g. Atkinson y Bench-Capon, 2021; Braet, 2005; Green, 2016; Macagno y Walton, 2015; Macagno, Walton y Reed, 2017; Rubinelli, 2009; Vega, 2004, 2012). Sin embargo, es un campo de estudio relativamente problemático por carecer de una definición clara, lo que ha dado lugar a una gran cantidad de tipologías y clasificaciones de estos tópicos o esquemas de argumentación en función criterios muy variados. A grandes rasgos se considera que son patrones generales y estereotipados de razonamiento cotidiano que las personas emplean espontáneamente al conectar ideas para resolver problemas o construir una prueba. Esas conexiones pueden estar regidas por distintos tipos de relación, como relaciones causales, definicionales, temporales, etc. En un discurso, un hablante se puede servir de esos patrones de relación comúnmente empleados y aceptados para mostrar la pertinencia de unos datos como razón en apoyo de una afirmación ante un oyente. Antes de entrar en un intento de definición en la que haya relativo acuerdo entre los distintos autores, es conveniente organizar cierto mapa conceptual en el que situar esta noción.

Los esquemas de argumentación se realizan por medio de la premisa mayor de los entimemas. La premisa mayor del entimema garantiza la inferencia mediante un contenido específico que combina términos de la premisa menor y de la conclusión. Esta premisa mayor consiste en una regla general condicional que puede ser analizada en diferentes niveles de abstracción. Pasemos a verlos.

1. *Esquemas de conocimientos definidos y tópicos específicos*

En un primer nivel, la premisa mayor consiste en una regla genérica formada a partir de la abstracción de patrones de coocurrencias que se dan en los eventos cotidianos de la *experiencia*. Esto es, si en la mayoría de los eventos singulares que componen una clase de experiencias resulta que A y B coocurren, entonces se abstrae una regla que los vincule. Es una regla genérica ligada al contenido de A y de B en el contexto de una clase, género o dominio de conocimiento definido. Representan aquello que ocurre la mayoría de las veces en determinadas disciplinas o en determinados entornos; por ejemplo, que cuando hay nubarrones, llueve. En tanto que productos inductivos ligados al entorno y circunscritos a un dominio de conocimiento, estas reglas genéricas pueden cambiar a través del espacio, del tiempo, de la cultura y de las comunidades epistémicas específicas. En cualquier caso, si son productos de la inducción y ésta depende de la frecuencia o repetición, también gozarán de cierta solidez y estabilidad en la mente, de tal modo que puedan

contar efectivamente como *conocimientos esquemáticos compartidos* por una comunidad dada en un momento dado.

Estos conocimientos pueden ser de orden muy cotidiano, como, por ejemplo, el hecho de saber que, la mayoría de las veces, si alguien está en el salón, entonces está en el sofá. Asimismo, pueden ser de orden mucho más específico, como cuando se refieren a lo que pasa la mayoría de las veces en dominios definidos como la ética, la jurisprudencia, la política, la economía, la medicina, etc. En estos casos tan especializados, tales esquemas pueden adoptar la forma de argumentos prefabricados y empaquetados para apoyar puntos de vista prototípicos en la argumentación. Contienen principios de inferencia específicos para el campo en cuestión. En este papel de casilleros profesionales, estos esquemas de géneros o dominios reciben el nombre de *tópicos específicos* (Macagno y Walton, 2014; Rigotti, 2006). El problema es que los dominios definidos posibles son muchos, con muchas ramificaciones, derivando en un número de tópicos específicos aparentemente ilimitado, sin casi criterios de admisión o exclusión. A veces se olvida su carácter mínimo de estructura estereotipada en un razonamiento argumentativo, llegando a haber tantas propuestas como autores (Zagar, 2010).

2. Principios tópicos generales

Estos esquemas cotidianos y tópicos específicos que acaban de ser comentados pueden estar ejemplificando, a su vez, un siguiente nivel de abstracción. Este nivel de abstracción consiste en reglas generales de conexión de tipo semántico y ontológico (Macagno y Walton, 2014; Rigotti, 2006, 2009; Rigotti y Greco-Morasso, 2006), desligadas de dominios específicos de conocimientos. Representan lo que sucede siempre o la mayoría de las veces en las relaciones que se producen entre categorías funcionales abstractas; por ejemplo, que, cuando se da una causa, normalmente se produce un efecto. En este nivel se encontrarían los principios *tópicos generales* que rigen los esquemas de argumentación. Estos principios tópicos generales pueden ser de naturaleza cualitativamente distinta (p. ej. causalidad, semejanza, contrariedad, etc.). Como se verá más abajo, cada tipo de relación tiene su propia estructura. Esta estructura determina el tipo de operación inferencial que debe llevarse a cabo para conectar coherentemente los elementos de la relación. Este sería el máximo nivel de abstracción relativo al *contenido* del argumento. Es, de hecho, un nivel cuasi formal.

Los esquemas de argumentación no suelen hacerse explícitos en los argumentos, sino que subyacen al contenido temático de la premisa mayor (que, por otro lado, también se puede dejar implícita). En este sentido, se consideran como reglas de relación metalingüísticas (Prakken, 2010) o herramientas mentales muy accesibles que reflejan nuestro modo de organizar la realidad (Macagno y Walton, 2014). El número de estas herramientas cognitivas o principios tópicos generales es limitado. Pero este número es aún muy variable dependiendo del criterio de clasificación que se emplee (para una revisión de clasificaciones de los esquemas tópicos generales, ver Macagno, Walton y Reed, 2017).

Su reducción a un número relativamente bajo y estable es deseable para que puedan ser considerados como la pieza central de la productividad de las creencias. En efecto, una de las virtudes que se atribuye a los esquemas tópicos generales de argumentación es la de lograr obtener, con pocos recursos, abundancia de resultados en la composición de argumentos (Candel-Sanmartín, 2008) para la justificación de creencias. En psicología existen algunos trabajos que proponen maximizar esta reducción traduciendo los tópicos a la fórmula del teorema de Bayes (e.g. Hahn y Hornikx, 2016). En esencia, proponen que todos los tópicos pueden quedar subsumidos a las probabilidades condicionadas. A mi juicio, es posible que las credibilidades se computen y combinen entre sí de una manera aproximada al cálculo probabilístico bayesiano. Sin embargo, el coste de esta reducción cuantitativa es la pérdida de cualquier distinción cualitativa. Y, en la comprensión, las ideas se conectan unas con otras en función de criterios cualitativos y semánticos (e.g. Asher y Lascarides, 2003; Hobbs y Agar, 1985; Kehler, 2002; Khalid, Alikhani, Fellner, McMahan, y Stone, 2020). Es decir, que los componentes de una representación sólo pueden conectarse legítimamente cuando haya alguno de dichos criterios. Una de las características cruciales de esas conexiones es que varían en fuerza. Es en este punto donde las probabilidades pueden entrar a operar como un factor de la credibilidad: esa fuerza se puede expresar como una probabilidad condicionada (e.g. Khalid et al., 2020; Pagin, 2014). En este sentido, el establecimiento de los criterios cualitativos que gobiernan las relaciones tópicas de los esquemas de argumentación son una precondition necesaria para la adscripción de probabilidades por encima de 0.5 y para su dinámica. Desde esta perspectiva, el problema sobre el catálogo de las relaciones tópicas generales parece estar en un nivel de discusión distinto al de las operaciones probabilísticas sobre las magnitudes de credibilidad. Aunque todo lo que haya detrás de la dinámica de credibilidades sean probabilidades condicionadas, esto no resuelve el problema del número y la clasificación de las relaciones tópicas generales. Por lo demás, salvo que indique lo contrario, cuando hable de esquemas tópicos de argumentación en lo que sigue me estaré refiriendo a este nivel general de abstracción.

3. *Diagramas de argumentos*

Estos esquemas de argumentación en los que reside el principio de relación tópico constituyen una especificación del *garante* (Rigotti y Morasso, 2010; Vega, 2012), conectando la premisa menor con la conclusión. El garante, a su vez, consiste en una descripción funcional de la premisa mayor como elemento de un esquema del argumento en su conjunto. En efecto, el modelo de Toulmin (sec. 4.2., cap. 1, pt. I) y otras formas de diagramación de argumentos son abstracciones esquemáticas de las funciones que llevan a cabo los distintos componentes de todo argumento en su objetivo de justificar aseveraciones, opiniones o creencias. Así, en el modelo de Toulmin, vimos que la premisa menor alberga los *datos* (o hechos) sobre los que se basa el argumento; que la conclusión es la *tesis* controvertida o hipótesis, también llamada *aseveración* (Bermejo-Luque, 2012), cuya aceptación o rechazo se pretende lograr. Y, la premisa mayor es el garante del paso inferencial de los datos a la conclusión.

4. *Esquemas inferenciales lógico-formales*

Finalmente, estos esquemas funcionales de los argumentos se integran sobre estructuras formales lógicas globales que configuran las reglas de inferencia, ya sean deductivamente válidas, como el MP o el MT o ciertas formas silogísticas, o inválidas. Estas reglas son limitadas, contables y no son modificables. Son las reglas lógicas que rigen la disposición formal de los términos o proposiciones del argumento. No están destinadas a incrementar la aceptabilidad de la conclusión partiendo de la aceptabilidad del contenido de las premisas. De hecho, están absolutamente libres de cualquier contenido o sistema externo a su propia forma (p. ej. lingüístico u ontológico). Y su forma hace referencia a la posición de los términos, a los conectores y a los cuantificadores. Representan el mayor nivel de abstracción, reuniendo bajo principios más generales estructuras argumentales diferentes y similares, necesarias o derrotables (Macagno y Walton, 2015). En este sentido, un mismo tipo de relación tópica que rijan un esquema de argumentación puede ser colocado sobre cualquier esquema de inferencia lógica puramente formal. Como he mencionado más arriba, el que se empleará en los experimentos de esta investigación es el de la primera figura silogística en modo indefinido.

5. *Conclusión*

Un primer nivel de abstracción se daría en la regla genérica condicional expresada en la premisa mayor relativa a un dominio de conocimiento definido. En un segundo nivel de abstracción se encontraría el esquema tópico de argumentación, normalmente implícito, que contiene el tipo de relación abstracta que permite conectar los hechos con la aseveración controvertida. Dicho esquema de argumentación es una especificación del garante, el cual está situado en un tercer nivel de abstracción. El garante funciona por medio de la premisa mayor de una estructura o esquema inferencial formal, situado en un siguiente nivel de abstracción (Walton y Macagno, 2015), propio de un razonamiento de tipo deductivo en un entimema. En este sentido, los esquemas de argumentación entimemáticos son el resultado de la combinación del nivel de abstracción ejemplificado por el contenido del condicional general y de la forma o figura del esquema inferencial formal de tipo deductivo. Las relaciones semánticas o tópicos constituyen el criterio que permite vincular legítimamente los elementos de la representación. Por otro lado, el hecho de que cada una de esas relaciones tenga una naturaleza distinta permite identificar y clasificar los distintos tipos de argumentos. Todos estos niveles de abstracción no se solapan, sino que se combinan de formas extremadamente complejas (Macagno y Walton, 2015).

Concepto de esquema tópico de argumentación

Existe cierto acuerdo en que los esquemas de argumentación consisten en patrones abstractos, generales y estereotipados de las inferencias más usuales y comúnmente empleadas en los razonamientos que se expresan en los discursos cotidianos (e.g. Racionero, 1990; Vega, 2012; Walton y Macagno, 2015). Son prototipos argumentativos. Representan la estructura genérica cuasi

formal de los tipos de argumentos más naturales. Lo que les hace ser *cuasi* formales es que albergan, esencialmente, *tipos de relaciones* de carácter cualitativamente distinto. Estas son las relaciones tópicas. Tales relaciones constituyen los principios rectores que permiten la inferencia, conectando legítimamente los elementos de las premisas con los de la conclusión. De este modo, permiten comprender situaciones (Dwyer, Hogan, y Stewart, 2013), resolver problemas (Miller, 2000) y construir pruebas (Warnick, 2000). Funcionan como reglas generales de relaciones ontológicas, semánticas o epistémicas entre objetos, proposiciones o conceptos, ideas y creencias (e.g. Braet, 2005; Macagno et al., 2017; Walton y Macagno, 2015; Warnick, 2000).

Los principios que rigen estos enlaces pueden ser de naturaleza muy variada. Así, hay inferencias basadas en el tópico de lo posible, del tiempo, de la cantidad, de la definición, del género y la especie, de la propiedad, de la contrariedad, del signo, de la causalidad, la reciprocidad, la semejanza, la consecuencia y un largo etcétera, incluyendo, según algunos, tópicos basados en la autoridad del experto, o en la opinión de la mayoría. Son principios inferenciales basados en las apelaciones últimas del mensaje. En ese sentido decía más arriba que representan el nivel más abstracto relativo al contenido del argumento. La estructura de estos principios abstractos se expresa en la forma de *máximas inferenciales* (ver siguiente apartado). Por tanto, la naturaleza de estos principios determina la estructura de estas máximas. De este modo, cada esquema particular queda definido por el tipo de relación que conlleva, que determina su estructura propia expresada en la máxima. Esto permite identificarlos para poder estudiarlos. Es en estas estructuras donde, ulteriormente, reside la fuerza o el poder de convicción del argumento (Candel-Sanmartín, 2008; Racionero, 1990; Vega, 2004). Así, los esquemas de argumentación quedan constituidos como reglas de cobertura legal que garantizan que el criterio tópico pase de las premisas a la conclusión (Macagno et al., 2017; Vega, 2004), permitiendo así la propagación del flujo de credibilidad a través de las redes de creencias. Son las llaves de paso del sistema.

Los esquemas tópicos de argumentación expresan relaciones de transmisión de propiedades entre conceptos e ideas (Katzav y Reed, 2004). Tales propiedades son, principalmente, la valencia emocional o actitudinal y, sobre todo, la credibilidad. Son, por tanto, generalizaciones abstractas pero accesibles empleadas en el razonamiento cotidiano que describen el procedimiento de derivación mediante el cual se transfiere la justificación de unas ideas a otras. Son reglas o pautas de justificación inferencial (Vega, 2004). Según el parecer de varios autores (e.g. Grimaldi, 1972; Macagno y Walton, 2015; Prakken, 2010; Vancil, 1979), los esquemas tópicos de argumentación describen los procesos cognitivos y epistemológicos que realizan los individuos en su razonamiento cotidiano. Describen, por tanto, los procesos psicológicos de la justificación en la dinámica de creencias.

Forma y contenido en los esquemas tópicos de argumentación

Puede decirse que los esquemas de argumentación codifican el aspecto morfosintáctico de las creencias. Son matrices estructurales que conectan funcionalmente las variables relevantes del principio rector tópico del razonamiento en el que se basa el argumento. Por ejemplo, conectan causas con efectos, la ocurrencia con la no ocurrencia de eventos más probables con menos probables, el género con su especie, el todo con sus partes, lo semejante con lo semejante, lo contrario con su contrario, etc. Estas variables pueden entenderse como categorías de celdillas vacías que pueden ser rellenadas por infinidad de contenidos específicos a modo de términos constantes. Cuando esto sucede, tales contenidos quedan presentados como expresión particular de una inferencia de carácter general que garantiza la aceptabilidad de la conclusión. La estructura de esta inferencia es la de una relación binaria que se expresa en enunciados condicionales (Dyck, 2002; Rigotti y Greco-Morasso, 2010; Vega, 2012). De este modo, según Dyck (2002), la relación tópica (T) reemplaza la implicación ($\rightarrow$) en la premisa mayor para derivar en un entimema con la forma de un “Modus Ponens Entimemático”, también llamado a veces “Modus Ponens Derrotable” (e.g. Walton, 2002; Walton, 2006; Walton y Reed, 2002). Este tipo de modus ponens se puede expresar mediante las siguientes máximas inferenciales (donde “T (,)” sustituye a “ $\rightarrow$ ”):

- a) Si P y T (P, Q), entonces Q.
- b) Si P (x), y T (x, y), entonces P (y).
- c) Si P (x), y T (P, Q), entonces Q (x)

Estas expresiones se pueden traducir de manera general como “si se da una premisa menor (P), y hay una relación tópica (T) entre la premisa menor y la conclusión (Q), entonces puede aceptarse la conclusión”. Más tarde, estos predicados pueden ir combinándose unos con otros y con más tipos de relaciones tópicas, tejiendo una red de relaciones de justificación integradas en un discurso más amplio (Macagno et al., 2017) o en un sistema de creencias. Siguiendo a Dyck (2002), pasemos a ver la estructura de los esquemas de argumentación correspondientes a algunos tópicos representativos (para una muestra exhaustiva, ver el artículo original del autor):

Esquema de argumentación “de la causa al efecto”:

- Forma de la máxima inferencial: Si $P \wedge Cs (P, Q) \rightarrow Q$

Esto se lee como “si se da un evento, y ese evento es causa (Cs) de otra cosa, entonces también se dará la otra cosa como efecto suyo”. Estas variables podrían quedar rellenadas mediante los siguientes términos constantes:

- a) Si hay nubarrones, y los nubarrones causan la lluvia, entonces se producirá la lluvia.
- b) Si no tenemos aliados, y no tener aliados causa perder guerras, entonces perderemos la guerra.

Tomando una de las variaciones anteriores, también se puede encontrar como:

- Si $P (x) \wedge Cs (P, Q) \rightarrow Q (x)$

Esto puede leerse como “si un evento o propiedad P se predica de un objeto x , y P es causa de otro evento o propiedad Q , entonces Q también se predicará de ese objeto x ”. Por ejemplo:

- a) Si los medios de transporte emiten gases invernadero, y la emisión de gases invernadero causa la extinción de los osos polares, entonces los medios de transporte causarán la extinción de los osos polares.
- b) Si los avances en medicina provocan un aumento del número de personas mayores, y el aumento del número de personas mayores causa, a su vez, un aumento del número de personas con alguna enfermedad, entonces los avances en medicina provocarán un aumento en el número de personas con alguna enfermedad.

Se observa, pues, que el esquema de causalidad está en un nivel de abstracción mayor, pudiendo así ser ejemplificado por un sinnúmero de contenidos particulares. Pasemos a ver, brevemente, dos esquemas de argumentación más.

Esquema de argumentación “del más y el menos”, también llamado “de la mayor y menor probabilidad”, o “a fortiori”:

- Forma de la máxima inferencial: Si $P(x) \wedge Pr(P(x) < P(y)) \rightarrow P(y)$

Esto se lee como “si ha pasado lo que es menos probable (Pr), entonces también pasará lo que es más probable.” Como ejemplo, podríamos poner: Si el Nàstic ha ganado al Real Madrid, y que el Nàstic gane al Real Madrid es menos probable que la posibilidad de que el Barça gane al Real Madrid, entonces el Barça ganará al Real Madrid”.

También podríamos encontrar su forma inversa:

- Si $\neg P(x) \wedge Pr(P(x) > P(y)) \rightarrow \neg P(y)$

Que se lee como “si no ha ocurrido lo que es más probable, entonces tampoco ocurrirá lo que es menos probable”. Por ejemplo: Si ni los epidemiólogos saben cómo evolucionará la pandemia, y la probabilidad de que lo sepan los epidemiólogos es mayor que la de que lo sepan los de tu grupo de WhatsApp, entonces los de tu grupo de WhatsApp no saben cómo evolucionará la pandemia. Ya se puede ir observando lo artificial e incómodo de presentar explícitamente todas las premisas en relaciones tópicas más complejas que las de la causalidad. Veamos un último ejemplo.

Esquema de argumentación “de los contrarios”:

- Forma de la máxima inferencial: Si $P(x) \wedge Ct((x, y) \wedge (P, Q)) \rightarrow Q(y)$

Se puede leer como: “si el atributo P se predica del objeto x , y x es el contrario (Ct) del objeto y , y a su vez, P es el contrario del atributo Q , entonces Q se predicará de y ”. En este caso, puede verse en seguida cómo la explicitación competa de la premisa mayor en un ejemplo concreto convertiría al argumento en un mensaje altamente farragoso y difícil de seguir. Por ejemplo: Si el autocontrol es beneficioso, y el autocontrol es lo contrario de la impulsividad, y beneficioso es un atributo contrario a perjudicial, entonces la impulsividad debe ser perjudicial.

Pese a la aparente complejidad de algunas relaciones tópicas (por lo menos, en su notación), son relaciones que se captan al vuelo y sin esfuerzo aparente en el curso de la comprensión cuando se presentan en su forma abreviada. En todo caso, que la complejidad de esta notación esté reflejando complejidad algorítmica y que se traduzca, a su vez, en procesos psicológicos más complejos es una cuestión empírica que aún no se ha abordado.

Seguidamente, presento algunos de los ejemplos anteriores en un formato abreviado:

- a) Los medios de transporte emiten gases invernadero. Por tanto, causarán la extinción de los osos polares.
- b) El Nàstic ha ganado al Real Madrid. Por tanto, el Barça también lo hará.
- c) Ni los epidemiólogos saben cómo evolucionará la pandemia. Por tanto, aún menos los de tu grupo de WhatsApp.
- d) El autocontrol es beneficioso. Por tanto, la impulsividad debe ser perjudicial.

A continuación, presento brevemente un concepto psicológico hermanado con el concepto de esquema típico de argumentación que se está tratando aquí.

El Concepto de Molécula Implicacional en Psicología

En las teorías sobre la organización esquemática del conocimiento desarrolladas en psicología social desde los años sesenta se habla de un concepto extremadamente similar (si bien, menos completo) al de esquemas de argumentación. Hablan de la existencia de unos conjuntos de proposiciones psicológicamente relacionadas, que están unidas por implicación psicológica, denominados *moléculas implicacionales* (e.g. Abelson y Reich, 1969; Bear y Hodun, 1975; Kruglanski, 1989; Wyer, 2004, 2007; Wyer y Carlston, 1994). Estas moléculas funcionan como esquemas (Schank y Abelson, 1977) que reflejan las generalizaciones acerca de los eventos que ocurren frecuentemente en el mundo real (Wyer y Albarracín, 2005). Por ejemplo, siguiendo a estos últimos autores, la convicción general de que fumar causa cáncer de pulmón puede expresarse en la siguiente forma molecular *causal*:

- P fuma; P tiene o tendrá cáncer de pulmón.

La creencia generalizada de que una persona obtiene lo que se merece revestiría la siguiente forma molecular del *merecimiento justo*:

- P hace algo malo (bueno); a P le sucederán cosas malas (buenas).

La creencia generalizada de que las personas que comparten gustos se gustan mutuamente quedaría expresada por la forma molecular de *semejanza-atracción*:

- A P1 le gusta X; a P2 le gusta X; P1 y P2 se gustan mutuamente.

La creencia, en fin, de que todas las propiedades de un grupo se generalizan a sus miembros se expresaría como la forma molecular de *estereotipos*:

- $P \in G$; los miembros de G tienen el atributo X ; P tiene el atributo X .

Estas moléculas implicacionales tienen efectos sobre la interpretación de la realidad social. Son un tipo de estructuras esquemáticas de relaciones binarias regidas por un principio de compleción según el cual, una vez que se rellena una de sus variables, automáticamente se infiere la otra completando así el esquema. Los términos constantes que rellenan estas variables son hechos procedentes de la experiencia. Así, cuando un hecho específico ejemplifica una de las variables de la molécula, la otra se infiere tomándose por verdadera. Este principio se aplica independientemente de la dirección de la relación. Por ejemplo, en la molécula de merecimiento justo, al observar que una persona ha cometido una mala acción podrá inferirse que será castigada; pero si observamos que ha recibido un castigo o desgracia, esa información se empleará para inferir que había hecho algo malo (Lerner y Miller, 1978; Lerner y Simmons, 1966; Walster, 1966; Wyer y Albarracín, 2005). Esto último le distinguiría hasta cierto punto de los esquemas de argumentación, puesto que en muchos de éstos la direccionalidad de la relación entre los elementos importa.

Lugares comunes: en qué sentido son lugares y en qué sentido son comunes

Teniendo en cuenta lo que sabemos ahora sobre las particularidades de la estructura formal de los esquemas tópicos de argumentación, de su contenido material y de las relaciones que establecen entre sí forma y contenido, se puede entender mejor por qué se les ha denominado *lugares comunes* a lo largo de la historia.

Primero, porque cada esquema está compuesto por lugares o habitáculos vacíos (i.e. variables, celdillas) dispuestos en una estructura común para todo argumento de un mismo tipo, siendo múltiplemente rellenable. Sintetizan los patrones de razonamiento comunes a la multiplicidad de argumentos pertenecientes a una misma clase de relación.

Relacionado con esto, otras veces se les ha denominado así por considerar que esas clases de relación representan los lugares más frecuentes y comunes a los que dirigirse para construir argumentos: apelando al principio de causalidad, de contrariedad, de la definición, etc. Después de todo, la detección de estas clases de relación ha procedido del análisis de gran diversidad de textos (e.g. Perleman y Olbretch Tyteca, 1958; Schellens y De Jong, 2004; Walton, Reed y Macagno, 2008; Warnick, 2000).

Y, a su vez, relacionado con los dos puntos anteriores, también se les ha llamado lugares comunes por el hecho de que estas estructuras cuasi formales se pueden aplicar “en común lo mismo a cuestiones de justicia que de física, de política o de otras muchas materias” (*Ret.* 1358a10). Son reglas de justificación inferencial independientes de la materia que se esté tratando, aplicables a cualesquiera de ellas y, por tanto, no le son propias a ninguna en particular. Si bien apoyarse en el teorema de la energía cinética puede servir para mostrar la validez de una afirmación en física, pero no en ética, o apoyarse en el principio de *ignorantia juris non excusat* puede hacerlo en derecho, pero no en química, los lugares comunes no pertenecen a materia particular alguna, sino al modo

en que razona la mente de forma natural. No pertenecen al objeto de aplicación, sino al sujeto que las aplica. En este sentido, los esquemas de argumentación son comunes a todas las formas de razonamiento, encontrándose tanto en razonamientos deductivos necesarios o derrotables, en razonamientos inductivos y en razonamientos abductivos (Macagno, Walton y Reed, 2017).

Finalmente, esto nos lleva a una última manera en la que también son comunes. Según Figa Faura (1993), los esquemas tópicos consisten en conocimientos, presunciones o creencias de carácter estructural y relacional comúnmente aceptadas, ya sea de forma explícita o implícita. En este sentido, varios estudios (e.g. Maio, Hahn, Frost, Kuppens, Rehman, y Kamble, 2014; Mercier, Deguchi, Van der Henst, y Yama, 2016; Warnick; 2000), parecen indicar que este tipo de conocimientos esquemáticos y estructurales no está adscrito a una comunidad epistémica particular, sino que es independiente del tiempo y de la cultura. Tales conocimientos parecen reflejar, más bien, nuestra forma de acceder, entender, relacionar y ordenar la realidad que percibimos (Macagno y Walton, 2014). Pese a que se aplican al mundo y, en cierta medida, dependen del entorno para generarse o activarse, sobre todo hablan de nuestra forma común de funcionar en el razonamiento cotidiano (Grimaldi, 1972; Vancil, 1979). De ahí su alta credibilidad y aceptación implícita general. La estructura del esquema no está en discusión. En una línea similar, Mercier (2011; 2013) y Mercier y Sperber (2011) defienden que las disposiciones cognitivas relativas al razonamiento argumentativo son universales e innatas, recibidas por la herencia genética y activadas por el entorno para la adaptación al medio. Están indisolublemente unidas a la racionalidad humana.

Credibilidad y niveles de abstracción: efectos de los conocimientos, de la naturaleza del tópico, y de la validez de las estructuras formales

Desde el inicio de esta sección sobre esquemas tópicos de argumentación, se ha puesto de manifiesto que su función consiste en permitir la transmisión de la credibilidad de la premisa menor hacia la conclusión mediante una conexión justificada por un principio o regla de relación. Es decir, que se parte de una credibilidad inicial correspondiente al hecho o dato y, mediante la regla de justificación inferencial tópica, se transmite a la aseveración de la conclusión. Por otro lado, antes también se dijo que esa regla expresada mediante la premisa mayor también debía ser creída para que el entimema pudiera tener efectos. Se mencionó que esas reglas binarias hacen referencia a lo que los individuos saben o creen que suele ocurrir por lo general o siempre. La credibilidad que se le atribuya dependerá en gran medida de la magnitud estimada de esa ocurrencia. Ahora bien, el hecho de que en esa premisa mayor se den distintos niveles de abstracción a la vez complica considerablemente el análisis. En este apartado, el más largo del apéndice C, me dispongo a analizar las relaciones que guardan los distintos niveles de abstracción de la premisa mayor del argumento con la credibilidad. En particular, me centraré en el papel que juegan los contenidos temáticos materiales y los esquemas tópicos generales de argumentación para, finalmente, ir algo más allá de la premisa mayor y examinar también el rol de la validez de las estructuras formales del argumento.

Empecemos recordando que la premisa mayor del entimema consiste en una proposición con una estructura condicional, ya sea formulada como tal o subyacente. Ese condicional es la expresión particular de una regla o generalización acerca del contenido temático de algún dominio de conocimiento definido. El contenido temático de esa regla general está ejemplificando un esquema de argumentación regido por alguna relación tópica. Las relaciones tópicas son de naturaleza variada y denotan la apelación justificativa del argumento. Este nivel de abstracción está desligado ya de ningún dominio de conocimiento definido pese a seguir ligado al contenido. Es el nivel en el que se relacionan las categorías funcionales abstractas del argumento que permiten la justificación. Esto es, es el nivel en el que se sitúa el garante del argumento. Por su parte, el argumento en su conjunto también posee una estructura formal abstracta, un esquema formal, que surge de la combinación de la premisa mayor con la menor. Un ejemplo de tales esquemas formales serían las figuras silogísticas (válidas o inválidas). Pues bien, tanto el contenido material que ejemplifica el esquema de argumentación, como la naturaleza del principio rector de la relación tópica de dicho esquema, como la validez del esquema inferencial formal tendrán efectos sobre la credibilidad. Como se mencionó más arriba, según Macagno y Walton (2015), estos niveles de abstracción interactúan entre ellos combinándose de formas muy complejas. A continuación, trataré de exponer la contribución de cada uno de estos niveles de abstracción a la credibilidad, yendo de menor a mayor abstracción.

Relaciones de la credibilidad con el conocimiento sobre el contenido material del esquema tópico

Esquemas temáticos cristalizados como creencias explícitas

El contenido material hace referencia a los términos constantes que rellenan las variables del esquema tópico en la premisa mayor condicional del argumento. Es parte del material sobre el que propiamente se razona. Son los contenidos generales propios del primer nivel de abstracción mencionado más arriba. Se trata de conocimientos obtenidos a partir de procesos inductivos sobre experiencias episódicas particulares de las situaciones cotidianas producidas en el entorno físico, o sobre los mensajes recibidos de nuestro entorno social⁷⁶. Los conocimientos abstraídos del entorno por inducción ya no son particulares, sino generales y descontextualizados. Los referentes de esas abstracciones son *clases de objetos y situaciones*. Este tipo de información se almacena por dominios de conocimiento en la memoria semántica. Puesto que proceden del entorno físico y social, son conocimientos compartidos por la mayoría de los miembros competentes de una sociedad o de una comunidad epistémica determinada. En ese sentido, estos conocimientos genéricos abstractos se corresponden con los esquemas, marcos y guiones abordados por la psicología y la inteligencia artificial (Bartlett, 1932; Minsky, 1974; Rumelhart, 1980; Schank y

⁷⁶ Pese a que destaco el papel de la experiencia directa, este tipo de creencias también puede formarse por otros tipos de frecuencias: por ejemplo, porque, la mayoría de las veces, nos hayan dicho que A es B, o por la ratio en la que he pensado en A y B juntas respecto de separadas. Estas coocurrencias aumentan el grado de asociación entre A y B, así como la familiaridad del par.

Abelson, 1977). Y los esquemas, marcos y guiones son parte de las *creencias explícitas* que se presentaron en el capítulo 3 de la parte I.

Son, pues, estructuras esquemáticas con un referente relativo a un género o dominio definido de conocimiento que se obtienen a partir de muestras limitadas de experiencias concretas con el objeto de la generalización. Por ejemplo, en el caso de la experiencia directa, a partir unas pocas experiencias con las tormentas, me formo la representación general de que “si se produce un rayo, entonces luego viene un trueno”. En estos casos, la experiencia directa no sólo constituye la base para la formación de esa representación, sino también para su justificación, conllevando, por defecto, un alto grado de credibilidad. Por tanto, en este ejemplo, la probabilidad subjetiva para la ocurrencia estimada de “trueno” dado que he percibido un rayo será alta.

Puesto que no conozco la totalidad de los elementos del conjunto (por ejemplo, el universo de todas las instancias de tormenta), las generalizaciones obtenidas como conclusión del proceso inductivo se consideran *ampliativas* (van más allá de lo que pueden garantizar los datos). Como consecuencia, estos esquemas de objetos o situaciones *pueden ser modificados por la acumulación de más experiencia*. Es decir, pese a que los contenidos de este primer nivel de abstracción puedan llegar a ser bastante estables, nada impide que continúen actualizándose con la entrada sucesiva de nueva información. Por ejemplo, si resulta que, a partir de un momento dado, todas o la mayoría de las veces que observo rayos ya no veo que le sigan truenos, entonces se debilitará la conexión entre ‘trueno’ y ‘rayo’, pudiendo llegar a extinguirse. La credibilidad del condicional anterior disminuirá. En tanto que las relaciones entre los componentes de estas representaciones son probabilísticas, subjetivamente consideraremos que la $P(\text{trueno} \mid \text{rayo})$ pasará a ser baja. De este modo, los esquemas temáticos de contenidos materiales pueden ser modificados a partir de los datos del entorno recogidos por el sistema cognitivo, con sus correspondientes efectos sobre la credibilidad.

Además de las variaciones derivadas de las limitaciones de conocimiento sobre los referentes, los esquemas también varían de unos entornos a otros en función del *espacio* y del *tiempo* en el que nos encontremos. Consecuentemente, también de unas sociedades y comunidades epistémicas a otras. Por tanto, el contenido de una misma proposición condicional genérica no será creído por igual por diferentes personas en diferentes sociedades, lugares y tiempos (esta idea se desarrolló con más detalle la sec. 1.3., cap. 1, pt. II).

Por lo dicho, queda claro que los procesos inductivos se aplican a regularidades epistémicas acerca de lo que observamos que pasa la mayoría de las veces para derivar en unas generalizaciones materiales falibles, unas leyes no estrictas que albergan excepciones. Son leyes contingentes, subjetivamente probables y argumentativamente rebatibles. Guardan la forma de “la mayoría de los casos A, son B” o “A es, si nada interfiere, B” o “si A, entonces normalmente B” y similares, pese a que no siempre se añada explícitamente el cualificador modal. Así, la mayoría de las relaciones condicionales sobre eventos cotidianos que se nos puedan ocurrir, como “si algo es un pájaro, entonces vuela”, son generalizaciones falibles de este tipo. Estas generalizaciones sirven en el

argumento como leyes de cobertura para dar cuenta y explicación de lo aseverado en la conclusión. Funcionan como reglas generales hipotéticas conectadas con un dominio de conocimiento del mundo que ejemplifican el esquema tópico de argumentación.

Generalizaciones construidas ad hoc como creencias implícitas

Los contenidos materiales de la premisa mayor que ejemplifican el esquema tópico de argumentación no siempre tendrán que consistir en esquemas cristalizados en la memoria obtenidos inductivamente. Esto es, no siempre tienen que consistir en las llamadas creencias explícitas (ver. Sec. 2, Cap. 3, Pt. I). Esto vale para premisas abstractas y generales como la mencionada de la relación causal entre los gases de efecto invernadero y la desaparición de los osos polares, o la relación *a fortiori* del más y del menos entre el conocimiento de los epidemiólogos y los que no lo son respecto a la pandemia, y muchos otros casos por el estilo en otras relaciones tópicas. No se puede afirmar sin más que esos contenidos materiales se hayan obtenido por inducción derivando en un esquema material cristalizado. Sin embargo, siguen siendo generalizaciones que ejemplifican esquemas tópicos de argumentación, aceptables para una audiencia particular, pero falibles, a las que se le puede asignar grados de probabilidad. Estas creencias se corresponden con las llamadas *creencias implícitas* (*ibid.*). Como Wyer y Albarracín (2005) ponen de manifiesto, “no es necesario que las creencias sean parte del conocimiento y esté almacenado en la memoria como tal. Más bien, las creencias pueden ser vistas simplemente como estimaciones de probabilidad subjetiva que se computan *on-line* en el momento en que devienen necesarias para lograr un objetivo para el cual son relevantes (p. ej. para comunicar información a otros, o para tomar una decisión conductual)” (p. 274). La historia psicológica que haya detrás para dar cuenta de su construcción y efectividad puede ser muy amplia y va más allá de las ambiciones de esta tesis. Sólo apuntaré una serie de posibilidades que no son excluyentes.

Por un lado, una vez que se le presenta el contenido condicional ‘ $A \rightarrow B$ ’ al oyente, éste activa sus conocimientos sobre A y sobre B. En virtud de los conocimientos activados, reconoce la legitimidad del condicional como un caso particular de un esquema tópico general. Esto es, reconoce que la manera en que A y B se relacionan es por causalidad, o frecuencia, o definición, o contrariedad, etc. Así pues, estos conocimientos facilitan el *acceso* a la relación tópica. A la relación tópica le corresponde una estructura inferencial que, al aplicarla, permite generar la *expectativa* de que produciéndose (o no) uno de tales componentes, probablemente se haya producido o se produzca (o no) el otro. De este modo, este condicional *ad hoc* se hace creíble pese a que nunca se hubiera pensado sobre él, ni se haya dado en la experiencia previa. La probabilidad de esta creencia implícita se estima en el acto.

Por otro lado, es posible que la activación de A y B remita, ulteriormente, a un esquema cristalizado como creencia explícita en un dominio de conocimiento más amplio. Mediante esta activación, A y B se vuelven vinculables y el condicional, creíble. Por ejemplo, que la emisión de gases invernadero pueda provocar la extinción de los osos polares podría ser aceptable para una comunidad epistémica que albergue la creencia explícita esquemática de que, la mayoría de las

veces, la destrucción de un hábitat extingue a sus habitantes. Así, la fuerza de este tipo de contenidos temáticos de la premisa mayor descansaría tanto sobre el esquema de argumentación como sobre la creencia explícita esquemática a la que remite. De este modo, los dominios de conocimiento pueden actuar como el marco de justificación para sostener otras ideas.

Relaciones de la credibilidad con la naturaleza del principio rector tópico

Todos los esquemas de argumentación tienen la función de establecer una conexión relevante entre los datos de la premisa menor y la tesis de la conclusión, pero algunos ofrecen conexiones con más garantías que otros. Esto es así debido a la naturaleza de la relación que rige el tópico.

Hay esquemas de argumentación cuyos principios de relación tópica tienen un carácter pragmático y social o testimonial, otros tienen un carácter más epistémico y empírico y, finalmente, otros tienen un carácter intuitivo más bien apriorístico (i.e. independiente de la experiencia). Y, en este mismo orden, ofrecen unas garantías para la credibilidad que van de más débiles (más fáciles de rebatir) a más fuertes (más difíciles de rebatir). En cualquier caso, para mayor facilidad en la explicación, empezaré tratando las relaciones entre la credibilidad y los esquemas de argumentación empíricos, para después pasar a los pragmáticos o testimoniales y, finalmente, abordar los apriorísticos.

Credibilidad de los tópicos empíricos de base epistémica

Los esquemas de argumentación regidos por principios tópicos empíricos son típicamente rebatibles o derrotables. Esto es, contienen la posibilidad de que su aplicación no sea universal, sino que albergue excepciones en tanto que se conectan directamente con los dominios de conocimiento de los contenidos temáticos del nivel de abstracción inmediatamente anterior. De hecho, se podría decir que heredan la magnitud de credibilidad obtenida en el nivel anterior. En este sentido, las garantías que ofrecen para la credibilidad no dependen sólo de ellos. En cualquier caso, estos niveles son interdependientes, puesto que el que ahora nos ocupa es el que permite enlazar los términos de la relación condicional material de un modo funcionalmente relevante.

Los principios que rigen los esquemas empíricos o epistémicos reflejan *tipos de relaciones* que se dan entre los objetos y eventos del mundo (Braet, 2005; Warnick, 2000). En su interacción con el entorno, el ser humano observa que los objetos se relacionan entre sí de diversas maneras. Las experiencias concretas de relaciones en algún sentido similares se agrupan en clases abstractas mediante procesos inductivos produciendo *reglas generales de relación* (ver sec. 1.2.2., cap. 1; y 2.2., cap. 4, pt. II). Así, estos esquemas sintetizan las relaciones observadas en las situaciones ordinarias (Wyer, 2007; Wyer y Albarracín, 2005). Como hemos visto, un primer resultado de los procesos inductivos consiste en la obtención de esquemas temáticos de objetos y situaciones (como ‘restaurante’, o como ‘si una mujer tiene leche, entonces ha dado a luz’) correspondientes con el nivel anterior. Pero el proceso de abstracción no se detiene ahí. En el presente nivel, acaban colapsando en categorías máximamente generales de las que resultan unos esquemas que versan sobre relaciones (p. ej. x causa y ; x se asemeja a y ; x es más r que y ; x es anterior a y ; x e y son

recíprocos, etc.). En este sentido, aquí la abstracción ya no se refiere a objetos y situaciones sino a *relaciones*. Las abstracciones de relaciones pueden, a su vez, estar incluidas en las de objetos y situaciones⁷⁷, pues son más abstractas y generales. Sus características son más funcionales que temáticas. Por ello mismo, no pertenecen ningún género o dominio definido de conocimiento.

Así pues, el esquema de argumentación queda definido por su estructura funcional, donde sus categorías son necesarias para la relación abstraída. Pese a que se adquiere o se activa a partir de la observación y de la experiencia, *es inmodificable*. Las relaciones abstraídas son más fruto de la comprensión que de la mera asociación. No tiene sentido discutir si las causas producen efectos, o si lo más probable ocurre más que lo menos probable, o si las cosas que se parecen comparten características en algún aspecto. Todo ello es necesario por su significado mismo. Ahora bien, que su estructura sea definitiva e inmodificable no significa que sus garantías de conexión sean infalibles. Estas estructuras relacionales se aplican sobre contenidos representacionales acerca de las cosas del mundo. Y en tanto que el mundo es cambiante y los individuos solo poseen un conocimiento limitado de él, la garantía que ofrecen es falible. Al consolidarse a partir de las regularidades observadas, asumen ese carácter regular. Así, en este nivel de abstracción, *lo que se sabe* que ocurre la mayoría de las veces es que cuando se da uno de los dos elementos de la relación binaria, *suele* darse el otro: la mayoría de las veces que se produce un evento considerado como una posible causa de algo, suele darse su efecto; la mayoría de las veces que ocurre lo considerado como menos probable respecto a algo, también ocurre lo que es considerado como más probable a su mismo respecto, sea lo que sea; y así sucesivamente. Por tanto, aunque A sea considerada como causa de B, el esquema tópico no garantiza la ausencia de excepciones en la representación de la relación. La aplicación adecuada del esquema enlaza pertinentemente A con B, pero *no es suficiente*, por sí misma, para la transmisión de una credibilidad total. Eso depende del conocimiento del individuo acerca del *contenido material* del dominio referido por la premisa, pudiendo encontrar o no excepciones en su representación subjetiva de la relación. Así pues, este tipo de esquemas de argumentación sólo *permite* la transición de la credibilidad mediante el enlace pertinente. Pero la

⁷⁷ Hay que ser cuidadosos en este punto: si por representación abstracta de objetos y situaciones nos referimos a marcos o matrices con celdillas o variables vacías del tipo: agentes, fines, acciones, marco espacial y temporal, instrumentos, relaciones causales entre elementos, etc. entonces, nos encontraríamos exactamente en el mismo nivel de abstracción. Esto es, los esquemas de argumentación, los esquemas de Schank y Abelson (1977) e incluso los modelos mentales o de situación (Johnson-Laird, 1983; Van Dijk y Kintsch, 1983), *en tanto que matrices estructurales*, se encontrarían en el mismo nivel de abstracción. Un nivel relativamente alto. Sólo se diferencian cuando se toman en cuenta sus referentes. El referente del esquema de argumentación sigue siendo una *relación absoluta*. El referente del esquema de Schank y Abelson (1977) es una 'cosa' (como 'restaurante') o una 'situación' o 'secuencia de acciones' (como 'compraventa') general. Estas unidades generales, a su vez, están compuestas por elementos. Por tanto, su esquema correspondiente consiste, nuevamente, en la *generalización de las relaciones regulares* entre estos elementos (Wyer y Albarracín, 2005). Estos referentes siguen siendo abstractos, pues son generalizaciones. Pero son menos abstractos que los referentes de los esquemas de argumentación. Finalmente, el referente del modelo de situación consiste en una 'cosa o situación específica y singular' (como '*este* restaurante' o '*esa* visita que hice a tal restaurante'). Esto es, sus referentes *no* son abstractos, pese a que su matriz sí que lo sea. Las cosas concretas del mundo también están compuestas por elementos relacionados entre sí.

magnitud de credibilidad que transcurra por él depende enteramente del conocimiento sobre el contenido material.

Credibilidad de los tópicos pragmáticos de base testimonial

Existen otros esquemas de argumentación actualmente aceptados que también dependen del entorno y de la experiencia, pero en este caso se trata de la experiencia social y vicaria que se produce como resultado de la comunicación y la cultura.

Este tipo de esquemas tienen como principios inferenciales de justificación la relación de fiabilidad que se establece entre una idea o aseveración con la fuente testimonial que la asevera. Son, por ejemplo, los argumentos “ad populum” (i.e. si lo dice la mayoría, entonces es verdad), “ad verecundiam” (i.e. si lo dicen los expertos, entonces es verdad) o “ad hominem” (i.e. si lo dice un particular inexperto, entonces no es verdad). No siempre se los ha considerado propiamente como esquemas de argumentación sino como algo externo a la técnica argumentativa. Por ejemplo, los dos primeros se pueden considerar la base del éndoxa argumentativo, esto es, como precondiciones para que los contenidos probables pasen a ser plausibles, de tal modo que puedan tener un uso dialéctico y retórico (ver sec. 5.2., cap. 1, pt. I). Por otra parte, los otros tipos de esquemas de argumentación (i.e. los de base empírica que hemos visto y los de base apriorística que veremos luego) también podrían ser implícita o explícitamente aceptados por todos, la mayoría o los expertos. Además de estas consideraciones, otras veces han aparecido en catálogos de falacias (e.g. Hamblin, 1970). Y las falacias se consideran, en principio, argumentos diferentes de los entimemas. En cualquier caso, es un hecho el que, en muchas ocasiones, aceptamos ciertas ideas como verdaderas por la mera credibilidad atribuida a la fuente comunicativa. Uno puede creer en la existencia de agujeros negros porque es lo que dicen ciertos astrónomos reputados, pese a no entender las razones que éstos esgrimen. Otro, puede creer que el papel higiénico es un bien muypreciado en tiempos de pandemia y confinamiento, pues así parece serlo para la mayoría. Las creencias de origen social-testimonial han sido tratadas con mayor profundidad las sec. 1.1., 1.3., cap. 1, pt. II; y 1.1.1.3., cap. 2, pt. II). Así pues, si fuera un esquema tópico de argumentación, la estructura de este esquema podría ser como sigue:

- Si $P \wedge Ts(P, Q) \rightarrow Q$

Que se leería como “si ha ocurrido P, y una fuente (experto, mayoría) atestigua (Ts) que P está unido a Q, entonces ha ocurrido, ocurre u ocurrirá Q”. Baste aquí llamar la atención entonces sobre el hecho de que la relación que se establece entre las variables P y Q es la más arbitraria de todas, por lo que su refutabilidad es mayor. En estos esquemas, no se evalúa tanto el contenido del tipo ‘A es B’ como la *credibilidad de la fuente* comunicativa y la *relación* entre esa fuente con el contenido en cuestión. Uno puede acabar atribuyéndole una credibilidad total o relativamente alta a una idea muy contraintuitiva del tipo ‘A es B’ si es sostenida por la mayoría o por alguien a quien considere experto en el tema. No es que uno considere que siempre o la mayoría de las veces A sea, de hecho, B. Más bien, considera que siempre o la mayoría de las veces que la opinión popular o el experto sostienen que ‘A es B’, llevan razón. Si bien los tópicos de base empírica eran el resultado de un

proceso de comprensión de relaciones por parte del individuo, en este caso, el individuo cree que 'A es B' porque confía en que esa relación ha sido comprendida por un tercero a quien otorga alta credibilidad. Como se dijo al final de capítulo 2 (i.e. la *asignación de expertos* de Gaifman, 1988), si la aseveración crítica es sobre un tema desconocido por el individuo, puede acabar creyéndola en la misma medida en que es creída por una fuente fiable. En cualquier caso, en tanto que su garantía es más débil que en los casos anteriores, puede afectar a su refutabilidad y, por tanto, a la estabilidad de su credibilidad (para algunos estudios empíricos en psicología sobre estas cuestiones véase Hahn et al., 2009; Hahn et al., 2013; Harris, Hahn, Madsen, y Hsu, 2016; Harris, Hsu, y Madsen, 2012).

Credibilidad de los tópicos apriorísticos de base inferencial o intuitiva

Este tipo de esquemas son los que mayores garantías ofrecen para la credibilidad puesto que no admiten la posibilidad de albergar excepciones. Su credibilidad es total e independiente de la experiencia. Ahora bien, esto no implica que cuando un argumento se remita a este tipo de principios tópicos no haya variaciones y derive siempre en niveles máximos de credibilidad para la conclusión. Siempre puede haber variaciones, pero éstas no se deberán a excepciones en la regla de relación tópica sino a la adecuación de su aplicación a los contenidos en cuestión. Serían máximas inferenciales como:

- Dos predicados contrarios no pueden ser verdaderos a la vez, pero sí falsos a la vez.
- El todo es mayor que la parte y la parte es inferior al todo.
- Lo que se predica universalmente del género, también de la especie; lo que se predica de la especie, se predica del género particular o universalmente.
- De dos cosas que sobrepasan a una tercera, la que lo hace en mayor medida es la mayor de las dos.
- Si algo se predica de A, y A es la definición de B, entonces también se predica de B.

Todos estos esquemas descansan sobre la base de relaciones verdaderas y primordiales; esto es, de relaciones que tienen credibilidad por sí mismas sin necesidad de nada externo ni anterior a ellas. Son principios cognitivos fundacionales que no requieren de contrastación empírica con el mundo para ser creídos y, por lo general, se toman como necesarios. Al ser rellenados o aplicados adecuadamente a los contenidos, garantizan la total transmisión de la credibilidad del contenido de las premisas al de la conclusión. Así pues, como siempre, sí, se evaluará la verdad del contenido temático de las premisas. Sin embargo, en el empleo de este tipo de esquemas, la evaluación principal recae sobre la *adecuación* del contenido al esquema. Por otra parte, la verdad de la relación tópica del esquema está fuera de discusión. Por ejemplo, imaginemos un argumento construido sobre el esquema de la definición, el cual afirma que 'A es B'. La conclusión que surja de ese argumento se sostiene sobre la base de aceptar que 'B es (la definición de) A'. Para que este argumento sea efectivo, el oyente no solo deberá creer que 'A es (de hecho) B', sino que deberá evaluar hasta qué punto B es realmente la definición de A, y no un atributo no definitorio de A. Esto es, no solo deberá creer que es verdad que "el ser humano es un bípedo implume", sino que deberá

evaluar si ese predicado es propiamente aquello que define al sujeto o no. En efecto, esta proposición puede ser verdadera sin que coincida con su definición. Este juicio sobre la adecuación podrá producir variabilidad en la credibilidad que se otorgue a la conclusión del argumento. Si se considera que el tópico está bien adecuado, la transferencia de credibilidad será total.

Relaciones de la credibilidad con la validez del esquema inferencial formal

Los dos niveles de abstracción que acabamos de ver se circunscriben a la premisa mayor, ya sea a su contenido o a su estructura interna como dispositivo relacional. Sin embargo, el esquema inferencial formal surge de la combinación de la forma de la premisa mayor con la de la menor. Es decir, es un esquema de la estructura formal del argumento en su conjunto. La característica más relevante de estos esquemas en relación con su capacidad de transmitir propiedades a la conclusión es la validez de la forma.

Los esquemas inferenciales formales lógicamente válidos derivan en verdades formales, necesarias y, por tanto, irrefutables. Su carácter irrefutable es formal. Por ejemplo, si “todos los felinos son animales voladores”, y “los perros son felinos”, la conclusión de que “los perros son animales voladores” es *formalmente* irrefutable. Sin embargo, decir que “todos los lobos son cánidos y todos los zorros son cánidos, por tanto, los zorros son lobos” es un argumento formalmente inválido y siempre será refutable. He escogido premisas falsas para el argumento válido y verdaderas para el inválido a fin de ilustrar esta cuestión. Cuando se emplean premisas (consideradas por los oyentes como) verdaderas en ambos, el argumento válido será más fuerte que el inválido, puesto que este último podría ser refutado *por* su forma. Esta opción no está disponible en el válido. Es posible que ambos logren producir credibilidad total en un individuo, pero sólo el segundo será susceptible de ser rebatido. Observemos otro ejemplo de argumento inválido: “si llueve, la calle estará mojada; observo que la calle está mojada; por tanto, concluyo que ha llovido”. Una persona podría creer totalmente en esta conclusión a partir de este argumento, pues parece un buen argumento y su contenido es probable. Y dependiendo de la situación contextual podría ser lo más probable. En todo caso, siempre podría someter su razonamiento a un mayor escrutinio y darse cuenta de que el que la calle esté mojada no implica necesariamente que haya llovido, por lo que no se puede extraer esa conclusión, pese a que su contenido sea posible. La lluvia es sólo una de las múltiples causas por las que el suelo puede estar mojado. Si uno entra a considerar otras posibilidades, podría asignar valores de credibilidad algo más bajos a las conclusiones surgidas de formas inválidas.

Con todo esto se llega a la hipótesis de que la disposición formal del argumento en su conjunto debe tener efectos sobre la credibilidad de la conclusión. Ahora bien, es posible que tales efectos sólo se observen en determinadas circunstancias de interacción con la consideración subjetiva sobre la verdad de las premisas (y la conclusión). En este sentido, hay que tener en cuenta que la forma de los argumentos puede ser válida o inválida, y que las premisas pueden ser consideradas como más o menos verdaderas o más o menos falsas. Del mismo modo, las conclusiones pueden partir de un estado inicial de alta o baja credibilidad.

Imaginemos que la conclusión parte de un estado inicial de baja credibilidad. En ese caso, cuando se considera que las premisas son hasta cierto punto verdaderas, la validez formal del argumento garantiza que la conclusión también pasará a ser considerada hasta cierto punto verdadera. Sin embargo, en las mismas condiciones de verdad de las premisas, la invalidez formal del argumento podría bloquear la transmisión de esa verdad a la conclusión. Por otro lado, cuando se considera que las premisas son más o menos falsas, la validez formal del argumento perderá su relevancia sobre la credibilidad de la conclusión. Si la conclusión no era creída, seguirá sin serlo independientemente de que el argumento sea válido o inválido.

Ahora imaginemos que la conclusión parte de un estado inicial de alta credibilidad, por las razones que sean. En ese caso, cuando se considera que las premisas son verdaderas hasta cierto punto y el argumento es válido podría suceder que la credibilidad de la conclusión se viera reforzada al mostrar nuevas razones que convergen sobre la conclusión para su aceptación. De este modo, aumentaría la *estabilidad* de esa conclusión, aunque no necesariamente su credibilidad. Dependiendo del grado de credibilidad de las premisas, la credibilidad de la conclusión podría permanecer igual que antes, o podría aumentar (este punto se desarrolló brevemente en sec. 1.3.3., cap. 2, pt. II). Sin embargo, si el argumento fuera inválido, la transmisión de la verdad de las premisas podría verse bloqueada, no afectando ni a la estabilidad y ni a la credibilidad de la conclusión. Por otro lado, cuando se considera que las premisas son más o menos falsas, el argumento en su conjunto (y su validez) perderá su relevancia sobre la credibilidad de la conclusión. Si la conclusión ya era creída por las razones que fueran, seguirá siendo creída independientemente del argumento⁷⁸.

Por tanto, la validez del argumento es una condición deseable, pero, por sí misma, no es necesaria ni suficiente para garantizar la credibilidad de la conclusión. Por su parte, la alta credibilidad de las premisas es una condición necesaria, pero no es suficiente. Finalmente, la credibilidad de las premisas añadida a la validez del argumento sí que serían condiciones conjuntamente suficientes.

Esquema de argumentación de la causa al efecto

Este esquema de argumentación es el escogido para los estudios experimentales realizados en esta tesis, razón por la cual lo presentaré aquí con algo más de detalle. La primera vez que se habla de un esquema tópico parecido es en la Retórica de Aristóteles, siendo uno de los cinco esquemas tópicos relativos al empleo retórico del principio de causalidad. Este esquema dice que “otro lugar común procede de la causa: si se predica pertinentemente se dice que algo es, y si no, que no es, porque la causa y aquello de lo que es causa se dan juntos; y sin causa, nada existe.” (*Ret.* 1400a 30-

⁷⁸ Evidentemente, lo que se está considerando aquí son argumentos que pretenden apoyar la conclusión, no argumentos que pretenden refutar las premisas o la conclusión. La refutación merecería un capítulo aparte, y no será considerada en esta tesis más que en momentos puntuales.

35). En todo caso, aquí hay que advertir que la dirección con que se presenta la relación de causalidad es doble: de la causa al efecto, y del efecto a la causa. Según Perelman et al. (1958), los argumentos basados en el nexo causal se construyen sobre la ocurrencia de un evento condicionante que se puede tomar como efecto o como causa en un terreno mecánico o teleológico. En el presente trabajo no abordaré el terreno teleológico de la causalidad, mucho más complejo, por lo que queda excluido de todo lo que diga a continuación.

Cuando la ocurrencia del evento se toma como efecto, se trata de descubrir o mostrar la existencia de una causa que haya podido determinarlo. Como ya vimos en el capítulo 3, ir de los efectos a las causas es característico de la inducción. Es un razonamiento de tipo diagnóstico o retrodictivo, orientado hacia el pasado. Al tomarlo como efecto, su formalización probabilística se correspondería con alguna expresión análoga a $P(\text{Causa} \mid \text{Efecto})$ y la magnitud resultante dependerá de la *necesidad* de la causa para el efecto. Si es totalmente necesaria, la magnitud de esa probabilidad será de 1. Por otro lado, cuando el evento se toma como causa, se procura evidenciar la existencia del efecto que debe resultar de ello. En este caso, se trata de un razonamiento predictivo, orientado hacia el futuro. Puesto que la ocurrencia del efecto queda condicionada a la ocurrencia de la causa, su formalización probabilística se correspondería con $P(\text{Efecto} \mid \text{Causa})$ y la magnitud resultante dependerá de la *suficiencia* de la causa para el efecto. Si es totalmente suficiente, la magnitud de esa probabilidad será de 1.

El moderno esquema de argumentación “de la causa al efecto”, definido por Walton (e.g. Macagno y Walton, 2014; Walton, 1996; Walton y Macagno, 2015; Walton et al., 2008), dice que “allá donde esté presente una causa, se puede esperar un efecto” y adopta la forma de una simple implicación del tipo $\text{Causa} \rightarrow \text{Efecto}$. Según estos mismos autores, la estructura del argumento se presenta como:

- Generalmente, si A ocurre, entonces B ocurrirá o podría ocurrir.
- En este caso, A ocurre o puede ocurrir.
- En este caso, B ocurrirá o podría ocurrir.

En este esquema de argumentación⁷⁹, la dirección de la inferencia se hace explícita en su propio nombre, así como en su estructura. Walton (e.g. Macagno y Walton, 2014; Walton, 1996) lo considera un esquema de tipo deductivo que descansa sobre una generalización causal derrotable, consistiendo en un *Modus Ponens Derrotable*. Y debido a esta dirección inferencial, el criterio de evaluación de la fuerza del argumento estará basado en la suficiencia de la causa para el efecto. En este sentido, Walton (1996) propone una serie de preguntas críticas para realizar dicha evaluación de este esquema de argumentación:

⁷⁹ Una posible crítica a esta formulación del esquema es la de que la relación tópica de la causalidad está totalmente ausente. Este mismo esquema podría ser compartido por una gran cantidad de esquemas de argumentación puesto que, en general, tratan de convencer acerca de la existencia de algo a partir de la constatación de la existencia de otra cosa.

1. ¿Cuán fuerte es la generalización causal (si es que, de hecho, es verdad en algún sentido)?
2. La evidencia aducida (si es que hay alguna) ¿es suficientemente fuerte para garantizar la generalización en la que se basa el argumento?
3. ¿Existen otros factores que podrían interferir (o que vayan a interferir) o impedir la producción del efecto en este caso?

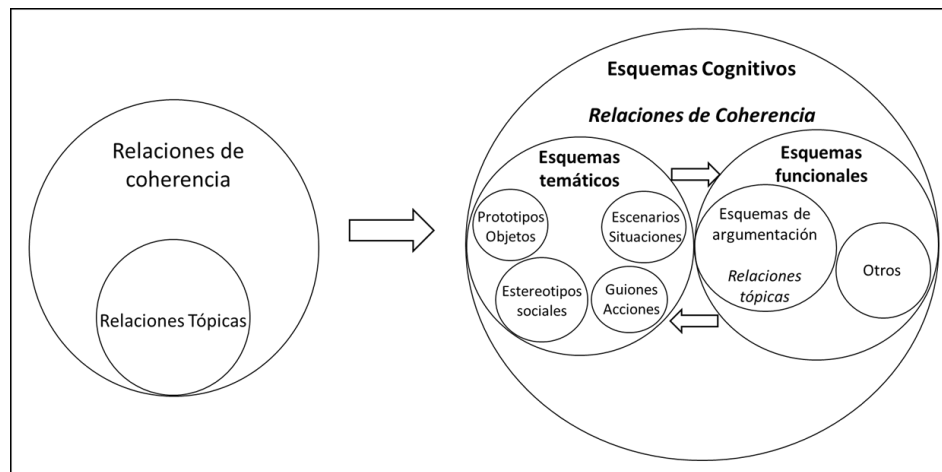
Las tres preguntas críticas me parecen interdependientes. Si la generalización causal es máximamente fuerte, entonces se puede esperar que siempre que ocurra la causa, vaya a ocurrir el efecto. En estas circunstancias, si la evidencia aducida es la causa, entonces se esperará que la causa sea suficiente para garantizar la obtención del efecto y, por tanto, para asegurar la generalización causal. Si la causa es suficiente para el efecto, entonces se espera que no haya factores que puedan interferir o impedir la producción del efecto. Sin embargo, si hubiera dicho factor, entonces la causa ya no podría considerarse como suficiente para el efecto, por lo que la fuerza de la generalización causal no sería máxima. En este sentido, se esperaría que, cuanto mayor sea la frecuencia de aparición de esos factores interferentes, menor sea la suficiencia juzgada de la causa para el efecto y más débil sea la generalización causal. Como ya vimos en el capítulo 4, este tipo de factores ha recibido el nombre de rebatidores refutativos en epistemología (Pollock, 1987), y de condiciones desactivadoras en psicología del razonamiento causal (Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003). Tal y como se dijo en ese capítulo, las condiciones desactivadoras disminuyen la suficiencia percibida del condicional causal, derivando en una menor aceptación de las conclusiones obtenidas en argumentos con la estructura *Modus Ponens* (MP) y *Modus Tollens* (MT). Y los argumentos de la primera figura con los que se trabaja en los experimentos de esta tesis funcionan de un modo similar al MP. Por otro lado, también vimos en el capítulo 4 que existe una relación entre el razonamiento cotidiano con condiciones desactivadoras y el razonamiento probabilístico. Los resultados de la investigación empírica son compatibles con la hipótesis de que detrás de un juicio subjetivo de probabilidad acerca de una conclusión MP y MT se esté produciendo la activación o preactivación de condiciones desactivadoras en la memoria a largo plazo. En este sentido, Hahn y Hornikx (2016) consideran que, si bien es cierto que la perspectiva probabilística es ciega a la diversidad de las relaciones tópicas de los esquemas de argumentación, es requerida por las preguntas críticas que Walton et al. (2008) formulan para cada uno de dichos esquemas. Es decir, la probabilidad entra en juego a la hora de evaluar la fuerza establecida por las relaciones tópicas en función de los refutadores activados. Es, básicamente, una forma de medir la fuerza de esas relaciones.

Como se ha ido viendo a lo largo de este apéndice, los entimemas son silogismos cuya validez es más psicológica que lógica, y cuyas premisas son representaciones más o menos creídas o, si se quiere, subjetivamente verdaderas más que verdaderas. No importa si en la realidad son o no verdaderas. Del mismo modo, se puede apelar a los diversos criterios tópicos de relación entre dos elementos en tanto que los sujetos acepten que se da esa relación entre ellos, sin importar si en

realidad se relacionan por ese criterio. No importa si realmente A es la causa de B. Para que un entimema basado en tópico de la causalidad sea exitoso en la modificación de creencias lo relevante es que el individuo acepte A como causa de B. De este modo, en un argumento con el esquema 'de la causa al efecto' pueden emplearse elementos que, tal vez, en la realidad sólo estén correlacionados, o que tengan una causa común. Lo único importante es que los individuos creen que uno es la causa del otro, de tal modo que su creencia en el efecto en presencia de la supuesta causa sea mayor que en su ausencia. Y lo mismo pasa con los refutadores. Por un lado, no es tan importante que determinado evento impida o dificulte la ocurrencia de otro en la realidad, ni tampoco que determinada información refute legítimamente la aseveración acerca de una ocurrencia. Lo único importante es que el individuo crea que la refuta. Por otro lado, no bastará con eso, sino que, crucialmente, el individuo deberá poder acceder a dicha información en su memoria a largo plazo. Ese acceso depende de cuan fuertemente asociados estén los elementos pertinentes en las estructuras de memoria activadas en una situación o contexto concreto. Cuanto mayor sea el acceso a los refutadores, menor será la credibilidad atribuida al condicional y, consecuentemente, menor será la capacidad de transmisión de la credibilidad de la premisa menor a la conclusión, resultando en una menor credibilidad de la conclusión.

Esquemas, marcos, guiones y esquemas de argumentación

Llegados a este punto, tenemos ya una idea bastante aproximada de lo que es un esquema de argumentación. En todo caso, me he limitado a señalar superficialmente que, en cierto modo, los esquemas de argumentación están relacionados con los conceptos de esquemas temáticos, marcos y guiones propuestos en psicología e inteligencia artificial (Bartlett, 1932; Minsky, 1975; Rumelhart, 1980; Schank y Abelson, 1977). En esta tesis sostengo que los esquemas de argumentación son otra clase más de esquema cognitivo. A saber, un tipo de esquema más abstracto, con relaciones como referentes, sin pertenencia a dominios específicos de conocimiento. Los esquemas de argumentación pueden estar presentes también en la estructura de los esquemas temáticos clásicos referenciados más arriba. En este sentido, se pueden tomar como pautas o reglas de razonamiento apresadas en representaciones complejas que vinculan sus microelementos representacionales. Mediante estas reglas, la activación de uno de tales elementos consigue activar el concepto complejo completo, tal como operan las reglas de producción. Por otro lado, mientras que los esquemas temáticos clásicos están vertebrados por relaciones de coherencia (Sanders, 2005), los esquemas de argumentación lo están por relaciones tópicas. A este respecto, también considero que las relaciones tópicas son un subconjunto de las relaciones de coherencia. De este modo, los conceptos de esquema de argumentación y de relaciones tópicas procedentes del campo de la argumentación quedan naturalmente integrados en el aparato conceptual de la psicología cognitiva, tal como pretende representar la siguiente figura. Por este motivo, me parece conveniente inspeccionar un poco más de cerca cómo pueden organizarse las relaciones mencionadas.



En la figura se representa a todas las relaciones tópicas como pertenecientes al conjunto de relaciones de coherencia. Tanto los esquemas temáticos como los esquemas funcionales pertenecen al conjunto más amplio de esquemas cognitivos. Todos los esquemas cognitivos están vertebrados por relaciones de coherencia. En particular, el tipo de relaciones de coherencia que funciona en los esquemas de argumentación es el de las relaciones tópicas. En ocasiones, los esquemas de argumentación pueden subyacer a la estructura de algunos esquemas temáticos, de tal modo que ciertas relaciones tópicas también puedan estar interviniendo en la configuración de tales esquemas.

No nos vamos a detener en especificar los detalles de los conceptos de esquemas cognitivos, marcos y guiones más allá de lo directamente relevante para seguir esta explicación. Partiremos de la idea de que todo esquema tiene las siguientes propiedades básicas:

- 1) Es una unidad informativa compleja; esto es, está compuesta de elementos más simples. Crucialmente, no consiste en un mero listado de componentes, sino en la *estructura* que se genera a partir de las *relaciones* tópicas que se establecen entre esos componentes.
- 2) Es una estructura abstracta, pese a que se puedan dar variaciones en su grado de abstracción en función del tipo de esquema. En este sentido, sus componentes representan *variables*, categorías o celdillas vacías que pueden ser rellenas por un número indeterminado de *constantes*, esto es, de instancias más concretas de la categoría en cuestión. Por tanto, son matrices estructurales generales de aplicación a muchos casos específicos.
- 3) Puede ser apriorístico o posterior a la experiencia. En ambos casos consisten en conocimientos implícitos o explícitos sobre estructuras estereotipadas o prototípicas. Pero sólo los esquemas posteriores a la experiencia se obtienen mediante procesos inductivos aplicados sobre el flujo de información entrante. Estos procesos abstraen sus características más frecuentes, las descontextualizan y permiten su generalización y aplicación a nuevas instancias particulares.

- 4) Es un tipo de representación del conocimiento que tiene la función de facilitar la comprensión enriqueciendo y estructurando la información entrante y sirviendo de base para la generación de inferencias.

Como se ha explicado en las páginas anteriores, el contenido de la premisa mayor de los entimemas está ejemplificando una regla general de cobertura que legitima la realización de inferencias que vayan de la premisa menor a la conclusión. Esa regla general se concibe en esta tesis como un esquema. Y un esquema consiste en un conocimiento general, abstracto y estereotipado sobre la estructura de aquello que pasa siempre o la mayoría de las veces. Para abordar estas similitudes y diferencias entre los esquemas de argumentación y los esquemas temáticos cognitivos clásicos deberé volver otra vez sobre la cuestión de los diferentes niveles de abstracción.

Niveles de abstracción en los esquemas: temáticos y funcionales

Tal y como ya se mencionó más arriba, se puede hablar de unos esquemas temáticos con un referente material relativo a entidades del mundo externo y pertenecientes a géneros o dominios de conocimiento relativamente definidos como los esquemas de objetos (p. ej. de una cara, de un pájaro, de un aula, etc.) o como los esquemas de eventos y situaciones concretas, pero no singulares (p. ej. de ir a un restaurante, de un acto comercial, de una tormenta, de un proceso legal, etc.). Así, sabemos que la mayoría de las veces, una cara se compone de ojos, nariz y boca dispuestos de un modo determinado, o que un restaurante se compone de comensales, camareros, comida, bebida, dinero y de una serie de acciones ordenadas de un modo determinado. Estos esquemas se podrían aplicar a una gran diversidad de instancias específicas y singulares, como la cara de este niño, o de este perro, o como aquella vez que fui a aquel restaurante, etc. En estos casos, cuando se singularizan a modelos de experiencias, se convierten en modelos de situación.

También hay esquemas con referentes más abstractos, vacíos, casi formales. Algunos de ellos también pertenecen a cierto tipo de géneros, como los esquemas de los distintos géneros discursivos (p. ej. del descriptivo, del narrativo, del expositivo o del persuasivo) que, a su vez, tienen subgéneros (p. ej. el dramático, el periodístico, el científico, el político, etc.). Por ejemplo, sabemos que, la mayoría de las veces, el discurso narrativo se compone de una introducción o presentación, un nudo y un desenlace y, la mayoría de las veces, en ese orden. Luego, ese esquema se puede aplicar a subgéneros como la tragedia o la comedia, etc. Y como esto, muchos otros géneros que se desarrollan en una secuencia temporal también se componen de la presentación de las piezas, la mezcla y el producto final. Por ejemplo, los esquemas pertenecientes a la gastronomía, a la construcción y montaje, y a una gran cantidad de actividades humanas. Ese orden esquemático contribuye a la comprensión de la actividad. Sin embargo, otros esquemas no pertenecen a ningún género ni dominio definido, como el esquema general de 'situación' (compuesto por las variables de espacio, tiempo, causación, agentes, metas, acciones, etc.) aplicable a situaciones de dominios más específicos (como restaurante, acto comercial, tormenta, etc.). Los esquemas de argumentación se encontrarían también en esta categoría de esquemas de mayor abstracción, casi formales y sin un género o dominio particular. Sus referentes no son entidades, sino estrictamente

relaciones binarias. En tanto que no pertenecen a ningún dominio de conocimiento específico, se pueden aplicar a cualquiera de ellos. A mi juicio, eso es lo que ha llevado a algunos teóricos de la argumentación a decir que los esquemas de argumentación no son conocimiento en ningún sentido (e.g. Walton, 2001). Como hemos visto, algunos de estos esquemas de argumentación son: el “de la causa al efecto”, el “de las consecuencias”, el “a fortiori”, el “de contrarios”, el de “semejanza”, y muchos otros. En cualquier caso, considero que sigue siendo cierto tipo de conocimiento, pues se sabe que, la mayoría de las veces, cuando se produce uno de los miembros de la relación, se suele producir también el otro. Por ejemplo, la mayoría de las veces que se da una posible causa, se da también su efecto, o que cuando se produce lo menos probable, también se produce lo más probable acerca de una misma cosa. Su aplicación genera expectativas, permite hacer predicciones y realizar inferencias. En este sentido, los esquemas de argumentación son un subtipo funcional y relacional de la noción más amplia y general de esquema cognitivo. Ahora bien, estos esquemas funcionales siempre se darán ejemplificados por contenidos materiales pertenecientes a dominios de conocimientos genéricos más definidos.

De hecho, muchas veces encontramos que unos esquemas se integran con otros. Por ejemplo, el esquema material de una tormenta implementa un esquema más abstracto de una situación con elementos que se deben situar en algún espacio, donde se producen algunas acciones o fenómenos que deben transcurrir en una línea temporal y que se relacionan entre sí mediante relaciones causales que permiten hacer predicciones sobre posibles efectos.

Esquemas como generadores de creencias

Por todo lo dicho, cabe decir que la premisa mayor de los entimemas puede consistir en un contenido que esté activando y ejemplificando algún esquema como los siguientes:

- a) Un esquema material cualquiera de dominio temático en la forma de una creencia explícita. Por ejemplo, “si alguien está en el salón, entonces está en el sofá”.
- b) Un esquema de argumentación propiamente dicho, de tal modo que:
 - a. El contenido sea una creencia explícita que ejemplifique un esquema material de dominio temático que se superpone a la estructura del esquema de argumentación. Por ejemplo, “si hay nubarrones, entonces lloverá”.
 - b. El contenido sea una creencia implícita que ejemplifique directamente un esquema de argumentación. En estos casos, pese a que sus dos elementos M y P no se han dado repetidamente en la experiencia, quedan vinculados en virtud de la relación tópica. Por ejemplo, “si se siguen emitiendo gases invernadero, entonces se extinguirán los osos polares”.

Es decir, no es necesario que la premisa mayor implemente siempre esquemas de argumentación, pese a que sea lo más común; basta con que sea un esquema. Es importante destacar que cuando la premisa mayor consista en una creencia implícita formada *ad hoc* por dos elementos que no habían coincidido antes en la experiencia, se hacen vinculables y aceptables *por estar ejemplificando un esquema de argumentación*. Sigue tratándose, pues, de un esquema. Y todo

esquema permite, por su estructura, interpretar la información entrante y producir inferencias. En palabras de Kintsch (1974) a propósito de las implicaciones e inferencias lógicas y pragmáticas: “*p* implica *q* si y solo si dada una *memoria semántica S* que contiene un conjunto de *reglas de inferencia I*, y una proposición dada *p*, *q* es derivable de *p*” (p. 57), donde esas reglas de inferencia contenidas en la memoria semántica pueden interpretarse como esquemas, ya sean temáticos o funcionales.

Volvamos a aquel ejemplo de María que vimos varias páginas atrás. Supongamos que sabemos de qué María se trata. Sabiendo que está en el salón (*p*), podremos creer que está en el sofá (*q*) al aplicar el esquema pertinente de que, la mayoría de las veces, si alguien está en el salón, está en el sofá (i.e. la regla *I* que está en *S*). Al menos, eso es lo que todos o la mayoría de los miembros de nuestra cultura actual sabe. No hay aquí ningún esquema de argumentación propiamente dicho, sino simplemente un esquema temático sobre situaciones cotidianas. Aun así, unido al dato inicial, permite producir una creencia.

Incluso hay casos más extremos donde los esquemas temáticos clásicos producen creencias en ausencia aparente de argumento alguno, aunque sí que se daría un razonamiento automático (que irá entre paréntesis). Dado que la información entrante atendida *S* siempre es parcial, al identificarla como una *clase M* (i.e. $S \in M$), activa el esquema pertinente (i.e. $M \rightarrow P$). Al hacerlo, los individuos incorporan automáticamente información de ese esquema para completarla y así interpretarla (i.e. $S \wedge P$). Más tarde, recuerdan la información inferida del esquema como si realmente hubiera estado presente en el input original de información entrante (Bransford, Barclay y Franks, 1972; Graesser, Gordon y Sawyer, 1979). En este sentido, la información inferida representa una nueva creencia.

Según Loftus (1996), esto explica que, en algunos casos judiciales, los testigos afirmen recordar hechos *típicos* del caso genérico (como los casos de abuso sexual, o de robos o de asesinatos) que realmente no ocurrieron en su caso particular. Pero, crucialmente, los individuos *creen* que sí ocurrieron. De modo similar, en un estudio de Loftus y Palmer (1974) se mostraba a los participantes una imagen de un accidente de tráfico y se les pedía que estimaran la velocidad del coche cuando “se estrelló” vs cuando “se golpeó” contra el árbol; además se les pidió que recordasen la imagen y que la describieran. Al activarse el esquema correspondiente con la entrada verbal, los de la condición “estrellarse” no sólo estimaron una velocidad más alta que los del otro grupo, sino que ‘recordaron’ haber visto cristales rotos y otros elementos consistentes con el esquema de estrellarse que realmente no estaban en la imagen. Pero ellos *creyeron* que sí. Este ejemplo guarda exactamente la misma estructura que el anterior: al darse un hecho particular y considerarlo como perteneciente a una clase, se le añaden características típicas de la clase como conclusión. Así, lo que finalmente se recuerda es esa conclusión, esa información enriquecida. Es esa información la que será finalmente creída.

Como explican Harris y Monaco (1978), estas inferencias derivadas del esquema pueden llegar a producirse, simplemente, al procesar una oración. Por ejemplo, al procesar “la hambrienta pitón *cogió* al ratón”, los individuos recuerdan haber leído u oído que “la hambrienta pitón *se comió* al

ratón”⁸⁰. Sin embargo, ese efecto no sucede si es “el hambriento señor” el que “cogió al ratón”. La inferencia se deriva del conocimiento general acerca de las serpientes pitón y sus posibles presas e incluso del contexto producido por el adjetivo “hambrienta”. Y, en tanto que la inferencia es recordada, constituye una creencia: los individuos *creen* haber recibido esa información⁸¹. Así pues, se infieren antecedentes, consecuentes, causas, efectos, espacios, tiempos, instrumentos, agentes, intenciones, etc. que realmente no estaban en la entrada informativa, pero sí en el esquema activado por esa entrada. Si bien, en condiciones normales, la mayoría de las veces es un mecanismo útil para la comprensión, en determinadas circunstancias puede derivar en creencias incorrectas. En cualquier caso, lo relevante aquí es que producen creencias.

Relaciones de coherencia y relaciones tópicas

La activación de los esquemas juega un papel destacado en la obtención de representaciones coherentes de la información parcial entrante (Brandsford y Johnson, 1972). Los esquemas son estructuras de relaciones coherentes entre componentes. Las celdillas o variables correspondientes con las etiquetas de, por ejemplo, agente, metas y acciones se relacionan coherentemente mediante conexiones causales, temporales, espaciales, de agencialidad, de intención, de reciprocidad, de consecuencia, de género y especie, etc. (Graesser, Gernsbacher y Goldman, 2000/1997). Las relaciones de coherencia son los vínculos que articulan y vertebran la representación esquemática (Sanders, 2005). De este modo, en la comprensión, el flujo de información entrante va quedando situado en estas estructuras independientemente del orden de entrada. Esto es lo que pasa, por ejemplo, en las analepsis, como ocurre en el tan comentado fragmento de *Sueños de Einstein* de Alan Lightman (1993), donde se narran unos acontecimientos en orden inverso, pero que se pueden entender correctamente mediante la activación del esquema de género pertinente, en este caso, el de la ‘distribución de fruta’ (Graesser et al., 2000/1997; Graesser, Millis y Zwaan, 1997).

En este sentido, según Pagin (2014), las relaciones de coherencia son relaciones binarias entre los contenidos de una representación. Tienen la función de mantener unidos los distintos componentes de la representación de tal modo que formen un todo organizado, en lugar de un listado de elementos inconexos. Este autor considera que el enriquecimiento pragmático (i.e. adición de información inferida) que se produce durante la comprensión tiene por objetivo aumentar el grado de coherencia de la representación. Esta idea presupone que la coherencia se da en una escala de grados de fuerza (Pagin, 2014; van den Broek, Bohn-Gettler, Kendeou, Carlson y White, 2011). Esto le lleva a proponer una jerarquía de relaciones de coherencia. Según esta jerarquía, algunos tipos de relaciones de coherencia son más débiles, como las de contigüidad espaciotemporal, y otros son progresivamente más fuertes, como las de causa-efecto (ver también

⁸⁰ Pese a que sería consistente con un esquema de argumentación que dice que “si algo estaba a punto de pasar, entonces ha pasado”.

⁸¹ Adviértase, en todo caso, que uno puede creer que le han dicho algo sin que tenga que creer en la verdad de ese algo. No es que crea o deje de creer que, de hecho, una pitón se comió a un ratón. Sino que cree que lo que leyó o escuchó fue que una pitón se comió a un ratón.

Sanders, Spooren y Noordman, 1992, 1993). En estos últimos casos, el segundo contenido del par binario viene preparado o posibilitado por lo representado en el primer contenido. De este modo, el primer contenido hace la función del antecedente (A) de un condicional y el segundo, de consecuente (C). El grado de fuerza de la relación puede variar desde la mera posibilidad hasta la necesidad, dependiendo, la mayoría de las veces, de los conocimientos previos del individuo que lo evalúa. Según el autor, estas relaciones pueden ser adecuadamente caracterizadas en términos de probabilidades condicionadas, donde la $P(C | A) \neq P(C)$. En los casos de relaciones fuertes de coherencia, los vínculos están permitiendo la transición de justificación entre los pares de contenidos, donde un miembro del par sirve de razón explicativa para sostener el otro (Pagin, 2014; Sanders, 2005). El modelo teórico que se desarrolla en esta tesis no comparte necesariamente la jerarquía según *tipos* de relación que propone Pagin, pero comparte totalmente la idea de que las relaciones que permiten la justificación son un subtipo de relaciones de coherencia que funcionan como grados de probabilidad subjetiva condicionada.

En los argumentos, las relaciones tópicas conectan de forma relevante una representación ya aceptada expresada en la premisa menor con una hipótesis propuesta en la conclusión. Esta conexión inferencial garantiza la transmisión de la justificación de la primera a la segunda. Y, como hemos visto, la transmisión que permiten estas conexiones puede variar en fuerza, pudiendo expresarse adecuadamente en términos de probabilidades condicionadas subjetivas. Por tanto, queda claro que las relaciones tópicas que rigen los esquemas de argumentación son un subtipo del concepto más amplio de relaciones de coherencia. Encontramos una idea similar en Rocci (2005) y Macagno y Walton (2014) cuando sostienen que los tópicos son relaciones argumentativas entre las premisas y la conclusión que pueden interpretarse en términos de coherencia del discurso. Conectan las secuencias del argumento de tal modo que unas sirven de razones para creer en otras y así poder atribuir justificadamente un predicado a un sujeto gramatical. Son relaciones semánticas y ontológicas que reflejan la estructura a través de la cual organizamos la realidad.

Conclusión

Ahora me dispongo a recapitular las principales conclusiones de este apéndice. Empezaré, entonces, recordando que los entimemas son silogismos, y que los silogismos son mecanismos de convicción que pretenden mostrar la unión del sujeto gramatical de la conclusión con el predicado de esta mediante la conexión de ambos con un término medio que los vincula. Ese vínculo se produce sobre la forma del argumento, esto es, en función de la disposición de sus tres términos o variables en las premisas dando lugar a las tres figuras silogísticas clásicas. Estas figuras pueden combinarse con una serie de constantes lógicas como la negación, los conectores y los cuantificadores. La ausencia de negación denota que el enunciado es afirmativo, y su presencia, que es negativo. Por otro lado, los cuantificadores denotan si la extensión de los enunciados es universal o particular. La combinación de estas constantes con las figuras da lugar a formas silogísticas válidas o inválidas. Ahora bien, como vimos, no siempre es necesario que en la expresión cotidiana se

incluyan explícitamente los cuantificadores. Cuando no los llevan, se les llama silogismos con modo indefinido.

Al describir los silogismos se mencionó que la validez de su forma garantiza la preservación de la verdad de las premisas a la conclusión. Esto es, que, si las premisas son verdaderas y la forma del argumento es válida, entonces la conclusión es necesariamente verdadera. Los entimemas, al ser silogismos, heredan esta propiedad de transmisión, pero sustituyendo la verdad por la credibilidad.

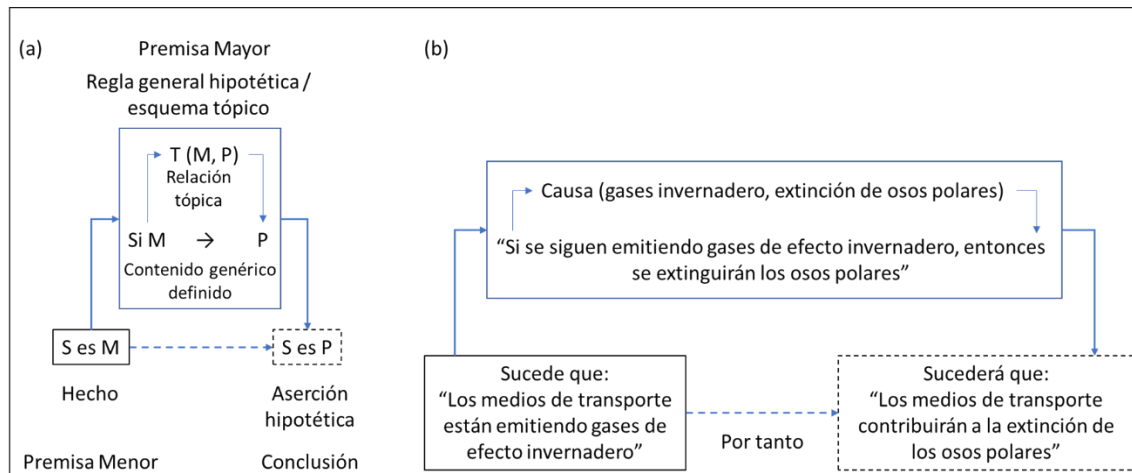
En una situación comunicativa, argumentativa o retórica, los entimemas se pueden emplear tanto para establecer como para refutar una idea. El presente trabajo se limita al establecimiento de ideas, dejando la refutación de lado. Obviamente, establecer una idea implica que no estaba previamente establecida. En este sentido, el estado inicial de credibilidad del que parte esa idea es relativamente bajo. Idealmente inexistente. Entonces, lo que se pretende lograr con los entimemas es transmitir credibilidad a esa idea; una credibilidad que debe proceder de otras que ya la tengan. Estas otras ideas ya creídas funcionarán como premisas del razonamiento retórico, y la idea a establecer, como su conclusión. De este modo, los entimemas preservan la credibilidad de las premisas a la conclusión.

Los entimemas, como todo silogismo, constan de una premisa menor y una premisa mayor. En la premisa menor aparece el sujeto gramatical de la conclusión. Esta premisa tiene la función de constituir la fuente original de la credibilidad, debiendo estar, de algún modo, relacionada con el entorno del individuo. El individuo debe poder identificar de qué elemento particular se está hablando. Normalmente, suele tratarse de hechos establecidos, probados o considerados como reales, altamente creíbles o probables por parte de los oyentes. Por otro lado, en la premisa mayor aparece el predicado de la conclusión. Esta premisa tiene la función de constituir una regla general de relación que permite la transmisión de la justificación de la premisa menor a la conclusión. Se considera que su estructura profunda es la de un condicional. Como se dijo, esta regla general también debe ser previamente creída y aceptada hasta cierto punto por los oyentes. Debe formar parte de sus conocimientos generales acerca de lo que pasa la mayoría de las veces. Así pues, en un entimema, las premisas deben ser más creídas que la conclusión de tal modo que la credibilidad previa residente en aquéllas se transfiera a ésta a través de la forma del argumento.

Las conexiones que permiten la justificación son las relaciones tópicas, que vertebran los esquemas de argumentación. El esquema de argumentación es una estructura casi formal relativamente compleja que permite establecer relaciones legítimas entre el contenido de la premisa menor y el de la conclusión. Esa legitimidad le viene dada por el tipo específico de relación esgrimido. Si no se puede establecer ningún acuerdo sobre ninguna relación entre las dos ideas contenidas en la premisa menor y la conclusión, entonces no será posible transmitir nada. Debe establecerse una relación tópica adecuada subsumida bajo la apariencia de un condicional. En tanto que la mayoría de las veces no queda expresada de forma explícita, esa relación tópica debe ser lo suficientemente accesible para que los participantes puedan captar de qué tipo se trata. En un ejemplo tal como “si piso el pedal del freno, entonces el coche irá más lento”, podemos captar sin

dificultades que la relación entre antecedente y consecuente es de causalidad, y no a fortiori, ni de contrariedad, ni semejanza, ni ninguna otra. El cómo la mente hace eso, es una pregunta que requiere una investigación aparte por propio derecho.

En la siguiente figura, se presenta la estructura completa y el funcionamiento de los entimemas (a), acompañada de un ejemplo concreto (b):



En el esquema (a), las letras S, M, P y T se corresponden con el sujeto, el término medio, el predicado y la relación tópica del argumento respectivamente. En la caja correspondiente a la premisa mayor, se muestra la estructura del esquema tópico de argumentación. El antecedente M del condicional se conecta con su consecuente P mediante una relación tópica T subyacente. Tanto en (a) como en (b), he empleado una línea continua en las cajas de las premisas para indicar que su contenido ya reside en la memoria del oyente; son contenidos ya creídos y aceptados. Correspondientemente, la línea discontinua de la caja de la conclusión indica que su contenido es el que se pretende establecer mediante el argumento. Por su parte, las flechas con líneas continuas muestran el flujo de la transmisión de la credibilidad, mientras que la flecha con la línea discontinua indica que el hecho justifica la creencia en la aserción hipotética contenida en la conclusión de manera indirecta. A lo largo de la Parte II, este esquema ha ido apareciendo en sucesivas ocasiones, con otras formas, enriquecido con una aparato conceptual cognitivo creciente, introduciendo conceptos relevantes como el de explicación, inducción por ejemplificación, y deducción probabilística.

En conjunto, se podría considerar que hay cuatro variables generales que pueden afectar la transmisión de la credibilidad en los entimemas:

- El conocimiento sobre los contenidos materiales de la premisa menor: antes ya se ha dicho que debe consistir en hechos o datos ligados al entorno del sujeto, probados o perceptibles, altamente creíbles. Pero uno no siempre tiene una certeza total acerca de lo que pasa a su alrededor debido a la conciencia de las propias limitaciones cognitivas, epistémicas, comunicativas o memorísticas. Cuanta menos credibilidad le otorgue al contenido de la premisa menor, menos cantidad de credibilidad podrá serle transferida a la conclusión.

- b) El conocimiento sobre los contenidos materiales de la premisa mayor que pueden ejemplificar los esquemas de argumentación. Esta premisa puede estar constituida por dos tipos distintos de creencias:
- a. Creencias explícitas: son esquemas temáticos, con referentes relativos a objetos (prototípicos), situaciones o escenarios típicos, estereotipos sociales, o guiones o pautas generales de acción ante determinados estímulos, en determinadas situaciones. Son creencias generales abstraídas mediante procesos inductivos a partir de lo que pasa la mayoría de las veces en la experiencia episódica. Su construcción y estabilidad se ve claramente afectada por la frecuencia de coocurrencia de los elementos que conforman la representación, lo cual afecta directamente a la probabilidad que se le asigna a dicho contenido. Es decir, a su grado de credibilidad. Pese a ser relativamente estables, pueden ir actualizándose y cambiar a la vez que los cambios ambientales en la frecuencia de esa coocurrencia. Esto es, son modificables con la experiencia, y así también lo es su credibilidad.
 - b. Creencias implícitas: a diferencia de las anteriores, no consisten en creencias cristalizadas en la memoria, sino que se derivan de éstas. Son generalizaciones construidas *ad hoc* para el caso en cuestión. El grado de credibilidad que se les atribuya depende de la combinación de las creencias explícitas. Si las explícitas eran probables, estas heredarán también esa probabilidad. Por tanto, son generalizaciones falibles, a las que se les puede asignar grados de probabilidad. Se construyen mediante la activación de los conocimientos sobre los elementos pertinentes a la representación, evaluando *on-line* la credibilidad de la conexión resultante.
- c) La naturaleza del principio rector tópico de los esquemas de argumentación: del mismo modo en que la comunicación, la experiencia y la intuición o inferencia se consideran las tres fuentes de la generación de las creencias, hay tópicos de base testimonial (o pragmática o comunicativa), empírica (o experiencial o epistémica) e inferencial (o apriorística o intuitiva). Cada uno de ellos ofrece distintos tipos de garantías para el establecimiento de la conclusión.
- a. Los tópicos de base testimonial pueden ser los más débiles puesto que no están basados en la comprensión de la relación que se establece entre los elementos de la representación bajo cuestión por parte del individuo. Más bien, están basados en la confianza que el individuo atribuye a una fuente comunicativa como interlocutor legítimo para realizar aseveraciones acerca de esa representación bajo cuestión.
 - b. Los tópicos de base empírica permiten el flujo adecuado de la credibilidad en el argumento conectando los elementos de la representación mediante algún criterio funcional y organizativo que le dote de legitimidad. Pero toda su capacidad para transferir la credibilidad de la menor a la conclusión les viene impuesta o heredada

de los conocimientos sobre los contenidos materiales de la premisa mayor que ejemplifican el tópico. El esquema como tal no es modificable, pero su aplicación al mundo por parte del sujeto admite excepciones, y por tanto puede variar en el mismo modo en que lo hacían los esquemas temáticos explicados más arriba.

- c. Los tópicos de base inferencial se construyen sobre una línea de razonamiento que es primordialmente verdadera, independientemente del entorno. Si se aplica adecuadamente sobre contenidos considerados verdaderos, permitirá la transición total de credibilidad. Pero no será así si no se aplica adecuadamente, o si los contenidos no son considerados totalmente verdaderos. Por tanto, pese a su naturaleza intuitiva y apriorística, también pueden derivar en conclusiones probables.
- d) La validez del esquema inferencial formal: una alta credibilidad de las premisas combinadas en una forma argumental válida son condiciones suficientes para garantizar la transmisión de esa credibilidad a la conclusión. Pero por separado no garantizan nada. Aun siendo alta la credibilidad de las premisas, si estas se montan sobre una estructura inválida el argumento será susceptible de ser rebatido y su efectividad se verá mermada. El argumento podría seguir funcionando, pese a su invalidez, si el individuo no es capaz de encontrar contraargumentos, si no tiene la capacidad ni la motivación suficiente para someter el razonamiento a un mayor escrutinio. Pero no hay garantías de que el argumento vaya a funcionar. Por otro lado, cuando la credibilidad de las premisas es baja, igual o inferior a la de la conclusión, la validez del argumento deja de ser relevante puesto que no habrá mucha credibilidad que transferir. En este caso, lo que sí está garantizado es que el argumento no será efectivo.

Más tarde se expuso una serie de condiciones que debe reunir cualquier esquema cognitivo, y se profundizó algo más en las similitudes y diferencias de los clásicos esquemas temáticos con los esquemas funcionales de argumentación, y cómo se combinan en las premisas mayores de los entimemas. En esa exposición, he intentado mostrar el encaje de los esquemas de argumentación como un subtipo abstracto del concepto más amplio de esquema cognitivo; un subtipo cuyos referentes consisten en relaciones. Son esquemas funcionales relacionales que, al no pertenecer a ningún dominio definido, pueden aplicarse a todos ellos por igual. Los esquemas de argumentación estarían en la base de la conexión de ideas que permite la justificación en cualquier dominio.

Después, se procedió a explicar cómo la mera percepción de una entrada informativa, al activar un esquema temático, puede derivar en una nueva creencia por inferencia sin necesidad aparente de un argumento. Se mostró, en todo caso, cómo la estructura de esa inferencia es análoga a la de un razonamiento. En la realidad se produce un hecho. El sujeto tiene la experiencia de observar ese hecho. Esa observación consiste entonces en una entrada informativa. La entrada informativa siempre es parcial e incompleta. Se reconoce identificándola como una instancia de una clase de evento. Esa clase de evento está almacenada en la memoria semántica. A esa clase de evento le corresponde un esquema. En ese esquema están contenidos todos los atributos que, para el

individuo, típicamente constituyen la clase. En la mente del individuo, esos atributos están altamente asociados unos con otros. De este modo, la entrada informativa parcial es completada o complementada con esos atributos que, en realidad, nunca se observaron. Se le añaden como conclusión de un razonamiento o inferencia automática, dando lugar a interpretaciones subjetivas de esa entrada informativa. Esa interpretación constituye una creencia.

Otro de los puntos que ha tratado de ponerse de manifiesto aquí es que todo esquema está estructurado por relaciones semánticas (u ontológicas) de coherencia. Así como los esquemas de argumentación son un subtipo de esquema, las relaciones tópicas son un subtipo de relaciones de coherencia: aquellas que rigen los esquemas de argumentación en la transmisión de la credibilidad. Tal como se dijo más arriba, los esquemas de argumentación pueden estar implicados en los esquemas temáticos, de tal modo que las relaciones tópicas también lo estén. En efecto, un esquema es una representación estructurada compuesta de elementos representacionales más simples. Por otro lado, las relaciones de coherencia son relaciones semánticas binarias que conectan los elementos que componen el esquema dotándolo así de estructura. Esa relación binaria se puede entender como la que se da entre un antecedente y un consecuente de un condicional. Las relaciones de coherencia pueden ser más fuertes o más débiles dependiendo de los conocimientos del individuo. Esas variaciones en la magnitud de la coherencia, esto es, en la fuerza del vínculo binario establecido, se pueden cuantificar como probabilidades condicionadas. En ocasiones, las relaciones de coherencia implicadas en el esquema temático se corresponden con las relaciones tópicas. Las relaciones tópicas son aquellas que permiten la transmisión de la justificación. En este sentido, las relaciones tópicas pueden transmitir más o menos justificación en función de los conocimientos del individuo. Del mismo modo, estas variaciones en la magnitud de la justificación transmitida se pueden cuantificar como probabilidades subjetivas condicionadas. El antecedente de ese condicional de la premisa mayor contiene un elemento M que aparece en la premisa menor, y lo vincula al elemento P que aparece en la conclusión. De este modo, la representación SP será tanto más aceptable cuanto mayor sea la justificación que vincule sus partes S y P; esto es, cuanto mayor sea el resultado de $P(P | S)$.

De este modo, se puede ver cómo los conceptos y estructuras existentes en las antiguas disciplinas de la retórica, la dialéctica y la argumentación pueden integrarse en el cuerpo de conceptos que maneja la investigación cognitiva y hacer así del entimema una herramienta muy útil para el estudio psicológico de la dinámica de la credibilidad.

REFERENCIAS

- Abelson, R. P., & Reich, C. M. (1969). Implicational molecules: A method for extracting meaning from input sentences. In *Proceedings of the International Joint Conference on Artificial Intelligence*. Bedford, MA.
- Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: Center for the Study of Language and Information.
- Adams, E. W., & Levine, H. P. (1975). On the uncertainties transmitted from premises to conclusions in deductive inferences. *Synthese*, 30(3/4), 429-460.
- Aikin, S. F. (2011). *Epistemology and the Regress Problem*. New York: Routledge.
- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. N. (2015). *Patterns of attachment: A psychological study of the strange situation*. Psychology press.
- Alba, J. W., Chromiak, W., Hasher, L., & Attig, M. S. (1980). Automatic encoding of category size information. *Journal of Experimental Psychology: Human Learning and Memory*, 6(4), 370.
- Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1459.
- Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1061.
- Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The journal of symbolic logic*, 50(2), 510-530.
- Alcock, J. E. (2018). Belief: What it means to believe and why our convictions are so compelling. *Humanist Perspectives*, (205), 40-41. Retrieved from

- link.gale.com/apps/doc/A555753943/AONE?u=anon~321c9ee5&sid=googleScholar&xid=6d06420e
- Aliseda, A. (2012). Abducción. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 17-22). Trotta.
- Álvarez, J. F. (1995). Dinámica deliberativa y valores epistémicos. *Isegoría*, (12), 137-148.
<https://doi.org/10.3989/isegoria.1995.i12.245>
- Álvarez, J. F. (2012). Explicación vs. argumentación. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 55-66). Trotta.
- Amossy, R. (2002). How to do things with doxa: Toward an analysis of argumentation in discourse. *Poetics Today*, 23(3), 465-487.
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, 22(3), 261-295.
- Andrews, K. (2020). *The animal mind: An introduction to the philosophy of animal cognition*. Routledge.
- Andrews, K., & Monsó, S. (2021). Animal cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). Retrieved from <https://plato.stanford.edu/archives/spr2021/entries/cognition-animal/>
- Angel, H. F., & Seitz, R. J. (2016). Process of believing as fundamental brain function: The concept of credition. *SFU Forschungsbulletin*, 1-20.
- Apter, M. J. (2007). *Reversal theory: The dynamics of motivation, emotion and personality* (2nd ed.). Oxford: Oneworld Publications.
- Areán Álvarez, L. F. (2012). *Fermat: El problema más difícil del mundo*. RBA Coleccionables.
- Areni, C. S. (2003). The effects of structural and grammatical variables on persuasion: An elaboration likelihood model perspective. *Psychology & marketing*, 20(4), 349-375.
- Aristóteles. (1982). *Tratados de lógica (Organon): I, Categorías, Tópicos, Refutaciones sofísticas* (M. Candel Sanmartín, Introducción, traducción y notas). Madrid: Editorial Gredos.
- Aristóteles. (1988). *Tratados de lógica (Organon): II, Sobre la interpretación; Analíticos primeros; Analíticos segundos* (M. Candel Sanmartín, Trad., Introducción y notas). Madrid: Gredos.

- Aristóteles. (1990). *Retórica* (Q. Racionero, Trad., Introducción, comentarios y notas). Madrid: Gredos.
- Aristóteles. (1995). *Física* (G. R. De Echandía, Introducción, trad. y notas). Madrid: Gredos.
- Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. New York: Routledge & Kegan Paul.
- Arnauld, A., & Nicole, P. (1662/2014). *La logique ou l'art de penser* (Ed. D. Descotes). Honoré Champion.
- Arnhart, L. (1981). The rationality of political speech: An interpretation of Aristotle's rhetoric. *Interpretation*, 9(2/3), 141-154.
- Arthur, W. B. (1994). Inductive reasoning and bounded rationality. *The American economic review*, 84(2), 406-411.
- Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
- Ashley, K. D. (1990). *Modeling legal arguments: Reasoning with cases and hypotheticals*. Cambridge, MA: MIT Press.
- Atienza, M. (2011). El derecho como argumentación. En M. Atienza & L. Ferrajoli, *Jurisdicción y argumentación en el Estado constitucional de derecho*. Ciudad de México: Universidad Nacional Autónoma de México.
- Atkinson, Katie, and Trevor Bench-Capon. "Value-based argumentation." *Journal of Applied Logics* 8.6 (2021): 1543-1588.
- Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.
- Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A*, 49(1), 5-28.
- Badura, C., & Berto, F. (2019). Truth in fiction, impossible worlds, and belief revision. *Australasian Journal of Philosophy*, 97(1), 178-193.
- Ballesteros, S. (1992). La representación del conocimiento en los sistemas conexionistas. *Psicothema*, 4(2), 343-354.
- Baratgin, J., & Politzer, G. (2016). Logic, probability and inference: A methodology for a new paradigm. In L. Macchi, M. Bagassi, & R. Viale (Eds.), *Cognitive unconscious and human rationality* (pp. 119-142). Cambridge, MA: MIT Press.
- Baron, J. (1995). Myside bias in thinking about abortion. *Thinking & Reasoning*, 1(3), 221-235.

- Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.
- Barthes, R. (1967). L'analyse rhétorique. En *Littérature et société* (pp. 31-45). Bruselas.
- Barthes, R. (1974). *La antigua retórica* (B. Dorriots, Trad.). Buenos Aires: Ediciones Buenos Aires.
- Barthes, R. (1975). structural Analysis of Narrative. *Criticism: The Major Statements*.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.
- Batanero, C., Cañizares, M. J., & Díaz, J. (1996). *Azar y probabilidad*. Madrid: Editorial Síntesis.
- Battaglia, F. P., & Pennartz, C. M. (2011). The construction of semantic memory: grammar-based representations learned from relational episodic information. *Frontiers in computational neuroscience*, 5, 36.
- Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical Transactions of the Royal Society of London*, 53, 370-418.
- Bear, G., & Hodun, A. (1975). Implicational principles and the cognition of confirmatory, contradictory, incomplete, and irrelevant information. *Journal of Personality and Social Psychology*, 32(4), 594.
- Bench-Capon, T. J., & Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial intelligence*, 171(10-15), 619-641.
- Berlo, D. K. (1960). *The process of communication: An introduction to theory and practice*. New York: Holt, Rinehart and Winston.
- Bermejo Luque, L. (2006). *Bases filosóficas para una teoría normativa integral de la argumentación: Hacia un enfoque unificado de sus dimensiones lógica, dialéctica y retórica* (Tesis doctoral, Universidad de Murcia, Departamento de Filosofía). Murcia: Universidad de Murcia. Recuperado de <http://hdl.handle.net/10201/33420>
- Bermejo-Luque, L. (2009). Argumentation Theory and the conception of epistemic justification. *Studies in Logic, Grammar and Rhetoric*, 16(29), 285-303.
- Bermejo-Luque, L. (2012). Modelo de Toulmin / Garantía / Respado / Reserva / Calificador. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 308-312). Trotta.

- Bermúdez, J. L. (2003). Ascribing thoughts to non-linguistic creatures. *Facta Philosophica*, 5(2), 313-334.
- Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three-and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, 107(3), 1018-1034.
- Blair, J. A., & Johnson, R. H. (1987). Argumentation as dialectical. *Argumentation*, 1(1), 41-56.
- Blair, J. A., & Johnson, R. H. (2000). Informal logic: An overview. *Informal Logic*, 20(2).
- Block, N. (1978). Troubles with functionalism. *Midwest Studies in the Philosophy of Science*, 9, 261-325.
- Bloom, C. P., Fletcher, C. R., Van Den Broek, P., Reitz, L., & Shapiro, B. P. (1990). An on-line assessment of causal reasoning during comprehension. *Memory & cognition*, 18(1), 65-71.
- Boden, M. A. (1988). *Computer models of mind: Computational approaches in theoretical psychology*. Cambridge University Press.
- Bonatti, L. (1998). Why it took so long to bake the mental-logic cake: Historical analysis of the recipe and its ingredients. En M. D. S. Braine y D. P. O'Brien (Eds.), *Mental Logic* (pp. 7-23). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Bonjour, L. (1984). Introduction to the theory of knowledge. *Teaching Philosophy*, 7(1), 64-66.
- BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- BonJour, L. (1989). Replies and Clarifications. En J. Bender (Ed.), *The Current State of the Coherence Theory* (págs. 129-137). Dordrecht: Kluwer.
- Bonnefond, M., Kaliuzhna, M., Van der Henst, J. B., & De Neys, W. (2014). Disabling conditional inferences: an EEG study. *Neuropsychologia*, 56, 255-262.
- Boole, G. (1854). *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities*. London: Walton and Maberly; Cambridge: Macmillan and Co. (Reimpreso en 1958, New York: Dover).
- Boster, J. (2002). Variación cultural. En R. A. Wilson & F. C. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (pp. 1358-1359). Madrid: Editorial Síntesis.
- Boukala, S. (2016). Rethinking topos in the discourse historical approach: Endoxon seeking and argumentation in Greek media discourses on 'Islamist terrorism'. *Discourse Studies*, 18(3), 249-268.
- Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. New York: Basic Books.

- Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago Press.
- Braet, A. C. (1999). The enthymeme in Aristotle's Rhetoric: From argumentation theory to logic. *Informal Logic*, 19(2), 101-117.
- Braet, A. C. (2004). The oldest typology of argumentation schemes. *Argumentation*, 18, 127-148.
- Braet, A. C. (2005). The common topic in Aristotle's *Rhetoric*: Precursor of the argumentation scheme. *Argumentation*, 19(1), 65-83.
- Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85(1), 1-21.
- Braine, M. D. S., & O'Brien, D. P. (1991). A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98(2), 182-203.
- Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic*. Psychology Press.
- Braine, M. D. S., Reiser, B. J., & Rumin, B. (1984). Some empirical justification for a theory of natural propositional logic. En G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 18, pp. 313-371). Academic Press.
- Braithwaite, R. B. (1932-1933). The nature of believing. *Proceedings of the Aristotelian Society*, 33, 129-146.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 717-726.
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3(2), 193-209.
- Brentano, F. (1973). Foreword. En A. C. Rancurello, D. B. Terrell, & L. L. McAlister (Trans.), *Psychology from an empirical standpoint*. Routledge & Kegan Paul. (Trabajo original publicado en 1874)
- Briñol, P., & Petty, R. E. (2015). Elaboration and validation processes: Implications for media attitude change. *Media Psychology*, 18(3), 267-291.
- Brisson, J., & Markovits, H. (2020). Reasoning strategies and semantic memory effects in deductive reasoning. *Memory & Cognition*, 48(6), 920-930.

- Brittain, C. (2006). Philo of Larissa. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2006 ed.). Retrieved from <https://plato.stanford.edu/archives/win2006/entries/philo-larissa/>
- Brown, D. E. (1991). *Human universals*. New York: McGraw-Hill.
- Budzyńska, K., & Kacprzak, M. (2008). A logic for reasoning about persuasion. *Fundamenta Informaticae*, 85(1-4), 51-65.
- Burdett, E. R., Lucas, A. J., Buchsbaum, D., McGuigan, N., Wood, L. A., & Whiten, A. (2016). Do children copy an expert or a majority? Examining selective learning in instrumental and normative contexts. *PloS one*, 11(10), e0164698.
- Burge, T. (1986). Individualism and psychology. *The Philosophical Review*, 95(1), 3-45.
- Burge, T. (1993). Content preservation. *The philosophical review*, 102(4), 457-488.
- Burge, T. (2003). Perceptual entitlement. *Philosophy and phenomenological research*, 67(3), 503-548.
- Burge, T. (2010). *Origins of Objectivity*. Oxford University Press, USA.
- Burnyeat, F. M. (1994). Enthymeme: Aristotle on the logic of persuasion. En D. J. Furley & A. Nehamas (Eds.), *Aristotle's Rhetoric: Philosophical essays. Proceedings of the twelfth Symposium Aristotelicum* (pp. 3-55). Princeton, NJ: Princeton University Press.
- Burnyeat, M. F. (1996). Zeno's challenge. In A. O. Rorty (Ed.), *Essays on Aristotle's rhetoric* (pp. 88-106). Berkeley: University of California Press.
- Burris, S. (2018). George Boole. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition). Recuperado de <https://plato.stanford.edu/archives/sum2018/entries/boole/>
- Byrne, R. M. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31(1), 61-83.
- Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, 40(3), 347-373.
- Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development*, 70(2), 381-395.
- Calvo, M. G., Eysenck, M. W., & Castillo, M. D. (1997). Interpretation bias in test anxiety: The time course of predictive inferences. *Cognition & Emotion*, 11(1), 43-64.
- Camp, E. (2007). Thinking with maps. *Philosophical Perspectives*, 21, 145-182.

- Candel Sanmartín, M. (1982). Introducción, traducción y notas. En Aristóteles, *Tratados de lógica (Organon): I, Categorías, Tópicos, Refutaciones sofísticas*. Madrid: Editorial Gredos.
- Casson, R. (2002). Antropología cognitiva. En R. A. Wilson & F. C. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (pp. 190-192). Madrid: Editorial Síntesis.
- Castañeda, L. E. G., & Knauff, M. (2021). 5. 4 Defeasible Reasoning and Belief Revision in Psychology. *The handbook of rationality*, 347
- Cela-Conde, C. J., & Marty, G. (1998). Introducción. El cerebro y el órgano del lenguaje. En N. Chomsky, *Una aproximación naturalista a la mente y al lenguaje* (pp. 11-67, A. Gomila Benejam, Trad.). Editorial Prensa Ibérica S.A.
- Chan, D., & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition*, 53(3), 217-238.
- Chater, N., & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. *Cognition*, 34(1), 93–107. [https://doi.org/10.1016/0010-0277\(90\)90033-G](https://doi.org/10.1016/0010-0277(90)90033-G)
- Chater, N., & Oaksford, M. (2000). The rational analysis of mind and behavior. *Synthese*, 122, 93-131.
- Chater, N., & Oaksford, M. (2004). Rationality, rational analysis, and human reasoning. In K. Manktelow & M. C. Chung (Eds.), *Psychology of reasoning: Theoretical and historical perspectives* (pp. 53-84). Psychology Press.
- Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Science*, 10, 287–291.
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17(4), 391-416.
- Cherniak, C. (1984). Computational Complexity and the Universal Acceptance of Logic. *Journal of Philosophy*, 81, 739-758.
- Chirkov, V. I., & Ryan, R. M. (2001). Parent and teacher autonomy-support in Russian and US adolescents: Common effects on well-being and academic motivation. *Journal of cross-cultural psychology*, 32(5), 618-635.
- Chisholm, R. M. (1957). *Perceiving*. Ithaca: Cornell University Press.
- Chomsky, N. (1957). *Syntactic structures*. Mouton.

- Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language*, 35(1), 26-58.
- Church, A. (1936). A note on the Entscheidungsproblem. *The journal of symbolic logic*, 1(1), 40-41.
- Clark, H. H. (1985). Language use and language users. *Handbook of social psychology*.
- Clark, H. H. (1996). *Using language*. Cambridge University Press
- Clegg, J. M., Kurkul, K. E., & Corriveau, K. H. (2019). Trust me, I'm a competent expert: developmental differences in children's use of an expert's explanation quality to infer trustworthiness. *Journal of Experimental Child Psychology*, 188, 104670.
- Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, 19(4), 360-379.
- Clifford, A., Holmes, A., Davies, I. R., & Franklin, A. (2010). Color categories affect pre-attentive color perception. *Biological psychology*, 85(2), 275-282.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
<https://doi.org/10.1037/0033-2909.112.1.155>
- Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. MIT press.
- Coletti, G., & Scozzafava, R. (2002). *Probabilistic logic in a coherent setting* (Vol. 15). Springer Science & Business Media.
- Collingwood, R. G. (1946). *The idea of history*. Oxford: Clarendon Press.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428.
- Collins, P. J., & Hahn, U. (2018). Fallacies of argumentation. In L. J. Ball & V. A. Thompson (Eds.), *The Routledge international handbook of thinking and reasoning* (pp. 88–108). Routledge/Taylor & Francis Group.
- Collins, P. J., Hahn, U., Von Gerber, Y., & Olsson, E. J. (2018). The bi-directional relationship between source characteristics and message content. *Frontiers in Psychology*, 9, 317842.
- Connors, M. H., & Halligan, P. W. (2015). A cognitive account of belief: a tentative road map. *Frontiers in psychology*, 5, 123666.
- Conway, A. R., & Engle, R. W. (1994). Working memory and retrieval: a resource-dependent inhibition model. *Journal of Experimental Psychology: General*, 123(4), 354.

- Conway, M. A. (1990). *Autobiographical memory: An introduction*. Buckingham, UK: Buckingham University Press.
- Conway, M. A. (2002). Sensory-perceptual episodic memory and its context: autobiographical memory. En A. Baddeley, J. P. Aggleton, & M. A. Conway (Eds.), *Episodic memory: New directions in research* (pp. 53-70). Oxford: Oxford University Press.
- Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*, 51(1-2), 26-49.
- Cook, S. A. (1971). The complexity of theorem proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing* (pp. 151-158).
- Copi, I. M., Cohen, C., & Rodych, V. (2019). *Introduction to logic* (15th ed.). Abingdon: Routledge.
- Corner, A., & Hahn, U. (2007). Evaluating the meta-slope: Is there a slippery slope argument against slippery slope arguments? *Argumentation*, 21(3), 349-359. <https://doi.org/10.1007/s10503-007-9047-x>
- Corner, A., & Hahn, U. (2013). Normative theories of argumentation: Are some norms better than others?. *Synthese*, 190, 3579-3610.
- Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, 111(2), 291-303.
- Couillard, N. L., & Woodward, A. L. (1999). Children's comprehension of deceptive points. *British Journal of Developmental Psychology*, 17(4), 515-521.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford Psychology Series, No. 26. New York, NY: Oxford University Press.
- Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.
- Cullison, A. (2010). What are Seemings? *Ratio*, 23(3), 260-274. doi:10.1111/j.1467-9329.2010.00466.x
- Cummins, D. D. (1992). Role of analogical reasoning in the induction of problem categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1103.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & cognition*, 23(5), 646-658.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & cognition*, 19(3), 274-282.

- Davidson, D. (1986). A Coherence Theory of Knowledge and Truth. En E. LePore (Ed.), *Truth and Interpretation* (pp. 307–319). Oxford: Blackwell.
- De Baere, B. (2003). Thought experiments, rhetoric, and possible worlds. *Philosophica*, 72(2), 105-130.
- De Echandía, G. R. (1995). Introducción. En Aristóteles, *Física* (pp. 7-68). Madrid: Gredos.
- De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7, 1-68; translated (1980) Foresight: Its logical laws, its subjective sources. In H. E. Kyburg, Jr. & H. E. Smokler (Eds.), *Studies in subjective probability*. New York, NY: Robert E. Krieger Publishing Company.
- De Moivre, A. (1733). *Approximatio ad summam terminorum binomii (a+ b) n in seriem expansi*.
- De Morgan, A. (1847). *Formal Logic: or, The Calculus of Inference, Necessary and Probable*. London: Taylor and Walton
- De Neys, W. (2010). Counterexample retrieval and inhibition during conditional reasoning: Direct evidence from memory probing. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 197-206). Oxford University Press.
- De Neys, W., Schaeken, W., & D'ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & cognition*, 30(6), 908-920.
- De Neys, W., Schaeken, W., & D'Ydewalle, G. (2003a). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & cognition*, 31(4), 581-595.
- De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology*, 15(2), 161-176.
- De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005a). Working memory and counterexample retrieval for causal conditionals. *Thinking & Reasoning*, 11(2), 123-150.
- De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005b). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, 11(4), 349-381.
- de Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 373.

- Deeb, S. S. (2006). Genetics of variation in human color vision and the retinal cone mosaic. *Current opinion in genetics & development*, 16(3), 301-307.
- Demetriou, A., Kui, Z. X., Spanoudis, G., Christou, C., Kyriakides, L., & Platsidou, M. (2005). The architecture, dynamics, and development of mental processing: Greek, Chinese, or Universal?. *Intelligence*, 33(2), 109-141.
- Demey, L., Kooi, B., & Sack, J. (2019). Logic and probability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Retrieved from <https://plato.stanford.edu/archives/sum2019/entries/logic-probability/>
- Dennett, D. C. (2017). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press. (Originalmente publicado en 1978)
- Devitt, M., & Sterelny, K. (1987). *Language and reality: An introduction to the philosophy of language*. Blackwell.
- Dewey, J. (1933). *How we think*. Boston: D.C. Heath & Co. (Obra original publicada en 1910).
- Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002). The relative contribution of content and context factors on the interpretation of conditionals. *Experimental Psychology*, 49(3), 181-195
- Dieussaert, K., Schaeken, W., De Neys, W., & d'Ydewalle, G. (2000). Initial belief state as a predictor of belief revision. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*.
- Díez, J. A. (2002). *Iniciación a la Lógica*. Ariel.
- Díez, J. A., & Moulines, C. U. (1999). *Fundamentos de filosofía de la ciencia*. Barcelona: Editorial Ariel.
- Dirks, K. T. (1999). The effects of interpersonal trust on work group performance. *Journal of applied psychology*, 84(3), 445.
- Douven, I., & Verbrugge, S. (2010). The Adams family. *Cognition*, 117(3), 302-318.
- Douven, I., & Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science*, 37(4), 711-730.
- Dreyfus, H. L. (1978). Cognitivism vs. Hermeneutics. *Behavioral and Brain Sciences*, 1(2), 233-234.
- Dubois, D., & Prade, H. (1991). Epistemic entrenchment and possibilistic logic. *Artificial Intelligence*, 50(2), 223-239.

- Dung, P. M. (1995). An argumentation-theoretic foundation for logic programming. *The Journal of logic programming*, 22(2), 151-177.
- Dutilh Novaes, C. (2022). Argument and argumentation. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition). Retrieved from <https://plato.stanford.edu/archives/fall2022/entries/argument/>
- Dwyer, C. P., Hogan, M. J., & Stewart, I. (2013). An examination of the effects of argument mapping on students' memory and comprehension performance. *Thinking Skills and Creativity*, 8, 11-24.
- Dyck, E. (2002). Topos and enthymeme. *Rhetorica*, 20(2), 105-117.
- Echevarría, J. (1999). *Introducción a la metodología de la ciencia. Filosofía de la ciencia en el siglo XX*. Colección Teorema. Editorial Cátedra. Madrid, España.
- Eibl-Eibesfeldt, I. (1989). *Human ethology*. New York: Aldine de Gruyter. (Reeditado en 2017 por Routledge).
- Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: is it spatial memory or a memory space?. *Neuron*, 23(2), 209-226.
- Elio, R., & Pelletier, F. J. (1997). Belief change as propositional update. *Cognitive Science*, 21(4), 419-460.
- Engel, P. (2018). The doxastic zoo. In E. Picardi (Ed.), *Eva Picardi on Language, Analysis and History* (pp. 297-316).
- Eriksson, L., & Hájek, A. (2007). What are degrees of belief? *Studia Logica*, 86(2), 183-213.
- Estabrook, E. M., & Yoder, J. I. (1998). Plant-plant communications: Rhizosphere signaling between parasitic angiosperms and their hosts. *Plant Physiology*, 116(1), 1-7.
- Evans, J. S. B. T. (2002). Logic and human reasoning: an assessment of the deduction paradigm. *Psychological bulletin*, 128(6), 978.
- Evans, J. S. B. T. (2005). The social and communicative function of conditional statements. *Mind & Society*, 4, 97-113.
- Evans, J. S. B. T., & Over, D. E. (1996). *Essays in cognitive psychology. Rationality and reasoning. Psychology/Erlbaum (Uk) Taylor & Fr.*
- Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

- Evans, J. S. B. T., & Over, D. E. (2004). *If: Supposition, pragmatics, and dual processes*. Oxford University Press, USA.
- Evans, J. S. B. T., & Over, D. E. (2013). Reasoning to and from belief: Deduction and induction are still distinct. *Thinking & Reasoning*, 19(3-4), 267-283.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 223-241.
- Evans, J. S. B. T., & Thompson, V. A. (2004). Informal reasoning: Theory and method. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 58(2), 69-74.
- Evans, J. S. B. T., Handley, S. J., Harper, C. N., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1495.
- Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal of Experimental Psychology*, 63(5), 892-909.
- Evans, J. S. B. T., Neilens, H., Handley, S. J., & Over, D. E. (2008). When can we say 'if'? *Cognition*, 108(1), 100-116.
- Evans, J. S. B. T., Thompson, V. A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, 6, 398.
- Ewing, A. C. (1934). *Idealism: A Critical Survey*. London: Methuen.
- Fairley, N., Manktelow, K., & Over, D. (1999). Necessity, sufficiency, and perspective effects in causal conditional reasoning. *The Quarterly Journal of Experimental Psychology Section A*, 52(3), 771-790.
- Farmer, G. D., Warren, P. A., & Hahn, U. (2017). Who "believes" in the Gambler's Fallacy and why?. *Journal of Experimental Psychology: General*, 146(1), 63.
- Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1327.
- Ferrando, P. J. (2003a). A Kernel density analysis of continuous typical-response scales. *Educational and psychological measurement*, 63(5), 809-824.
- Ferrando, P. J. (2003b). Analyzing retest increases in reliability: A covariance structure modeling approach. *Structural Equation Modeling*, 10(2), 222-237.

- Ferrater Mora, J. (2005). *Diccionario de filosofía* (J.-M. Terricabras, Revisión actualizada). Barcelona: RBA Coleccionables.
- Festinger, L. (1957). Social comparison theory. In *Selective Exposure Theory* (Vol. 16, pp. 401-420).
- Festinger, L. (1962). Cognitive dissonance. *Scientific American*, 207(4), 93-106.
- Figa Faura, L. (1993). *Lógica, tópica y razonamiento jurídico*. Real Academia de Jurisprudencia y Legislación.
- Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological science*, 25(1), 38-46.
- Firestone, C., & Scholl, B. J. (2015). Can you experience 'top-down' effects on perception?: The case of race categories and perceived lightness. *Psychonomic bulletin & review*, 22, 694-700.
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and brain sciences*, 39, e229.
- Fishbein, M. (1963). An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human relations*, 16(3), 233-239.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and language*, 27(3), 235-244.
- Fodor, J. A. (1975). *The language of thought* (Vol. 5). Cambridge, MA: Harvard University Press.
- Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind* (Vol. 2). MIT Press.
- Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford University Press, USA.
- Fodor, J. A., & Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's "ecological approach". *Cognition*, 9(2), 139-196.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2), 3-71.

- Fogelin, R. J. (1991). *Understanding arguments: An introduction to informal logic*. Harcourt Brace Jovanovich.
- Foley, R. (1994). Egoism in epistemology. In F. F. Schmitt (Ed.), *Socializing epistemology: The social dimensions of knowledge* (pp. 53-73). Lanham, MD: Rowman & Littlefield.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior research methods, instruments, & computers*, 35, 116-124.
- Franklin, J. (2001). *The science of conjecture: Evidence and probability before Pascal*. Johns Hopkins University Press.
- Freeman, J. B. (1991). *Dialectics and the macrostructure of arguments: A theory of argument structure* (Vol. 10). Walter de Gruyter. (Reimpreso en 2011).
- Frege, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle a. S.: Louis Nebert. Traducido como *Concept Script, a formal language of pure thought modelled upon that of arithmetic* por S. Bauer-Mengelberg en J. van Heijenoort (Ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931* (pp. 1-82). Cambridge, MA: Harvard University Press, 1967.
- Frege, G. (1953). *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl*. Breslau: W. Koebner. Traducido como *The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number* (2ª ed. revisada) por J. L. Austin. Oxford: Blackwell.
- Friedman, M. (1974). Explanation and scientific understanding. *the Journal of Philosophy*, 71(1), 5-19.
- Fumerton, R. (2006). Epistemic internalism, philosophical assurance and the skeptical predicament. In T. M. Crisp, M. Davidson, & D. Vander Laan (Eds.), *Knowledge and reality* (Vol. 103). Springer Science & Business Media. https://doi.org/10.1007/1-4020-4733-9_8
- Fumerton, R. (2016). The prospects for traditional internalism. En B. Coppenger & M. Bergmann (Eds.), *Intellectual Assurance: Essays on Traditional Epistemic Internalism* (pp. 239-258). Oxford, Reino Unido: Oxford University Press. doi:10.1093/acprof/9780198719632.003.0013
- Gabaudan, F. C. (1994). Formas y funciones del entimema en la oratoria ática. *Cuadernos de Filología Clásica: Estudios Griegos e Indoeuropeos*, (4), 205-226.
- Gaifman, H. (1988). A theory of higher order probabilities. In B. Skyrms & W. L. Harper (Eds.), *Causation, chance, and credence* (pp. 191-219). Kluwer Academic Publishers.

- Galesic, M., Olsson, H., Dalege, J., Van Der Does, T., & Stein, D. L. (2021). Integrating social and cognitive aspects of belief dynamics: towards a unifying framework. *Journal of the Royal Society Interface*, 18(176), 20200857.
- Gambetta, D. (Ed.). (1988). *Trust: Making and breaking cooperative relations*. Oxford: Basil Blackwell.
- Garavelli, B. M. (1997). *Manuale di retorica* (p. 253). Milan: Bompiani.
- García Secades, M. (2002). Antecedentes de la concepción subjetivista de la probabilidad. En A. H. E. P. E. (Ed.), *Historia de la probabilidad y de la estadística* (pp. 119-132). Madrid: Editorial AC.
- García Secades, M. (2005). Probabilidad filosófica y moral: Las primeras contribuciones a la idea de lo probable. *Estadística Española*, 47(160), 567-600.
- Gärdenfors, P., & Stephens, A. (2018). Induction and knowledge-what. *European Journal for Philosophy of Science*, 8(3), 471-491.
- Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability* (Vol. 174). San Francisco: freeman.
- Garnham, A., & Oakhill, J. (1996). *Manual de psicología del pensamiento: pensar y razonar*. Ediciones Paidós Ibérica.
- Garrod, S., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(2), 250.
- Gattis, M. (Ed.). (2003). *Spatial schemas and abstract thought*. Cambridge, MA: MIT Press.
- Gaviria, C., & Jiménez-Leal, W. (2014). Coherence and argument structure: An empirical comparison between plausible reasoning and the Bayesian approach to argumentation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).
- Gaviria, C., & Jiménez-Leal, W. (2014). Encuentros cercanos con argumentos del “tercer tipo”: razonamiento plausible y probabilidad subjetiva como modelos de evaluación de argumentos. *Crítica (México, DF)*, 46(137), 85-112.
- Gawronski, B., & Strack, F. (2012). Cognitive consistency as a basic principle of social information processing. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 1-16). The Guilford Press.

- Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & cognition*, 35(8), 2060-2074.
- Geiger, S. M., & Oberauer, K. (2010). Towards a reconciliation of mental model theory and probabilistic theories of conditionals. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 289-307). Oxford University Press.
- Geis, M. L., & Zwicky, A. M. (1971). On invited inferences. *Linguistic inquiry*, 2(4), 561-566.
- Gentner, D. (2001). Spatial metaphors in temporal reasoning. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 203-222). Cambridge, MA: MIT Press.
<https://doi.org/10.7551/mitpress/6392.003.0012>
- Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy: Introduction. *American psychologist*, 52(1), 32.
- Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space→time metaphors. *Language and cognitive processes*, 17(5), 537-565.
- Gentzen, G. (1934). Untersuchungen über das logische Schließen. I & II. *Mathematische Zeitschrift*, 39, 176–210, 405–431. doi:10.1007/BF01201353, doi:10.1007/BF01201363.
- George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology*, 86(1), 93-111.
- George, C. (1997). Reasoning from uncertain premises. *Thinking & Reasoning*, 3(3), 161-189.
- Gibbs, R. W. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge, UK: Cambridge University Press.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4), 650-669.
- Gilbert, D. T. (1991). How mental systems believe. *American psychologist*, 46(2), 107.
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of personality and social psychology*, 65(2), 221.
- Ginsberg, M. L. (Ed.). (1987). *Readings in nonmonotonic reasoning*. Morgan Kaufmann.
- Girón, F. J. (1994). Historia del cálculo de probabilidades: de Pascal a Laplace. *Historia de la Matemática*, 113-133.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20(1), 1-19.

- Glenberg, A. M. (1999). Why mental models must be embodied. In B. Kokinov (Ed.), *Advances in psychology* (Vol. 128, pp. 77-90). North-Holland.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.
- Glenberg, A. M., Jaworski, B., Rischal, M., & Levin, J. R. (2007). What brains are for: Action, meaning, and reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 77-102). Erlbaum.
- Glock, H. J. (2000). Animals, thoughts and concepts. *Synthese*, 123, 35-64.
- Glock, H.-J. (2010). Can animals judge? *Dialectica*, 64(1), 11-33.
- Godden, D., & Walton, D. (2007). Advances in the theory of argumentation schemes and critical questions. *Informal Logic*, 27(3), 267-292.
- Gödel, K. (1986). *Kurt Gödel: Collected works: Volume I: Publications 1929-1936* (Vol. 1). Oxford University Press, USA.
- Gohlke, P. (1943). *Die Entstehung der aristotelischen Ethik, Politik, Rhetorik*. Viena: Akademie der Wissenschaften.
- Goldman, A. I. (2001). Experts: Which ones should you trust?. *Philosophy and phenomenological research*, 63(1), 85-110.
- Goldman, S. R. (2018). Discourse of learning and the learning of discourse. *Discourse Processes*, 55(5-6), 434-453.
- Goldstone, R. (2002). Semejanza. En R. A. Wilson & F. C. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (pp. 1168-1169). Madrid: Editorial Síntesis.
- Gómez Camacho, F. S. J. (2002). Probabilismo y toma de decisiones en la Escolástica española. En A. H. E. P. E. (Ed.), *Historia de la probabilidad y de la estadística* (pp. 81-102). Madrid: Editorial AC.
- Gómez Villegas, M. A. (2001). EL «Ensayo encaminado a resolver un problema en la doctrina del azar». *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales (España)*, 95(1-2), 81-85.
- González Bedoya, J. (1989). Perelman y la retórica filosófica. *Prólogo a la edición española de C. PERELMAN y L. OLBRECHTSTYTECA, Tratado de la argumentación*. Madrid: Gredos.

- González Bedoya, J. (2006). Prólogo. En C. Perelman & L. Olbrechts-Tyteca, *Tratado de la argumentación: La nueva retórica* (pp. 7-26). Gredos. (Obra original publicada en 1958).
- Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 23-32). New York, NY: Bobbs-Merrill.
- Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive science*, 32(1), 108-154.
- Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory formation system. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 299–323). The MIT Press.
- Gottlieb, P. (2023). Aristotle on non-contradiction. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition). Retrieved from <https://plato.stanford.edu/archives/win2023/entries/aristotle-noncontradiction/>
- Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 16, pp. 59-109). Academic Press.
- Graesser, A. C., Gernsbacher, M. A., & Goldman, S. R. (2000). Cognición. En T. A. van Dijk (Comp.), *El discurso como estructura y proceso: Estudios sobre el discurso I: Una introducción multidisciplinaria* (pp. 417-452). Barcelona: Editorial Gedisa. (Obra original publicada en 1997 por Sage Publications of London, Thousand Oaks, and New Delhi).
- Graesser, A. C., Gordon, S. E., & Sawyer, J. D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer+ tag hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 18(3), 319-332.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48(1), 163-189.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review*, 101(3), 371.
- Grasso, F. (2003). *Characterising rhetorical argumentation* (Doctoral dissertation, Heriot-Watt University).
- Green, N. L. (2016). Implementing argumentation schemes as logic programs. In *Proceedings of the CMNA@IJCAI* (pp. 1-7).
- Grice, H. P. (1969). Utterer's meaning and intention. *The Philosophical Review*, 78(2), 147-177.

- Grice, H. P. (1975). Logic and conversation. En P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 41-58). Brill.
- Grice, H. P. (1989). Logic and conversation. En *Studies in the Way of Words* (pp. 22-40). Harvard University Press.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Grilli, M. D., & Verfaellie, M. (2014). Personal semantic memory: insights from neuropsychological research on amnesia. *Neuropsychologia*, 61, 56-64.
- Grimaldi, W. M. A. (1972). *Studies in the philosophy of Aristotle's rhetoric*. Hermes: Zeitschrift für Klassische Philologie, Einzelschriften, 25. Wiesbaden.
- Grimaldi, W. M. A. (1980). *Aristotle's Rhetoric: A commentary*. Nueva York: Fordham University Press.
- Groarke, L. (2017). Informal logic. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition). Recuperado de <https://plato.stanford.edu/archives/win2017/entries/logic-informal/>
- Groarke, L. F. (2022). Is Aristotle the Forefather of Informal Logic?. *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie*, 61(1), 139-159.
- Grosset, N., Barrouillet, P., & Markovits, H. (2005). Chronometric evidence for memory retrieval in causal conditional reasoning: The case of the association strength effect. *Memory & cognition*, 33, 734-741.
- Gutiérrez-García, A., & Calvo, M. G. (2017). Social anxiety and threat-related interpretation of dynamic facial expressions: sensitivity and response bias. *Personality and Individual Differences*, 107, 10-16.
- Hacking, I. (1995). *El surgimiento de la probabilidad: Un estudio filosófico de las ideas tempranas acerca de la probabilidad, la inducción y la inferencia estadística*. Madrid: Alianza Editorial.
- Haenni, R., Kohlas, J., & Lehmann, N. (2000). Probabilistic argumentation systems. In D. M. Gabbay & P. Smets (Eds.), *Handbook of defeasible reasoning and uncertainty management systems: Algorithms for uncertainty and defeasible reasoning* (Vol. 5, pp. 221-288). Dordrecht: Springer Netherlands.
- Hahn, U., & Hornikx, J. (2016). A normative framework for argument quality: Argumentation schemes with a Bayesian foundation. *Synthese*, 193, 1833-1873.

- Hahn, U., & Oaksford, M. (2006a). A Bayesian approach to informal argument fallacies. *Synthese*, 152, 207-236.
- Hahn, U., & Oaksford, M. (2006b). A normative theory of argument strength. *Informal Logic*, 26(1), 1-24.
- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3), 704–732. <https://doi.org/10.1037/0033-295X.114.3.704>
- Hahn, U., Bluhm, R., & Zenker, F. (2017). Causal argument. In M. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 475-494). Oxford, UK: Oxford University Press.
- Hahn, U., Harris, A. J., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, 29(4), 337-367.
- Hahn, U., Oaksford, M., & Harris, A. J. (2013). Testimony and argument: A Bayesian perspective. In F. Zenker (Ed.), *Bayesian argumentation: The practical side of probability* (pp. 15-38). Springer.
- Hájek, A. (2019). Interpretations of probability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition). Retrieved from <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>
- Hakim, M. A., Thontowi, H. B., Yuniarti, K. W., & Kim, U. (2012). The basis of children’s trust towards their parents in Java, ngemong: Indigenous psychological analysis. *International Journal of Research Studies in Psychology*, 1(2), 3-16.
- Halldorson, M., & Singer, M. (2002). Inference processes: Integrating relevant knowledge and text information. *Discourse Processes*, 34(2), 145-161.
- Hamblin, C. L. (1970). *Fallacies*. Advanced Reasoning Forum. (Reimpreso en 2022, marzo).
- Hanna, R. (2006). *Rationality and Logic*. Cambridge, MA: MIT Press.
- Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82(7), 335-349.
- Hardwig, J. (1991). The role of trust in knowledge. *The Journal of Philosophy*, 88(12), 693-720.
- Harkness, S., & Super, C. M. (Eds.). (1996). *Parents’ cultural belief systems: Their origins, expressions, and consequences*. Guilford Press.
- Harris, A. J., Hahn, U., Madsen, J. K., & Hsu, A. S. (2016). The appeal to expert opinion: Quantitative support for a Bayesian network approach. *Cognitive Science*, 40(6), 1496-1533.

- Harris, A. J., Hsu, A. S., & Madsen, J. K. (2012). Because Hitler did it! Quantitative tests of Bayesian argumentation using ad hominem. *Thinking & Reasoning*, 18(3), 311-343.
- Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), 1179-1187.
- Harris, R. J., & Monaco, G. E. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General*, 107(1), 1-22.
- Harris, S., Sheth, S. A., & Cohen, M. S. (2008). Functional neuroimaging of belief, disbelief, and uncertainty. *Annals of neurology*, 63(2), 141-147.
- Hart, H. L. A. (1949). The ascription of responsibility and rights. *Proceedings of the Aristotelian Society*, 49(1), 171-194.
- Hasan, A., & Fumerton, R. (2016). Foundationalist Theories of Epistemic Justification. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). Recuperado de <https://plato.stanford.edu/archives/win2016/entries/justep-foundational/>
- Hastings, A. C. (1962). *A Reformulation of the Modes of Reasoning in Argumentation*. Northwestern University.
- Hayes, B. K., & Heit, E. (2018). Inductive reasoning 2.0. *Wiley Interdisciplinary Reviews: Cognitive Science*, 9(3), e1459.
- Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley interdisciplinary reviews: Cognitive science*, 1(2), 278-292.
- Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, 7(4), 569-592.
- Heit, E. (2008). Models of inductive reasoning. *The Cambridge handbook of computational psychology*, 322-338.
- Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In *Psychology of learning and motivation* (Vol. 39, pp. 163-199). Academic Press.
- Heit, E., & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 805.
- Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(2), 411.

- Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(2), 411-422. <https://doi.org/10.1037/0278-7393.20.2.411>
- Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York, NY: Free Press.
- Hempel, C. G. (1967). The White Shoe: No Red Herring. *British Journal for the Philosophy of Science*, 18(3), 239-240.
- Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2), 135-175. <https://doi.org/10.1086/286983>
- Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, 69(4), 366-378.
- Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, 12(3), 123-135.
- Hintikka, K. J. J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Ithaca, NY: Cornell University Press.
- Hiser, J., & Koenigs, M. (2018). The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. *Biological psychiatry*, 83(8), 638-647.
- Hitchcock, C. (2002). Probability and chance. In *International Encyclopedia of the Social and Behavioral Sciences* (Vol. 18, pp. 12,089–12,095). London: Elsevier.
- Hobbs, J. R., & Agar, M. H. (1985). The coherence of incoherent discourse. *Journal of Language and Social Psychology*, 4(3-4), 213-232.
- Holland, D. C., & Quinn, N. (Eds.). (1987). *Cultural models in language and thought*. Cambridge: Cambridge University Press.
- Holyoak, K. J. (1985). The pragmatics of analogical transfer. In *Psychology of learning and motivation* (Vol. 19, pp. 59-87). Academic Press.
- Holyoak, K. J. (2002). Prefacio: Psicología. En R. A. Wilson & F. C. Keil (Eds.), *Enciclopedia MIT de Ciencias Cognitivas* (pp. 29-40). Madrid: Editorial Síntesis.
- Holyoak, K. J. (2012). Analogy and relational reasoning. *The Oxford handbook of thinking and reasoning*, 234-259.

- Hood, M. D. (1984). *Aristotle's enthymeme: Its theory and application to discourse*. University of Oregon.
- Horst, S. (2002). Teoría Computacional de la Mente. En H. Wilson & F. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (pp. 1264-1266). Madrid: Editorial Síntesis.
- Horty, J. F., Thomason, R. H., & Touretzky, D. S. (1990). A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial intelligence*, 42(2-3), 311-348.
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15(4), 635-650.
- Hovy, E., & Maier, E. (1995). Parsimonious or profligate: How many and which discourse relations. *Unpublished manuscript*.
- Huby, P. (2016). *Is Aristotle the Forefather of Informal Logic?* Dialogue: Canadian Philosophical Review, 55(1), 7-20. <https://doi.org/10.1017/S0012217316000123>
- Huemer, M. (1997). Probability and coherence justification. *Southern Journal of Philosophy*, 35, 463-472.
- Huemer, M. (2001). *Skepticism and the Veil of Perception*. Lanham, MD: Rowman & Littlefield.
- Huemer, M. (2007). Compassionate Phenomenal Conservatism. *Philosophy and Phenomenological Research*, 74(1), 30-55. doi:10.1111/j.1933-1592.2007.00002.x
- Huemer, M. (2011). Does Probability Theory Refute Coherentism? *Journal of Philosophy*, 108(1), 35-54.
- Huemer, M. (2013). Phenomenal conservatism Über Alles. In C. Tucker (Ed.), *Seemings and justification: New essays on dogmatism and phenomenal conservatism* (pp. 328-350). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199899494.003.0015>
- Huygens, C. (1980). *De ratiociniis in ludo aleae*. Ex officina J. Elsevirii.
- Ierodiakonou, K. (2020). Theophrastus. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition). Retrieved from <https://plato.stanford.edu/archives/win2020/entries/theophrastus/>
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 22). Psychology Press.
- Iranzo, V. (2012). Justificación. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 308-312). Trotta.

- Isberner, M. B., & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica*, 142(1), 15-22.
- Isberner, M. B., & Richter, T. (2014a). Comprehension and validation: Separable stages of information processing? A case for epistemic monitoring in language comprehension.
- Isberner, M. B., & Richter, T. (2014b). Does validation during language comprehension depend on an evaluative mindset?. *Discourse Processes*, 51(1-2), 7-25.
- Iyer, R. S., & Krishnan, V. R. (2002). Choice of influence strategies: Role of need for power, need for affiliation, and inhibition. *Prabandhan*, 1(3), 28-34.
- Jakobson, R. (1959). Sign and system of Language. *Verbal art, verbal sign, verbal time*, 28-33.
- Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350-377). MIT Press.
- Jaynes, E. T. (1968). Prior probabilities. *Institute of Electrical and Electronic Engineers Transactions on Systems Science and Cybernetics*, SSC-4, 227-241.
- Jeffrey, R. C. (1965). Ethics and the logic of decision. *The Journal of Philosophy*, 62(19), 528-539.
- Johnson-Laird, P. N. (1975). Models of deduction. En R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.
- Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition*, 50(1-3), 189-209.
- Johnson-Laird, P. N. (2001). Mental models. In R. A. Wilson & F. C. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 491-493). Cambridge, MA: MIT Press.
- Johnson-Laird, P. N. (2002). Modelos mentales. En R. A. Wilson & F. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (versión española, pp. 845-846). Madrid: Editorial Síntesis.
- Johnson-Laird, P. N. (2008). Mental models and deductive reasoning. *Reasoning: Studies in Human Inference and its Foundations*, 206-222.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N., & Lee, N. Y. (2006). Are there cross-cultural differences in reasoning?. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 28, No. 28).

- Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1(2), 134-148.
- Johnson-Laird, P. N., Gawronski, B., & Strack, F. (2012). Mental models and consistency. *Cognitive consistency: A fundamental principle in social cognition*, 225-243.
- Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological review*, 111(3), 640.
- Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19(4), 201-214.
- Johnstone, C. L. (1980). An Aristotelian trilogy: Ethics, rhetoric, politics, and the search for moral truth. *Philosophy and Rhetoric*, 13(1), 1-24.
- Jonauskaitė, D., Dael, N., Parraga, C. A., Chèvre, L., García Sánchez, A., & Mohr, C. (2020). Stripping# The Dress: the importance of contextual information on inter-individual differences in colour perception. *Psychological Research*, 84, 851-865.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. In *Advances in experimental social psychology* (Vol. 2, pp. 219-266). Academic Press.
- Jøsang, A. (2002). The consensus operator for combining beliefs. *Artificial Intelligence*, 141(1-2), 157-170.
- Kahane, H. (1971). *Logic and contemporary rhetoric: The use of reason in everyday life*. Belmont, CA: Wadsworth Publishing Company.
- Kalpokus, D. (2014). The experience not well lost. *Contemporary Pragmatism*, 11(1), 43-56. <https://doi.org/10.1163/18758185-90000275>
- Kamalski, J., Lentz, L., Sanders, T., & Zwaan, R. A. (2008). The forewarning effect of coherence markers in persuasive discourse: Evidence from persuasion and processing. *Discourse Processes*, 45(6), 545-579.
- Kant, I. (1787). *Crítica de la razón pura* (P. Ribas, Introducción, traducción, notas e índices, Edición 2005). Barcelona: RBA Colecciones.
- Karban, R. (2015). Plant sensing and communication. In *Plant sensing and communication*. University of Chicago Press.
- Katzav, J., & Reed, C. A. (2004). On argumentation schemes and the natural classification of arguments. *Argumentation*, 18, 239-259.

- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications.
- Keil, F. C. (2006). Explanation and understanding. *Annu. Rev. Psychol.*, 57, 227-254.
- Kelman, H. C., & Hovland, C. I. (1953). "Reinstatement" of the communicator in delayed measurement of opinion change. *The Journal of Abnormal and Social Psychology*, 48(3), 327.
- Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes*, 51(5-6), 374-397.
- Keynes, J. M. (1921). *A treatise on probability*. Mineola, N.Y.: Dover Publications.
- Khalid, B., Alikhani, M., Fellner, M., McMahan, B., & Stone, M. (2020). Discourse coherence, reference grounding and goal oriented dialogue. *arXiv preprint arXiv:2007.04428*.
- Khemlani, S. S., Byrne, R. M., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science*, 42(6), 1887-1924.
- Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. N. (2012). The probabilities of unique events. *PLoS ONE*, 7(10), Article e45975. <https://doi.org/10.1371/journal.pone.0045975>
- Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. N. (2014). A mental model theory of set membership. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).
- Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. N. (2015). Naive probability: Model-based estimates of unique events. *Cognitive science*, 39(6), 1216-1258.
- Kienpointner, M. (1986). Topische Sequenzen in argumentativen Dialogen. *Zeitschrift für germanistische Linguistik*, 14(3), 321-355. <https://doi.org/10.1515/zfgl.1986.14.3.321>
- Kienpointner, M. (1992). Alltagslogik. *Struktur und Funktion von Argumentationsmustern*. Stuttgart-Bad Cannstatt: Frommann-Holzboog.
- Kienpointner, M. (1997). On the Art of Finding Arguments: What Ancient and Modern Masters of Invention Have to Tell Us About the "Ars Inveniendi". *Argumentation*, 11, 225-236.
- Kienpointner, M. (2017). Rhetoric and argumentation. En J. Flowerdew & J. E. Richardson (Eds.), *The Routledge handbook of critical discourse studies* (pp. 14-30). Routledge.
- Kim, J., & Kendeou, P. (2021). Knowledge transfer in the context of refutation texts. *Contemporary Educational Psychology*, 67, 102002.
- Kintsch, W. (1974). *The representation of meaning in memory*. Lawrence Erlbaum.

- Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, 95(2), 163.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, 85(5), 363.
- Kirkpatrick, L. A., & Ellis, B. J. (2006). What is the evolutionary significance of self-esteem? The adaptive functions of self-evaluative psychological mechanisms. In M. H. Kernis (Ed.), *Self-esteem issues and answers: A sourcebook of current perspectives* (pp. 334-339). Psychology Press.
- Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation* (Vol. 13, pp. 410-505). Minneapolis, MN: University of Minnesota Press.
- Klein, P. (1999). Human knowledge and the regress of reasons. *Philosophical Perspectives*, 13, 297-325.
- Klein, P. (2007). Human knowledge and the infinite progress of reasoning. *Philosophical Studies*, 134, 1-17.
- Knowles, E. S., & Linn, J. A. (2004). *Resistance and persuasion*. Psychology Press.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological science*, 15(10), 694-698.
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer.
- Koons, R. (2017). Defeasible reasoning. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition). Recuperado de <https://plato.stanford.edu/archives/win2017/entries/reasoning-defeasible/>
- Kopperschmidt, J. (1997). Rhetorik-ein inter (multi-, trans-) disziplinäres Forschungsprojekt. *Rhetorica*, 15(1), 81-106.
- Kosslyn, S. M. (1980). *Image and mind*. Harvard University Press.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. The MIT Press.
- Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, 8(3), 439-453.

- Krifka, M. (2002). Composicionalidad. En H. Wilson & F. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (pp. 329-331). Madrid: Editorial Síntesis.
- Kripke, S. A. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York: Plenum.
- Kruglanski, A. W., & Orehek, E. (2009). Toward a relativity theory of rationality. *Social Cognition*, 27(5), 639-660.
- Kruglanski, A. W., & Shteynberg, G. (2012). Cognitive consistency as means to an end: How subjective logic affords knowledge. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 1-15). Guilford Press.
- Krzyżanowska, K., Collins, P. J., & Hahn, U. (2017). Between a conditional's antecedent and its consequent: Discourse coherence vs. probabilistic relevance. *Cognition*, 164, 199-205.
- Kubose, T. T., Holyoak, K. J., & Hummel, J. E. (2002). The role of textual coherence in incremental analogical mapping. *Journal of memory and language*, 47(3), 407-435
- Kuhl, J., & Fuhrmann, A. (1998). Decomposing self-regulation and self-control: The Volitional Components Inventory. In J. Heckhausen & C. S. Dweck (Eds.), *Motivation and self-regulation across the life span* (pp. 15-49). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511527869.003>
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological review*, 119(3), 573.
- Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *Neuroimage*, 33(1), 343-361.
- La Corte, V., & Piolino, P. (2016). On the role of personal semantic memory and temporal distance in episodic future thinking: The TEDIFT model. *Frontiers in human neuroscience*, 10, 385
- Lackey, J. (2008). *Learning from words: Testimony as a source of knowledge*. Oxford: Oxford University Press.
- Lamond, G. (2016). Precedent and analogy in legal reasoning. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition). Retrieved from <https://plato.stanford.edu/archives/spr2016/entries/legal-reas-prec/>

- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2), 211.
- Landro, A. H., & González, M. L. (2013). Bernoulli, De Moivre, Bayes, Price y los fundamentos de la inferencia inductiva. *Cuadernos del CIMBAGE*, (15), 33-56.
- Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science*, 16(4), 622-638.
- Laplace, P. S. de. (1795). *Essai philosophique sur les probabilités*. Paris: V. Courcier.
- Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and philosophy*, 16(5), 437-493.
- Lassonde, K. A., Kendeou, P., & O'Brien, E. J. (2016). Refutation texts: Overcoming psychology misconceptions that are resistant to change. *Scholarship of Teaching and Learning in Psychology*, 2(1), 62.
- Lea, R. B. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1469-1482.
- Lea, R. B., Mulligan, E. J., & Walton, J. L. (2005). Accessing distant premise information: how memory feeds reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 387.
- Lea, R. B., O'Brien, D. P., Fisch, S. M., Noveck, I. A., & Braine, M. D. S. (1990). Predicting propositional logic inferences in text comprehension. *Journal of Memory and Language*, 29(3), 361-387.
- Lee, M., Kim, H., & Kim, M. (2014). The effects of Socratic questioning on critical thinking in web-based collaborative learning. *Education as Change*, 18(2), 285-302.
- Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (2003). Models of consistency. *Psychological Science*, 14(2), 131-137.
- Lehrer, K. (1990). *Theory of Knowledge* (1st ed.). Boulder: Westview Press.
- Lehrer, K. (2003). Coherence, circularity and consistency: Lehrer replies. En E. J. Olsson (Ed.), *The Epistemology of Keith Lehrer* (pp. 309–356). Dordrecht: Kluwer.
- León Sáez de Ybarra, J. L. (2014). Validez y extensiones del Modelo de la Probabilidad de Elaboración (ELM). Para una teoría persuasiva en el campo de la publicidad. *adComunica*, 169-188.

- Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, 85, 1030–1051.
- Lerner, M. J., & Simmons, C. H. (1966). Observer's reaction to the "innocent victim": Compassion or rejection? *Journal of Personality and Social Psychology*, 4, 203–210.
- Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation*. La Salle, IL: Open Court.
- Lewis, C. I. (1952). The given element in empirical knowledge. *The Philosophical Review*, 61, 168-175.
- Lewis, D. (1989) *Causal Explanation*, en David Lewis (ed.), *Philosophical Papers Vol. II*. Oxford University Press. pp. 214-240 (1986)
- Lewis, D. (1994). Reduction of mind. En S. Guttenplan (Ed.), *A companion to the philosophy of mind* (pp. 412-431). Oxford: Blackwell.
- Lightman, A. (1993). *Einstein's Dreams*. Nueva York: Warner. [Sueños de Einstein. Barcelona: Tusquets, 1993].
- Loftus, E. F. (1996). Memory distortion and false memory creation. *Journal of the American Academy of Psychiatry and the Law Online*, 24(3), 281-295.
- Loftus, E. F., & Palmer, J. E. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13, 585–589.
- Lombrozo, T. (2006). The structure and function of explanations. *Trends in cognitive sciences*, 10(10), 464-470.
- Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, 99(2), 167-204.
- Lombrozo, T., & Gwynne, N. Z. (2014). Explanation and inference: Mechanistic and functional explanations guide property generalization. *Frontiers in Human Neuroscience*, 8, 700.
- Lombrozo, T., & Rutstein, J. J. (2004). Simplicity in explanation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 26, No. 26).
- Long, A. A., & Sedley, D. N. (Eds. & Trans.). (1987). *The Hellenistic philosophers* (Vols. 1-2, chs. 68-70). Cambridge: Cambridge University Press.
- Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology*, 1(1), 65-82.

- López González, E. (1998). Tratamiento de la colinealidad en regresión múltiple. *Psicothema*, 10(2), 491-507.
- Lorenzen, P. (1955). *Einführung in die operative Logik und Mathematik*. Berlin: Springer.
- Lorenzen, P., & Lorenz, K. (1978). *Dialogische Logik*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Lupyan, G. (2017). Objective effects of knowledge on visual perception. *Journal of experimental psychology: human perception and performance*, 43(4), 794.
- Macagno, F., & Walton, D. (2014). Argumentation schemes and topical relations. In G. Gobber & A. Rocci (Eds.), *Language, reason and education* (pp. 185–216). Bern: Peter Lang.
- Macagno, F., & Walton, D. (2015). Classifying the patterns of natural arguments. *Philosophy & Rhetoric*, 48(1), 26-53.
- Macagno, F., Walton, D., & Reed, C. (2017). Argumentation schemes: History, classifications, and computational applications. En F. Macagno, D. Walton, & C. Reed (Eds.), *Argumentation schemes: History, classifications, and computational applications* (pp. 2493-2556).
- Madrid-Casado, C. M. (2012). *Laplace: La mecánica celeste*. Barcelona: RBA Coleccionables.
- Madrugá, J. A. G., & Johnson-Laird, P. N. (1988). Entrevista a Philip N. Johnson-Laird. *Cognitiva*, 1(3), 311-333.
- Madsen, J. K., Hahn, U., & Pilditch, T. D. (2020). The impact of partial source dependence on belief and reliability revision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(9), 1795.
- Magliano, J. P., Radvansky, G. A., & Copeland, D. E. (2007). Beyond language comprehension: Situation models as a form of autobiographical memory. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 379-391). Erlbaum.
- Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, 91(4), 615-629.
- Maier, J., & Richter, T. (2016). Effects of text-belief consistency and reading task on the strategic validation of multiple texts. *European Journal of Psychology of Education*, 31(4), 479-497.
- Maier, M., & Abdel Rahman, R. (2019). No matter how: Top-down effects of verbal and semantic category knowledge on early visual perception. *Cognitive, Affective, & Behavioral Neuroscience*, 19, 859-876.

- Maio, G. R., Hahn, U., Frost, J. M., Kuppens, T., Rehman, N., & Kamble, S. (2014). Social values as arguments: similar is convincing. *Frontiers in Psychology*, 5, 99627.
- Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and instruction*, 4(2), 91-115.
- Marcus, R. B. (1990). Some revisionary proposals about belief and believing. *Philosophy and Phenomenological Research*, 50(1), 132–153.
- Markovits, H. (1984). Awareness of the ‘possible’ as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology*, 75(3), 367-376.
- Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children’s vigilance towards deception. *Cognition*, 112(3), 367-380.
- Mastropasqua, T., Crupi, V., & Tentori, K. (2010). Broadening the study of inductive reasoning: Confirmation judgments with uncertain evidence. *Memory & cognition*, 38(7), 941-950.
- Mateos-Aparicio Morales, G. (2002). Historia de la probabilidad (desde sus orígenes a Laplace) y su relación con la historia de la teoría de la decisión. En A. H. E. P. E. (Ed.), *Historia de la probabilidad y de la estadística* (pp. 1-18). Madrid: Editorial AC.
- Matheson, H. E., & Barsalou, L. W. (2018). Embodiment and grounding in cognitive neuroscience. In J. T. Wixted (Ed.), *Stevens’ handbook of experimental psychology and cognitive neuroscience* (Vol. 3, pp. 1-27). Wiley.
- McCarthy, J. M., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. En B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4*. Edinburgh: Edinburgh University Press.
- McClelland, J. L. (1998). Connectionist models and Bayesian inference. En M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 21-53). Oxford University Press.
- McClelland, J. L., & Thompson, R. M. (2007). Using domain-general principles to explain children’s causal reasoning abilities. *Developmental Science*, 10(3), 333-356.
- McDonald, J., Samuels, M., & Rispoli, J. (1996). A hypothesis-assessment model of categorical argument strength. *Cognition*, 59(2), 199-217.
- McGuire, W. J., & Papageorgis, D. (1962). Effectiveness of forewarning in developing resistance to persuasion. *Public Opinion Quarterly*, 26(1), 24-34.
- McRae, K., & Jones, M. (2013). Semantic memory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 206-224). Oxford: Oxford University Press.

- Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. L. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, 10(3), 517-532.
- Mercier, H. (2009). *La théorie argumentative du raisonnement* (Doctoral dissertation, Ecole pratique des hautes études-EPHE PARIS).
- Mercier, H. (2011). On the universality of argumentative reasoning. *Journal of Cognition and Culture*, 11(1-2), 85-113.
- Mercier, H. (2013). Introduction: Recording and explaining cultural differences in argumentation. *Journal of Cognition and Culture*, 13(5), 409-417.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57-74.
- Mercier, H., Deguchi, M., Van der Henst, J. B., & Yama, H. (2016). The benefits of argumentation are cross-culturally robust: The case of Japan. *Thinking & Reasoning*, 22(1), 1-15.
- Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The sense of confidence during probabilistic learning: A normative account. *PLoS computational biology*, 11(6), e1004305.
- Miceli, M., & Castelfranchi, C. (2014). *Expectancy and emotion*. OUP Oxford.
- Miller, C. R. (2000). The Aristotelian topos: Hunting for novelty. In A. Gross & A. Walzer (Eds.), *Rereading Aristotle's rhetoric* (pp. 130-146). Southern Illinois University Press.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97.
- Minsky, M. (1974). A framework for representing knowledge. En P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York, NY: McGraw-Hill.
- Misztal, B. (2013). *Trust in modern societies: The search for the bases of social order*. John Wiley & Sons.
- Moretti, L., & Piazza, T. (2023). Transmission of justification and warrant. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition). Retrieved from <https://plato.stanford.edu/archives/sum2023/entries/transmission-justification-warrant/>
- Murphy, P. (2006). Coherentism. *Internet Encyclopedia of Philosophy*. Recuperado de <https://www.iep.utm.edu/coherenc/>.

- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26(2-3), 131-157. <https://doi.org/10.1080/01638539809545042>
- Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 876.
- Newell, A., & Simon, H. A. (1956). Current developments in complex information processing (p. 0038). Rand Corporation.
- Nichols, S., & Stich, S. P. (2000). A cognitive theory of pretence. *Cognition*, 74(2), 115-147.
- Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford University Press.
- Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19(12), 1213-1218.
- Nilsson, N. J. (1986). Probabilistic logic. *Artificial Intelligence*, 28(1), 71-87. [https://doi.org/10.1016/0004-3702\(86\)90043-1](https://doi.org/10.1016/0004-3702(86)90043-1)
- Nilsson, N. J. (1993). Probabilistic logic revisited. *Artificial Intelligence*, 59(1-2), 39-42. [https://doi.org/10.1016/0004-3702\(93\)90101-6](https://doi.org/10.1016/0004-3702(93)90101-6)
- Nisbett, R. (2004). *The geography of thought: How Asians and Westerners think differently... and why*. Simon and Schuster.
- Nitsch, F. J., & Kalenscher, T. (2021). Influence of memory processes on choice-consistency. *Royal Society Open Science*, 8(10), 200308.
- Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(1), 54-65.
- O'Brien, D. P., Roazzi, A., Athias, R., Dias, M. G., Brandão, M. C., & Brooks, P. J. (2003). The language of thought and the existence of a mental logic: Experimental investigations in the laboratory and in the field. *Cognitive Science*, 27(4), 611-630.
- Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Cognitive Science*, 6(1), 1-38.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological review*, 101(4), 608.
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

- Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, 32(1), 69-84.
- Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, 19(3-4), 346–379.
- Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual review of psychology*, 71, 305-330.
- Oaksford, M., & Chater, N. (Eds.). (1998). *Rational models of cognition*. Oxford University Press.
- Oaksford, M., & Hahn, U. (2007). Induction, deduction and argument strength in human reasoning and argumentation. *Inductive reasoning*, 269-301.
- Oaksford, M., & Hahn, U. (2012). Why are we convinced by the ad hominem argument?: Bayesian source reliability and pragma-dialectical discussion rules. In *Bayesian argumentation: The practical side of probability* (pp. 39-58). Dordrecht: Springer Netherlands.
- O'Brien, D. P., Braine, M. D., Connell, J. W., Noveck, I. A., Fisch, S. M., & Fun, E. (1989). Reasoning about conditional sentences: Development of understanding of cues to quantification. *Journal of Experimental Child Psychology*, 48(1), 90-113.
- O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(4), 777.
- O'Brien, E. J., & Cook, A. E. (2016). Coherence threshold and the continuity of processing: The RI-Val model of comprehension. *Discourse Processes*, 53(5-6), 326-338.
- O'Brien, E. J., Cook, A. E., & Guéraud, S. (2010). Accessibility of outdated information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(4), 979.
- Olbrechts-Tyteca, L., & Perelman, C. (1989). Tratado de la argumentación. La nueva retórica. Trad.: Julia Sevilla Muñoz. Madrid: Editorial Gredos SA.
- Olmos, P. (2012). Tópico. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 588-595). Trotta.
- Olsson, A., & Ochsner, K. N. (2008). The role of social cognition in emotion. *Trends in cognitive sciences*, 12(2), 65-71.
- Olsson, E. J. (2002). What is the Problem of Coherence and Truth? *The Journal of Philosophy*, 99, 246–272.

- Olsson, E. J. (2017). Coherentism. En *The Routledge Handbook of Philosophy of Memory* (pp. 310-322). Routledge.
- O'Shaughnessy, J., & O'Shaughnessy, N. (2004). *Persuasion in advertising*. London: Routledge.
- Osherson, D. N. (1975). Logic and models of logical thinking. En R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Osherson, D. N. (1975). *Logical abilities in children: III. Reasoning in adolescence: Deductive inference*. Lawrence Erlbaum.
- Osherson, D. N., Smith, E. E., & Shafir, E. B. (1986). Some origins of belief. *Cognition*.
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological review*, 97(2), 185.
- Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, 15(4), 431-438. <https://doi.org/10.1080/13546780903266188>
- Over, D. E., & Evans, J. S. B. (2003). The probability of conditionals: The psychological evidence. *Mind & language*, 18(4), 340-358.
- Over, D. E., Hadjichristidis, C., Evans, J. S. B., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive psychology*, 54(1), 62-97.
- Pagin, P. (2014). Pragmatic enrichment as coherence raising. *Philosophical Studies*, 168(1), 59-100.
- Paivio, A. (1971). Imagery and language. En S. J. Segal (Ed.), *Imagery* (pp. 7-32). Academic Press.
- Pandelaere, M. (2022). Pursuing affiliation through consumption. *Current Opinion in Psychology*, 46, 101330.
- Parry, W. T., & Hacker, E. A. (1991). *Aristotelian logic*. Suny Press.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann.
- Pearl, J., & Russell, S. (2000). Bayesian networks. En *MIT Encyclopedia of the Cognitive Sciences* (pp. 62-64). MIT Press. Recuperado de http://www.info2.uqam.ca/~lounis_h/dic938G-hiv2011/bayes/pearl-bbn2000.pdf
- Pechey, R., & Halligan, P. (2012). Using co-occurrence to evaluate belief coherence in a large non clinical sample. *PloS one*, 7(11), e48446.

- Perelman, C., & Olbrechts-Tyteca, L. (1958). *Tratado de la argumentación: La nueva retórica* (J. Sevilla Muñoz, Trad.). Madrid: Editorial Gredos. (Obra original publicada en 1958).
- Pettigrew, R. (2015). Epistemic utility arguments for epistemic norms. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition). Retrieved from <https://plato.stanford.edu/archives/win2023/entries/epistemic-utility/>
- Petty, R. E., & Cacioppo, J. T. (1979). Effects of forewarning of persuasive intent and involvement on cognitive responses and persuasion. *Personality and Social Psychology Bulletin*, 5(2), 173-176.
- Petty, R. E., & Cacioppo, J. T. (1984). Source factors and the elaboration likelihood model of persuasion. *Advances in Consumer Research*, 11(1), 668-672.
- Pfau, M., Szabo, A., Anderson, J., Morrill, J., Zubric, J., & H-Wan, H. H. (2001). The role and impact of affect in the process of resistance to persuasion. *Human communication research*, 27(2), 216-252.
- Pfeifer, N. (2013). The new psychology of reasoning: A mental probability logical perspective. *Thinking & Reasoning*, 19(3-4), 329-345.
- Pfeifer, N. (2014). Reasoning about uncertain conditionals. *Studia Logica*, 102, 849-866.
- Pfeifer, N., & Kleiter, G. D. (2006). Is human reasoning about nonmonotonic conditionals probabilistically coherent? In *Proceedings of the 7th Workshop on Uncertainty Processing* (pp. 138-150).
- Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability logic. *Journal of Applied Logic*, 7(2), 206-217.
- Pfeifer, N., & Kleiter, G. D. (2010). Uncertain deductive reasoning. In K. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The science of reason: A Festschrift for Jonathan St. BT Evans* (pp. 145-166). Psychology Press.
- Piaget, J. (1971). *Biology and knowledge*. Edinburgh: Edinburgh University Press.
- Piest, B. A., Isberner, M. B., & Richter, T. (2018). Don't believe everything you hear: Routine validation of audiovisual information in children and adults. *Memory & Cognition*, 46, 849-863.
- Pijnacker, J., Geurts, B., Van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, 48(10), 2940-2951.

- Politzer, G. (2007). Reasoning with conditionals. *Topoi*, 26(1), 79-95.
- Politzer, G. (2016). Deductive reasoning under uncertainty: A water tank analogy. *Erkenntnis*, 81(3), 479-506.
- Politzer, G., & Baratgin, J. (2016). Deductive schemas with uncertain premises using qualitative probability expressions. *Thinking & Reasoning*, 22(1), 78-98.
- Politzer, G., & Bonnefon, J. F. (2006). Two varieties of conditionals and two kinds of defeaters help reveal two fundamental types of reasoning. *Mind and Language*, 21(4), 484-503.
- Politzer, G., & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking & Reasoning*, 7(3), 217-234.
- Pollock, J. L. (1987). Defeasible reasoning. *Cognitive science*, 11(4), 481-518.
- Pollock, J. L. (1991). A theory of defeasible reasoning. *International Journal of Intelligent Systems*, 6(1), 33-54.
- Pollock, J. L. (1995). *Cognitive carpentry: A blueprint for how to build a person*. Cambridge, MA: MIT Press.
- Pollock, J. L. (1998). Perceiving and reasoning about a changing world. *Computational Intelligence*, 14(4), 498-562.
- Pollock, J. L. (2001). Defeasible reasoning with variable degrees of justification. *Artificial intelligence*, 133(1-2), 233-282.
- Pollock, J. L. (2008). Defeasible reasoning. En L. J. Rips & J. Adler (Eds.), *Reasoning: Studies in human inference and its foundations* (pp. 451-470). Cambridge: Cambridge University Press.
- Pollock, J. L., & Cruz, J. (1999). *Contemporary Theories of Knowledge* (Vol. 35). Rowman & Littlefield.
- Poston, T. (2008). Foundationalism. *Internet Encyclopedia of Philosophy*. Recuperado de <https://www.iep.utm.edu/found-ep/>.
- Prakken, H. (2010). On the nature of argument schemes. In C. Reed & C. W. Tindale (Eds.), *Dialectics, dialogue and argumentation: An examination of Douglas Walton's theories of reasoning and argument* (pp. 167-185).
- Pryor, J. (2000). The skeptic and the dogmatist. *Nous*, 34, 517-549.
- Pujadas Torres, L. M. (2002). *La ascensión y la caída de la teoría funcionalista de la mente*. Palma de Mallorca: Universitat de les Illes Balears.

- Putnam, H. (1975). The meaning of 'meaning'. En K. Gunderson (Ed.), *Minnesota Studies in the Philosophy of Science* (Vol. 7, pp. 131-193). University of Minnesota Press.
- Putnam, H. (1975). What is mathematical truth? En H. Putnam, *Mathematics, matter and method: Philosophical papers* (Vol. 1) (pp. 60-78). Cambridge, UK: Cambridge University Press. (Original work published 1975).
- Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80(1), 1-24.
- Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. MIT Press.
- Quick, B. L., Shen, L., & Dillard, J. P. (2013). Reactance theory and persuasion. *The SAGE handbook of persuasion: Developments in theory and practice*, 2, 167-183.
- Quine, W. V., & Ullian, J. S. (1970). *The web of belief*. New York: Random House.
- Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition*, 68(3), B93-B101.
- Quinn, S., & Markovits, H. (2002). Conditional reasoning with causal premises: Evidence for a retrieval model. *Thinking & Reasoning*, 8(3), 179-191.
- Racionero, Q. (1990). Introducción, comentarios y notas. En Aristóteles, *Retórica*. Gredos.
- Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 608-620.
- Ragni, M., Khemlani, S., & Johnson-Laird, P. N. (2014). The evaluation of the consistency of quantified assertions. *Memory & cognition*, 42(1), 53-66.
- Ramsey, F. P. (1926). Truth and Probability. En R. B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays* (pp. 156-198). London: Kegan Paul, Trench, Trubner & Co.; New York: Harcourt, Brace and Company. (1999 electronic edition).
- Ramsey, F. P. (1928). Further Considerations. En R. B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays*. London: Kegan Paul, Trench, Trubner & Co.; New York: Harcourt, Brace and Company. (1999 electronic edition).
- Ramsey, F. P. (1929). Probability and Partial Belief. En R. B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays*. London: Kegan Paul, Trench, Trubner & Co.; New York: Harcourt, Brace and Company. (1999 electronic edition).

- Rapanta, C., & Walton, D. (2016). The use of argument maps as an assessment tool in higher education. *International Journal of Educational Research*, 79, 211-221.
- Reboul, O. (1990). Rhétorique et dialectique chez Aristote. *Argumentation*, 4(1), 35-52.
- Reboul, O., & Johnstone, H. W. Jr. (1988). Can there be non-rhetorical argumentation? *Philosophy & Rhetoric*, 21(3), 220-233. Penn State University Press.
- Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition*, 34(1), 3-16. <https://doi.org/10.3758/BF03193382>
- Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: the effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, 130(3), 323.
- Reiter, R. (1980). A logic for default reasoning. *Artificial intelligence*, 13(1-2), 81-132.
- Renoult, L., Davidson, P. S., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: at the crossroads of semantic and episodic memory. *Trends in cognitive sciences*, 16(11), 550-558.
- Rescher, N. (1973). *The Coherence Theory of Truth*. Oxford: Oxford University Press.
- Rescher, N. (1976). *Plausible Reasoning: An Introduction to the Theory and Practice of Plausibilistic Inference*. BRILL.
- Rescher, N. (1977). *Dialectics: A controversy-oriented approach to the theory of knowledge*. Suny Press.
- Rescher, N. (2013). *Dialectics: A classical approach to inquiry*. Berlin: Walter de Gruyter.
- Rescher, N. (2018). Aristotle's theory of modal syllogisms and its interpretation. In *Critical Approaches to Science and Philosophy* (pp. 152-177). Routledge.
- Rescorla, M. (2009a). Cognitive maps and the language of thought. *The British Journal for the Philosophy of Science*, 60(2), 377-407.
- Rescorla, M. (2009b). Chrysippus' dog as a case study in non-linguistic cognition. In R. Lurz (Ed.), *The philosophy of animal minds* (pp. 52-71). Cambridge University Press.
- Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review*, 74(1), 71-80.
- Revlin, R., Cate, C. L., & Rouss, T. S. (2001). Reasoning counterfactually: Combining and rendering. *Memory & Cognition*, 29, 1196-1208.

- Rhodes, M., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition*, 108(2), 543-556.
- Richter, T. (2003). *Epistemologische Einschätzungen beim Textverstehen* (Doctoral dissertation).
- Richter, T. (2011). Cognitive flexibility and epistemic validation in learning from multiple texts. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility* (pp. 125-140). Berlin: Springer. https://doi.org/10.1007/978-94-007-1793-0_7
- Richter, T. (2015). Validation and comprehension of text information: Two sides of the same coin. *Discourse Processes*, 52(5-6), 337-355.
- Richter, T., & Maier, J. (2018). The role of validation in multiple-document comprehension. In *Handbook of multiple source use* (pp. 151-167). Routledge.
- Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96(3), 538.
- Ricoeur, P. (1980). *La metáfora viva* (A. Neira, Trad.). Madrid: Ediciones Europa.
- Rigotti, E. (2006). Relevance of context-bound loci to topical potential in the argumentation stage. *Argumentation*, 20, 519-540.
- Rigotti, E. (2009). Whether and how classical topics can be revived within contemporary argumentation theory. En F. H. van Eemeren & B. Garssen (Eds.), *Pondering on problems of argumentation: Twenty essays on theoretical issues* (pp. 157-178). Dordrecht: Springer Netherlands.
- Rigotti, E., & Greco Morasso, S. (2010). Comparing the argumentum model of topics to other contemporary approaches to argument schemes: The procedural and material components. *Argumentation*, 24, 489-512.
- Rigotti, E., & Greco, S. (2006). Topics: The argument generator. In E. Rigotti et al. (Eds.), *Argumentation for financial communication, Argumentum eLearning module*. Recuperado de <http://www.argumentum.ch>
- Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, 90(1), 38-71
- Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Mit Press.
- Rives, B. (2010). Fodor, Jerry. En J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Recuperado de <https://www.iep.utm.edu/fodor/>

- Rizzella, M. L., & O'Brien, E. J. (2002). Retrieval of concepts in script-based texts and narratives: the influence of general world knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 780.
- Rocci, A. (2005). Are manipulative texts coherent? Manipulation, presuppositions, and (in-) congruity. In L. de Saussure & P. Schulz (Eds.), *Manipulation and Ideologies in the Twentieth Century: Discourse, Language, Mind* (pp. 85-112). John Benjamins Publishing Company.
- Rodríguez, M. P. A., & Rodríguez, T. R. F. (2000). El concepto de inferencia inconsciente de Helmholtz: los problemas de su interpretación empirista y de una lectura computacional. *Revista de Historia de la Psicología*, 21(2/3), 275-286.
- Romney, A. K., & Batchelder, W. H. (2002). Teoría del consenso cultural. En R. A. Wilson & F. C. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (pp. 1267-1268). Madrid: Editorial Síntesis.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Routledge.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4), 573-605.
- Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, 126(3), 211.
- Rotenberg, K. J. (1995). The socialisation of trust: Parents' and children's interpersonal trust. *International Journal of Behavioral Development*, 18(4), 713-726.
- Rotenberg, K. J. (2010). The conceptualization of interpersonal trust: A basis, domain, and target framework. *Interpersonal trust during childhood and adolescence*, 8-27.
- Rotenberg, K. J., Petrocchi, S., Lecciso, F., & Marchetti, A. (2013). Children's trust beliefs in others and trusting behavior in peer interaction. *Child Development Research*, 2013.
- Rott, H. (2008). Belief revision. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of Human Inference and Its Foundations* (pp. 514-534). Cambridge: Cambridge University Press.
- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American psychologist*, 26(5), 443.
- Rubin, D. C. (2006). The basic-systems model of episodic memory. *Perspectives on Psychological Science*, 1(4), 277-311.
- Rubinelli, S. (2006). The ancient argumentative game: τόποι and loci in action. *Argumentation*, 20(3), 253-272.

- Rubinelli, S. (2009). *Ars Topica: The classical technique of constructing arguments from Aristotle to Cicero* (Vol. 15). Springer Science & Business Media.
- Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. E. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Russell, B. (1945). *Historia de la filosofía* (J. Gómez de la Serna & A. Dorta, Trans.; J. Mosterín, Pról.). En *Grandes Obras de la Cultura*. Barcelona: RBA Coleccionables.
- Russo, A. (1962). *La filosofía della retorica di Aristotele*. Nápoles.
- Ryan, R. M., & Deci, E. L. (2002). Overview of self-determination theory: An organismic dialectical perspective. In *Handbook of self-determination research* (pp. 3-33).
- Ryle, G. (1949). *The Concept of Mind*. New York: Barnes & Noble.
- Saaty, T. L., & Ozdemir, M. S. (2003). Why the magic number seven plus or minus two. *Mathematical and computer modelling*, 38(3-4), 233-244
- Sachs, O., & Hirsch, J. (2008). A neurology of belief. *Annals of Neurology*, 63, 129-130.
- Salmon, W. C. (1966). The foundations of scientific inference. En R. G. Colodny (Ed.), *The foundations of scientific inference* (pp. 135-275). Pittsburgh, PA: University of Pittsburgh Press.
- Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
- Salmon, W. C. (1989). *Four decades of scientific explanation*. Minneapolis, MN: University of Minnesota Press.
- Sanders, T. (2005, November). Coherence, causality and cognitive complexity in discourse. In *Proceedings/Actes SEM-05, First International Symposium on the Exploration and Modelling of Meaning* (pp. 105-114).
- Sanders, T. J., Spooren, W. P., & Noordman, L. G. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15(1), 1-35.
- Sanders, T. J., Spooren, W. P., & Noordman, L. G. (1993). Coherence relations in a cognitive theory of discourse representation. *Discourse Processes*, 16(1-2), 35-60.
- Sant'Anna, A. (2018). Episodic memory as a propositional attitude: A critical perspective. *Frontiers in Psychology*, 9, 370748. <https://doi.org/10.3389/fpsyg.2018.03748>

- Santos del Cerro, J. (2002). Probabilismo moral y probabilidad. En A. H. E. P. E. (Ed.), *Historia de la probabilidad y de la estadística* (pp. 103-118). Madrid: Editorial AC.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Schacter, D. L., Dobbins, I. G., & Schnyer, D. M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience*, 5(11), 853-862.
- Schank, R. C., & Abelson, R. P. (1975, September). Scripts, plans, and knowledge. En *Proceedings of the International Joint Conference on Artificial Intelligence* (Vol. 75, pp. 151-157).
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Schellens, P. J., & De Jong, M. (2004). Argumentation schemes in persuasive brochures. *Argumentation*, 18, 295-323.
- Scholman, M. C. J. (2019). *Coherence relations in discourse and cognition: Comparing approaches, annotations, and interpretations* (Doctoral dissertation). Utrecht University.
- Schramm, W. (1954). How communication works. En W. Schramm (Ed.), *The process and effects of communication* (pp. 3-26). University of Illinois Press.
- Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, 59(3), 237-255.
- Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, 59(3), 237-255.
- Schroyens, W., Schaeken, W., & Handley, S. (2003). In search of counter-examples: Deductive rationality in human reasoning. *The Quarterly Journal of Experimental Psychology Section A*, 56(7), 1129-1145.
- Schuessler, R. (2015). Probability in medieval and renaissance philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 ed.). Retrieved from <https://plato.stanford.edu/archives/spr2015/entries/probability-medieval-renaissance/>
- Schuessler, R. (2019). Probability in medieval and renaissance philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Retrieved from <https://plato.stanford.edu/archives/sum2019/entries/probability-medieval-renaissance/>
- Schwitzgebel, E. (2019). Belief. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Recuperado de <https://plato.stanford.edu/archives/sum2019/entries/belief/>

- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424.
- Searle, J. R. (1990). Is the brain's mind a computer program?. *Scientific American*, 262(1), 25-31.
- Sedlmeier, P. (2002). Associative learning and frequency judgements: The PASS model. In P. Sedlmeier & T. Betsch (Eds.), *ETC. Frequency processing and cognition* (pp. 137-152). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198508632.003.0009>
- Seitz, R. J., & Angel, H. F. (2020). Belief formation—A driving force for brain evolution. *Brain and Cognition*, 140, 105548.
- Seitz, R. J., Paloutzian, R. F., & Angel, H. F. (2018). From believing to belief: a general theoretical model. *Journal of Cognitive Neuroscience*, 30(9), 1254-1264.
- Shafer, G. (1992). Influence diagrams, belief nets, and decision analysis. *International Journal of Approximate Reasoning*, 6(3), 445-480. Elsevier.
- Shah, N., & Velleman, J. D. (2005). Doxastic deliberation. *The Philosophical Review*, 114(4), 497-534.
- Shastri, L. (2002). Episodic memory and cortico–hippocampal interactions. *Trends in Cognitive Sciences*, 6(4), 162-168.
- Shipley, T. F., & Zacks, J. M. (Eds.). (2008). *Understanding events: From perception to action*. Oxford University Press.
- Shore, B. (1996). *Culture in mind: Cognition, culture, and the problem of meaning*. New York, NY: Oxford University Press.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54(4), 574-591.
- Singer, M. (2013). Causal bridging inferences: Validating consistent and inconsistent sequences. In *Reading and language processing* (pp. 212-231). Psychology Press.
- Singer, M. (2013). Validation in reading comprehension. *Current Directions in Psychological Science*, 22(5), 361-366.
- Singer, M. (2019). Challenges in processes of validation and comprehension. *Discourse Processes*, 56(5-6), 465-483.
- Singer, M., & Ritchot, K. F. (1996). The role of working memory capacity and knowledge access in text inference processing. *Memory & cognition*, 24(6), 733-743.
- Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language*, 31(4), 507-524.

- Skovgaard-Olsen, N., Collins, P., & Klauer, K. C. (2023). Possible worlds truth table task. *Cognition*, 238, 105507.
- Sloman, S. A. (1993). Feature-based induction. *Cognitive psychology*, 25(2), 231-280.
- Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgements of likelihood. *Cognition*, 52(1), 1-21.
- Sloman, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking & Reasoning*, 3(2), 81-110.
- Smith, B. C. (2002). Computación. En H. Wilson & F. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (versión española, pp. 331-334). Madrid: Editorial Síntesis.
- Smith, R. (2022). Aristotle's Logic. In: Zalta, E. N. & Nodelman, U. (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 ed.). URL = <<https://plato.stanford.edu/archives/win2022/entries/aristotle-logic/>>.
- Snoek Henkemans, A. F. (2001). Argumentation, explanation and causality: An exploration of current linguistic approaches to textual relations. In S. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text Representation: Linguistic and Psycholinguistic Aspects* (pp. 231-246). John Benjamins Publishing.
- Sober, E. (2000). *Philosophy of biology* (2nd ed.). Boulder, CO: Westview Press.
- Solmsen, F. (1924). *Die Entwicklung der aristotelischen Logik und Rhetorik*. Berlin.
- Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology*, 43(2), 129-169.
- Sperber, D. (1996). *Explaining culture: A naturalistic approach* (p. 97). Blackwell: Oxford.
- Sperber, D. (2001). An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, 29(1-2), 401-413.
- Sperber, D. (2013). Speakers are honest because hearers are vigilant: Reply to Kourken Michaelian. *Episteme*, 10(1), 61-71.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Cambridge, MA: Harvard University Press.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & language*, 25(4), 359-393.

- Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford University Press.
- Sprute, J. (1977). Der Zweck der aristotelischen Rhetorik. En R. Bubner (Ed.), *Logik, Ethik, Theorie der Geisteswissenschaften*. Hamburgo
- Sprute, J. (1982). *Die Enthymementheorie der aristotelischen Rhetorik*. Gotinga
- Stanovich, K. (2011). *Rationality and the reflective mind*. New York, NY: Oxford University Press.
- Stanovich, K. E., & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking and Reasoning*, 4, 193-230.
- Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of experimental psychology: general*, 127(2), 161.
- Sterelny, K. (2002). Lenguaje del Pensamiento. En H. Wilson & F. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (pp. 708-710). Madrid: Editorial Síntesis.
- Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology*, 48(3), 613-643.
- Sugiura, M., Seitz, R. J., & Angel, H. F. (2015). Models and neural bases of the believing process. *Journal of Behavioral and Brain Science*, 5(1), 12.
- Sutherland, R. J., & Rudy, J. W. (1989). Configural association theory: The role of the hippocampal formation in learning, memory, and amnesia. *Psychobiology*, 17(2), 129-144.
- Sweegers, C. C. G., Coleman, G. A., Van Poppel, E. A. M., Cox, R., & Talamini, L. M. (2015). Mental schemas hamper memory storage of goal-irrelevant information. *Frontiers in human neuroscience*, 9, 629.
- Sweegers, C. C., & Talamini, L. M. (2014). Generalization from episodic memories across time: A route for semantic knowledge acquisition. *Cortex*, 59, 49-61.
- Tardini, S. (2005). Endoxa and communities: Grounding enthymematic arguments. *Studies in Communication Sciences, Special Issue "Argumentation in Dialogic Interaction"*, 279-293.
- Tarski, A. (1954). Contributions to the theory of models. I. In *Indagationes Mathematicae (Proceedings)* (Vol. 57, pp. 572-581). Elsevier BV.
- Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In *Advances in experimental social psychology* (Vol. 11, pp. 249-288). Academic Press.

- Teira, D. (2012). Probabilidad. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 485-491). Trotta.
- Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and brain sciences*, 24(4), 629-640.
- Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in cognitive sciences*, 10(7), 309-318.
- Tešić, M., & Hahn, U. (2023). The impact of explanations as communicative acts on belief in a claim: The role of source reliability. *Cognition*, 240, 105586.
- Thagard, P. (2000). *Coherence in Thought and Action*. Cambridge, Mass.: The MIT Press.
- Thagard, P. (2002). Explicación. En H. Wilson & F. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas* (pp. 522-524). Madrid: Editorial Síntesis.
- Thagard, P. (2005). Testimony, credibility, and explanatory coherence. *Erkenntnis*, 63(3), 295–316.
- Thagard, P. (2007). Coherence, truth, and the development of scientific knowledge. *Philosophy of science*, 74(1), 28-47.
- Thagard, P. (2009). Why cognitive science needs philosophy and vice versa. *Topics in Cognitive Science*, 1(2), 237-254.
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, 22(6), 742-758.
- Thompson, V. A., Evans, J. S. B. T., & Handley, S. J. (2005). Persuading and dissuading by conditional argument. *Journal of Memory and Language*, 53(2), 238-257.
- Tindale, C. W. (2007). Revisiting Aristotle's topoi. *OSSA Conference Archive*, 141. Recuperado de <https://scholar.uwindsor.ca/ossaarchive/OSSA7/papersandcommentaries/141>
- Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.
- Toulmin, S. (2002). Prólogo. En *Los usos de la argumentación* (pp. 9-11). Barcelona: Editorial Gredos. (Obra original publicada en 1958; edición actualizada 2007).
- Toulmin, S. (2007). Los usos de la argumentación (María Morrás y Victoria Pineda, Trans.). *Barcelona, España: Ediciones Península (primera edición en inglés, 1958)*.
- Trabasso, T. (2005). The role of causal reasoning in understanding narratives. *From orthography to pedagogy: Essays in honor of Richard L. Venezky*, 81-106.

- Trabasso, T. (2014). The role of causal reasoning in understanding narratives. In T. R. Trabasso, J. P. Sabatini, D. W. Massaro, & R. Calfee (Eds.), *From orthography to pedagogy: Essays in honor of Richard L. Venezky* (pp. 81-106). New York: Psychology Press.
<https://doi.org/10.4324/9781410613028>
- Trabasso, T., Van den Broek, P., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse processes*, 12(1), 1-25.
- Trevarthen, C. (2002). Intersubjectivity. En R. A. Wilson & F. C. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. 667-669). MIT Press.
- Tucker, C. (2010). Why open-minded people should endorse dogmatism. *Philosophical Perspectives*, 24(1), 529-545. <https://doi.org/10.1111/j.1520-8583.2010.00202.x>
- Tullett, A. M., Prentice, M. S., Teper, R., Nash, K. A., Inzlicht, M., & McGregor, I. (2013). Neural and motivational mechanics of meaning and threat. *American Psychological Association*.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York, NY: Academic Press.
- Tulving, E., & Craik, F. I. (Eds.). (2000). *The Oxford handbook of memory*. Oxford University Press.
- Tulving, Endel (1983), *Elements of Episodic Memory*, New York: Oxford University Press.
- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42(1), 230-265.
- Twardy, C. (2004). Argument maps improve critical thinking. *Teaching Philosophy*, 27(2), 95-116.
- Van Bockstaele, B., Verschuere, B., Moens, T., Suchotzki, K., Debey, E., & Spruyt, A. (2012). Learning to lie: Effects of practice on the cognitive cost of lying. *Frontiers in psychology*, 3, 33246.
- Van Boekel, M., Lassonde, K. A., O'Brien, E. J., & Kendeou, P. (2017). Source credibility and the processing of refutation texts. *Memory & Cognition*, 45, 168-181.
- Van Camp, W. (2014). Explaining understanding (or understanding explanation). *European Journal for Philosophy of Science*, 4(1), 95-114.
- Van Cleve, J. (2005). Why Coherence is Not Enough: A Defense of Moderate Foundationalism. En M. Steup & E. Sosa (Eds.), *Contemporary Debates in Epistemology* (pp. 168-180). Oxford: Blackwell.

- Van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539-588). Academic Press.
- Van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54(5-6), 360-372.
- Van den Broek, P., Bohn-Gettler, C. M., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 123-139). Information Age Publishing.
- Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. *The construction of mental representations during reading*, 71-98.
- Van Dijk, T. A. (1977). Context and cognition: Knowledge frames and speech act comprehension. *Journal of Pragmatics*, 1(3), 211-231.
- Van Dijk, T. A. (1980). The semantics and pragmatics of functional coherence in discourse. *Speech act theory: Ten years later*, 49-65.
- Van Dijk, T. A. (1997). The study of discourse. *Discourse as structure and process*, 1(34), 703-52.
- Van Dijk, T. A. (1998). *Texto y contexto: Semántica y pragmática del discurso* (6a ed.). Cátedra.
- Van Dijk, T. A. (1999). *Ideología: Una aproximación multidisciplinaria*. Barcelona: Gedisa.
- Van Dijk, T. A. (2003). The discourse-knowledge interface. In *Critical discourse analysis: Theory and interdisciplinarity* (pp. 85-109). London: Palgrave Macmillan UK.
- Van Dijk, T. A. (2006). Discourse, context and cognition. *Discourse studies*, 8(1), 159-177.
- Van Dijk, T. A. (2007). The study of discourse: An introduction. In T. A. van Dijk (Ed.), *Discourse studies* (pp. xix-xlii). London: Sage.
- Van Dijk, T. A. (2008). *Discourse and context: A sociocognitive approach*. Cambridge: Cambridge University Press.
- Van Dijk, T. A. (2012). A note on epistemics and discourse analysis. *British Journal of Social Psychology*, 51(3), 478-485.
- Van Dijk, T. A. (2013). *Discurso y contexto*. Barcelona: Editorial Gedisa.

- Van Dijk, T. A. (2014). Discourse and knowledge. In *The Routledge handbook of discourse analysis* (pp. 587-603). Routledge.
- Van Dijk, T. A. (2014). *Discourse and knowledge: A sociocognitive approach*. Cambridge University Press.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- Van Eemeren, F. H. (2010). *Strategic maneuvering in argumentation discourse: Extending the pragma-dialectical theory of argumentation*. Amsterdam: Benjamins.
- Van Eemeren, F. H., & Grootendorst, R. (1984). *Speech acts in argumentative discussions: A theoretical model for the analysis of discussions directed towards solving conflicts of opinion*. Dordrecht/Berlin: Foris/Mouton de Gruyter.
- Van Eemeren, F. H., & Grootendorst, R. (2002). *Argumentación, comunicación y falacias: Una perspectiva pragma-dialéctica* (C. López & A. M. Vicuña, Trans.). Santiago de Chile: Universidad Católica de Chile. (obra original en inglés publicada en 1992).
- Van Eemeren, F. H., & Grootendorst, R. (2011). *Una teoría sistemática de la argumentación: La perspectiva pragmadialéctica* (C. López y A. M. Vicuña, Trad.). Buenos Aires: Editorial Biblos. (Obra original publicada en 2004).
- van Eemeren, F. H., Grootendorst, R., Snoeck Henkemans, A. F., Blair, J. A., Johnson, R. H., Krabbe, E. C. W., Plantin, C., Walton, D. N., Willard, C. A., Woods, J., & Zarefsky, D. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Lawrence Erlbaum Associates.
- Van Enschot-Van Dijk, R., Hustinx, L., & Hoeken, H. (2003). The concept of argument quality in the Elaboration Likelihood Model: A normative and empirical approach to Petty and Cacioppo's 'strong' and 'weak' arguments. In *Anyone who has a view: Theoretical contributions to the study of argumentation* (pp. 319-335). Dordrecht: Springer Netherlands.
- Van Fraassen, B. C. (1983). Calibration: Frequency justification for personal probability. In R. S. Cohen & L. Laudan (Eds.), *Physics, philosophy, and psychoanalysis*. Dordrecht: Springer.
- Van Fraassen, B. C. (1984). Belief and the will. *Journal of Philosophy*, 81, 235-256.
- Van Gelder, T. (2002). Argument mapping with reason! able. *The American philosophical association newsletter on philosophy and computers*, 2(1), 85-90.
- Van Gelder, T. (2007). The rationale for Rationale™. *Law, probability and risk*, 6(1-4), 23-42.

- Van Kampen, H. S. (2019). The principle of consistency and the cause and function of behaviour. *Behavioural Processes*, 159, 42-54.
- Vancil, D. L. (1979). Historical barriers to a modern system of topoi. *Western Journal of Communication (includes Communication Reports)*, 43(1), 26-37.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.
- Vega, L. (1993). Tà Éndoxa: argumentación y plausabilidad. *Éndoxa*, (1), 5-19.
- Vega, L. (1998). Aristotle's endoxa and plausible argumentation. *Argumentation*, 12(1), 95-113.
- Vega, L. (2004). Entimemas. *Cuadernos de Filosofía del Derecho*, 27, 283-315.
- Vega, L. (2012). Argumentación, teoría de la. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 55-66). Trotta.
- Vega, L. (2012). Esquema argumentativo. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 233-236). Trotta.
- Vega, L. (2012). Esquema argumentativo. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 233-237). Trotta.
- Vega, L. (2012). Inducción. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 294-297). Trotta.
- Vega, L. (2012). Informal, lógica. En L. Vega Reñón & P. Olmos Gómez (Eds.), *Compendio de Lógica, Argumentación y Retórica* (2ª ed. revisada y actualizada, pp. 308-312). Trotta.
- Vélez, N., Bridgers, S., & Gweon, H. (2019). The rare preference effect: Statistical information influences social affiliation judgments. *Cognition*, 192, 103994.
- Verhagen, A. (2005). *Constructions of intersubjectivity: Discourse, syntax, and cognition*. Oxford University Press.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.
- Von Wright, G. H. (1951). Deontic logic. *Mind*, 60(237), 1-15.
- Voss, J. F., Wiley, J., & Sandak, R. (1999). On the use of narrative as argument. In *Narrative Comprehension, Causality, and Coherence* (pp. 241-258). Routledge.

- Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in cognitive sciences*, 10(4), 141-142.
- Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, 20(1), 28-32.
- Wallace, A. F., & Fogelson, R. D. (1961). Culture and personality. *Biennial Review of Anthropology*, 42-78.
- Walster, E. (1966). Assignment of responsibility for an accident. *Journal of Personality and Social Psychology*, 3, 73-79.
- Walster, E., Aronson, E., & Abrahams, D. (1966). On increasing the persuasiveness of a low prestige communicator. *Journal of Experimental Social Psychology*, 2(4), 325-342.
- Walton, D. (1992). Rules for plausible reasoning. *Informal Logic*, 14(1), 33-51.
- Walton, D. (1996). *Argumentation schemes for presumptive reasoning*. Psychology Press.
- Walton, D. (1999). Historical origins of argumentum ad consequentiam. *Argumentation*, 13, 251-264.
- Walton, D. (2001). Enthymemes, common knowledge, and plausible inference. *Philosophy & rhetoric*, 34(2), 93-112.
- Walton, D. (2002). Are some Modus Ponens arguments deductively invalid? *Informal Logic*, 22(1), 1-13.
- Walton, D. (2004). Classification of fallacies of relevance. *Informal logic*, 24(1).
- Walton, D. (2006). Argument from appearance: A new argumentation scheme. *Logique et Analyse*, 49(195), 319-340.
- Walton, D. (2008). *Informal logic: A pragmatic approach*. Cambridge University Press.
- Walton, D. (2021). Aristotelian dialectic, argumentation theory and artificial intelligence. *Essays on argumentation in antiquity*, 245-277.
- Walton, D., & Macagno, F. (2005). Common knowledge in legal reasoning about evidence. *International Commentary on Evidence*, 3(1).
- Walton, D., & Macagno, F. (2006). Common knowledge in argumentation. *Studies in Communication Sciences*, 6(1), 3-26.

- Walton, D., & Macagno, F. (2009). Enthymemes, argumentation schemes and topics. *Logique et Analyse*, 52(205), 39-56.
- Walton, D., & Macagno, F. (2015). A classification system for argumentation schemes. *Argument & Computation*, 6(3), 219-245.
- Walton, D., & Reed, C. (2002, July). Argumentation schemes and defeasible inferences. In *Workshop on Computational Models of Natural Argument, 15th European Conference on Artificial Intelligence* (pp. 11-20).
- Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. Cambridge University Press.
- Walton, D., Tindale, C. W., & Gordon, T. F. (2014). Applying recent argumentation methods to some ancient examples of plausible reasoning. *Argumentation*, 28, 85-119.
- Warnick, B. (2000). Two systems of invention: The topics in the Rhetoric and The New Rhetoric. In A. Gross & A. Walzer (Eds.), *Rereading Aristotle's rhetoric* (pp. 107-129). Southern Illinois University Press.
- Waskan, J. A. (2003). Intrinsic cognitive models. *Cognitive science*, 27(2), 259-283.
- Wason, P. C. (1966). *New horizons in psychology*. Penguin Books.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content* (Vol. 86). Harvard University Press.
- Weidenfeld, A., Oberauer, K., & Hörnig, R. (2005). Causal and noncausal conditionals: An integrated model of interpretation and reasoning. *The Quarterly Journal of Experimental Psychology Section A*, 58(8), 1479-1513.
- Weller, S. C. (1987). Shared knowledge, intracultural variation, and knowledge aggregation. *American Behavioral Scientist*, 31(2), 178-193.
- Wellner, B., Pustejovsky, J., Havasi, C., Rumshisky, A., & Sauri, R. (2006, July). Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue* (pp. 117-125).
- Wertgen, A. G., & Richter, T. (2020). Source credibility modulates the validation of implausible information. *Memory & Cognition*, 48, 1359-1375. <https://doi.org/10.1016/j.cognition.2010.08.015>
- West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of educational psychology*, 100(4), 930.

- Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, 36(2), 109-129.
- Williams, A. J., & Danovitch, J. H. (2019). What does Mickey Mouse know about food? Children's trust in favorite characters versus experts. *Journal of Experimental Child Psychology*, 187, 104647.
- Wilson, D., & Sperber, D. (2004). Relevance theory. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 607-632). Oxford: Blackwell.
- Winocur, G., Moscovitch, M., & Sekeres, M. (2007). Memory consolidation or transformation: context manipulation and hippocampal representations of memory. *Nature neuroscience*, 10(5), 555-557.
- Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Norwood, NJ: Ablex Publishing Corporation.
- Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A manual for the embedded figures tests*. Palo Alto, CA: Consulting Psychologists Press.
- Wittgenstein, L. (1922). *Tractatus logico-philosophicus* (C. K. Ogden, Trans.). New York: Routledge.
- Wodak, R. (2016). Dialogue matters: Transcending dichotomies. *Journal of Multicultural Discourses*, 11(4), 367-374.
- Wolfe, C. R., & Britt, M. A. (2008). The locus of the myside bias in written argumentation. *Thinking & reasoning*, 14(1), 1-27.
- Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication*, 26(2), 183-209.
- Wyer Jr, R. S. (2007). Principles of mental representation. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 285-307).
- Wyer, R. S., & Albarracín, D. (2005). Belief formation, organization, and change: Cognitive and motivational influences. *The handbook of attitudes*, 273, 322.
- Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current directions in psychological science*, 16(2), 80-84.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.
[https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Žagar, I. (2010). Topoi in critical discourse analysis. *Lodz Papers in Pragmatics*, 6(1), 3-27.

- Zarefsky, D. (1995). Argumentation in the tradition of speech communication studies. In F. H. van Eemeren, R. Grootendorst, J. A. Blair, & C. A. Willard (Eds.), *Perspectives and approaches: Proceedings of the Third International Conference on Argumentation* (Vol. 1, pp. 32–52). Sic Sat.
- Zwaan, R. A. (1999). Five dimensions of narrative comprehension: The event-indexing model. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 93-110). Lawrence Erlbaum Associates Publishers.
- Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. En B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 44, pp. 35-62). Academic Press.
- Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, 23(4), 1028-1034.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162-185.
- Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, 10(4), 954-958.
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 386-397.

