What makes a word a good representative of the category of "emotion"? The role of feelings and interoception

Pilar Ferré [1*], Marc Guasch [1], Hans Stadthagen-González [2], José Antonio Hinojosa [3,4,5], Isabel Fraga [6], Javier Marín [7], Miguel Ángel Pérez-Sánchez [7]

[1] Universitat Rovira i Virgili. Departament de Psicologia and CRAMC, Tarragona, Spain.

[2] Department of Psychological Sciences, Benedictine College, United States.

[3] Departamento de Psicología Experimental, Procesos Cognitivos y Logopedia, Universidad Complutense de Madrid, Madrid, Spain.

[4] Instituto Pluridisciplinar, Universidad Complutense de Madrid, Madrid, Spain.

[5] Centro de Investigación Nebrija en Cognición (CINC), Universidad Nebrija, Madrid, Spain

[6] Departmento de Psicología Social, Psicología Básica y Metodología, Universidad de Santiago de Compostela, Santiago de Compostela, Spain.

[7] Departamento de Psicología Básica y Metodología, Universidad de Murcia, Murcia, Spain.

OSF repository link: https://osf.io/a2hcu

* Correspondence author:

Pilar Ferré

Departament de Psicologia and CRAMC. Universitat Rovira i Virgili

Carretera de Valls, s.n. 43007 Tarragona, Spain

Email: mariadelpilar.ferre@urv.cat

Abstract

The words we use to describe emotions vary in terms of prototypicality; that is, some of these words may be more representative of the semantic category of emotion than others (e.g., *anger* refers more clearly to an emotion than *boredom*). Based on a multi-componential conception of emotions, the aim of the present study was to examine the contribution of several variables to emotion prototypicality. Some of those variables are related to the distinct components of emotions: evaluation, action, body expression, internal body sensations (interoception), and feelings. Other variables are related to the concreteness/abstractness distinction: sensory experience, social interaction, thought and morality. We collected ratings for these variables for a large set of words (1,286) which varied in emotion prototypicality. A regression analysis revealed that the variables that most contributed to emotion prototypicality were feelings and interoception. Furthermore, a factor analysis identified two underlying factors: socio-emotional polarity and emotional experience. The scores of each word in both factors were used to create a two-dimensional space and a density plot which provides relevant information about the organization of emotion concepts in memory.

Keywords: emotion prototypicality, multi-componential conception of emotions, emotion words, feelings, interoception

A way to characterize an emotion lexicon, that is, a corpus that includes all the emotion words in a particular language (e.g., anger, fear), is in terms of prototypicality. This approach is inspired by Rosch's (1978) work and considers that emotion concepts, like other categories such as colours, vary in the degree to which they represent an emotion (e.g., Ortony et al., 1987; Shaver et al., 1987, 2001). That is, some words are more representative of the semantic category of emotions than others and therefore convey a particular emotional meaning more strongly.

The prototypicality approach has been followed in a series of studies conducted in the last decades. The common procedure is to provide participants with a list of potential emotion terms and ask them to rate the extent to which each word refers to an emotion (e.g., Alonso-Arbiol et al., 2006; Galati et al., 2008; Niedenthal et al., 2004; Pérez-Sánchez et al., 2021; Shaver et al., 1987, 2001; Zammuner, 1998). The results of these studies clearly show that some emotion terms (e.g., happiness) are perceived as more prototypical than others (e.g., fun). Interestingly, emotion prototypicality has been shown to influence word retrieval from the mental lexicon. For instance, responses are faster for words of higher emotional prototypicality in category verification tasks (Fehr & Russell, 1984; Niedenthal et al., 2004; Zammuner, 1998) as well as in lexical decision tasks (Haro et al., 2022).

An additional aim of many of these studies has been to identify the features or dimensions that contribute to emotion prototypicality. For instance, in a study conducted with 153 Italian words, Zammuner (1998) identified emotional charge (i.e., the degree to which a word denotes an affective experience, regardless of its pleasant/unpleasant nature), intensity (i.e., the amount of the affective experience), and duration (i.e., the time that the affective experience denoted by the term lasts), as the best predictors of emotion prototypicality. In a similar study conducted with 237 French

terms, Niedenthal et al. (2004) found that intensity was the most reliable predictor of prototypicality. In the largest study conducted from this approach, Pérez-Sánchez et al. (2021) collected emotion prototypicality ratings for 1,286 Spanish words. They found that emotional charge was the best predictor of prototypicality, followed by four discrete-emotion related variables (i.e., the extent to which the word was related to sadness, happiness, disgust, and anger), arousal (i.e., the degree of activation entailed by an affective experience, Bradley & Lang, 2000), and two lexical-semantic variables (i.e., age of acquisition of the word and lexical frequency).

All in all, these studies suggest that the most representative exemplars of the category of emotions are words that denote an intense affective experience (regardless of its positive or negative polarity; Niedenthal et al., 2004; Pérez-Sánchez et al., 2021; Zammuner, 1998). It should be noted, however, that the emotional experience is a complex process. In this regard, different approaches have been developed with the aim of providing a theoretical framework that accounts for the key features of emotions. The most influential views in this area can be classified as: basic emotion theories (e.g., Ekman, 2018), appraisal theories (e.g., Scherer & Moors, 2019), and constructionist theories (e.g., Lindquist, 2013). These proposals differ in their conception of emotions and in their major focus (Scherer, 2022). Basic emotion theories propose the existence of a small group of discrete emotions, which are elicited by innate affect programs. These emotions have characteristic patterns of behavioral expressions, as well as specific neurobiological markers. Appraisal theories state that emotions are adaptive responses to a stimulus or event, which is appraised as relevant in terms of the survival and well-being of the individual. Emotions involve changes in several organismic components. Under this view, appraisal influences the intensity and quality of the other components (i.e., action tendencies, physiological response, behavior and feelings),

contributing to the differentiation of emotions. Finally, constructionist theories posit that emotions are not natural kinds but rather psychological constructions or categories created by people. “Core affect” (i.e., the conceptualization of subjective feelings) is a central concept for some of these theories. According to this view, the dimensions of valence (pleasantness-unpleasantness) and arousal (activation-deactivation) (e.g., Russell & Barrett, 2009) would be at the heart of the representation of emotion concepts. Of note, since the seminal work of Bradley & Lang (1999), who provided normative ratings, valence and arousal have been the key variables in research on the affective characteristics of words and their role in cognitive processing.

Despite these divergences, there are several ideas that are shared among these different accounts. All of three approaches propose that emotions are elicited by a stimulus or event, that is appraised by the individual as relevant to his/her survival or well-being. Emotions involve a set of responses to deal with the stimulus, which include several dimensions: 1) a strong motivation to act, 2) a set of physiological changes that allow for rapid action, 3) a behavioral expression and 4) a subjective feeling. In sum, there seems to be a consensus between these models to characterize emotions as multi-componential phenomena, which involve at least five main components: appraisal mechanisms, behavioral or body expression, physiological reaction, action tendency (motivation to act) and feeling (i.e., a particular subjective experience, Sander et al., 2018; van Berkum, 2022).

The few studies which have explored the factors that contribute to emotion prototypicality have relied on participants’ assessments about emotional experiences in terms of valence, arousal, intensity, and emotional charge (Niedenthal et al., 2004; Pérez-Sánchez et al., 2021; Zammuner, 1998). No study to date has examined the contribution of each of the five emotion components mentioned above, so it is unkown

what information the participants have relied on to make their judgments about the prototypicality of verbal labels denoting emotions. This is the first goal of the present study. Through a multidimensional rating approach, we aimed to examine whether all these components contribute equally to the emotional prototypicality of a word or if some components have more weight than others in giving a word its emotional content. Knowing the key features of the underlying representation of emotion concepts has theoretical and applied implications, considering the critical role of emotion concepts in emotion perception (e.g., Brooks & Freeman, 2018) and emotion experience (Lindquist & Barrett, 2008).

To shed light in this question, we collected ratings for the whole set of 1,286 words included in the normative study of Pérez-Sánchez et al. (2021) on several variables and examined their contribution to emotion prototypicality. These variables were related to the five components mentioned above: Evaluation (related to the appraisal component), action (related to the action tendency component), body expression (related to the expression component), interoception (i.e., inner body sensations, related to the physiological component), and feelings (related to the feeling component or the subjective experience of emotion).

We also collected ratings for several additional characteristics, which are relevant to the mental organization of words in terms of concreteness/abstractness (e.g., Crutch et al., 2013; Troche et al., 2017). In particular, the variables considered here were: sensory experience (i.e., the extent to which the word is related to visual, olfactory, gustatory, auditory, and tactile sensations), social interaction (i.e., the extent to which a particular word is related to relationships between people), thought (i.e., the extent to which the word is related to mental activities, ideas, opinions, etc.), and morality (i.e., the extent to which the word is associated to the motivation to act in

accordance to certain social rules; Crutch et al., 2013). Sensory experience is known to play an important role in the representation of concrete concepts and their differentiation with respect to abstract concepts (Troche et al., 2014). Social interaction and introspection (thought) have been proposed to contribute to abstract word acquisition and representation (Barsalou, 1999; Borghi et al., 2022). Morality, like thought, denotes phenomena that are not directly observable but instead reflex complex social rules that govern behavior (e.g., honesty; Troche et al., 2014). The reason for including these variables is that emotional information has been proposed to play a crucial role in the representation and processing of abstract concepts (Kousta et al., 2011). In fact, most emotion words refer to abstract states (Crutch et al., 2013), although not all abstract words are affectively loaded (Borghi et al., 2018). Recent work in this area highlights the need to characterize the different kinds of abstract concepts (e.g., Desai et al., 2018, Fingerhut & Prinz, 2018). Considering that most words in the Pérez-Sánchez et al.'s (2021) dataset are abstract words and that they vary in terms of emotion prototypicality, these variables may provide valuable information about the properties that give an abstract word its affective content.

Our second goal was to examine the organizational structure of emotion words, given the large number of words and the large list of variables collected. To that end, we elaborated a high-dimensional semantic space wherein the emotion terms are represented. Some of the studies that have used a prototypicality approach have already described the structure of the emotion lexicon in different languages (e.g., Alonso-Arbiol et al., 2006, in Basque; Galati et al., 2008, in neo-Latin languages; Shaver et al., 1987; 2001, in English and Indonesian, respectively). The common approach in these studies has been to ask participants to group together words that refer to similar emotions, but following this procedure with a large set of words would make the task

unmanageable. For that reason, the number of emotion words included in these studies is not very large (around one hundred). In the present study, we used a different approach inspired by the methodology followed in research on the representation of concepts in semantic memory (Crutch et al., 2013; Troche, 2017; Troche et al., 2017). We generated a high-dimensional semantic space of the Spanish emotion words from speakers' ratings in the variables collected in the present study as well as in other affective variables collected by Pérez-Sánchez et al. (2021), namely, valence, arousal, and emotionality (i.e., emotional charge, or the hedonic load of a word, regardless of its polarity). The last variables were included because previous studies showed them to be relevant for the characterization of emotion words (Alonso-Arbiol et al., 2006, Galati et al., 2008, Pérez-Sánchez et al., 2021; Shaver et al., 1987; 2001). We also examined how emotion words are distributed in this space through a density plot, which allowed us to identify the regions of the space which are more populated, as well as which words are closer to each other in terms of their conceptual representation.

To sum up, we collected ratings for a large pool of words on a large set of variables, with the double aim of: (1) identifying the crucial features related to the components of emotions that define emotion prototypical concepts, and (2) characterizing the structural organization of these concepts. These goals not only have theoretical relevance in terms of our knowledge about the information that individuals use to identify a particular state or experience as an emotion, but also have important practical implications. For instance, for research involving clinical populations, especially for the study of disorders in which the processing of emotional information is anomalous and for neurological conditions in which there is a loss of semantic-affective knowledge.

Method

Participants

We collected ratings from 651 native speakers of Spanish. Their mean age was 22.2 years (range: 20–67; *SD*: 1.4) and 555 of them (85.3%) were women. The sample size was selected to attain at least 20 responses for each word and variable. This is the common minimum value in other normative studies that have collected ratings for a similar number of stimuli and variables[1] (e.g., Li et al., 2017; Stadthagen-Gonzalez et al., 2017; Xu et al., 2022). Most participants were students at universities in two different regions of Spain, namely, Universitat Rovira i Virgili (URV) in Tarragona (471 participants) and Universidad de Murcia (UMU) (178 participants). Two undergraduates from the Universidade de Santiago de Compostela (USC), and two from the Universidad Complutense de Madrid (UCM) also participated. All participants signed an informed consent form. Each participant completed up to five questionnaires in exchange for academic credit or as volunteers. Participants who completed more than one questionnaire did not repeat the same set of words. After data trimming (see the Results section), we collected a total of 1,135 valid questionnaires. All participants signed an informed consent form before starting the ratings. The research was conducted in line with the APA ethical standards. Approval was granted by the Ethics Committee for Research on People, Society and the Environment of the Universitat Rovira i Virgili (CEIPSA-2021-PR-0044).

[1] Considering the large number of words included in our questionnaires, 20 raters per item seems to be an optimal value to expect high reliability ratings. Indeed, if we chose the ICC as a reliability index, we can use the formulas by Zou (2012), which are implemented in the R package *ICC.Sample.Size* version 1, to estimate the number of items (N) needed to obtain a given ICC expected value from 20 raters. For instance, for an expected ICC = .8, with no ICC below .65 (which are values based on prior results), 20 raters, alpha = 0.05 with two tails, and power = 0.80 (code = *calculateIccSampleSize (p = 0.8, p0 = 0.65, k = 20, alpha = 0.05, tails = 2, power = 0.80)*), we obtain an estimated N = 31 items. Our questionaries had at least 238 items each (i.e., each participant rated at least 238 words), which is over seven times higher. Therefore, 20 participants rating such a large set of items seems to be enough to expect high ICC values.

Materials

All the words included in the Pérez-Sánchez et al. (2021) database were included in this study. These authors aimed to provide prototypicality ratings for a large pool of potential emotion words. The set comprises 1,286 words, which were selected from various sources following the definition provided by Pavlenko (2008, page 148), namely: "words that directly refer to particular affective states or processes and that function to either describe or express them". Emotion prototypicality scores for those words were obtained by asking participants to rate to what extent a word refers to an emotion in a 5-points Likert scale, where 1 meant "This word does not refer to an emotion" and 5 meant "This word clearly refers to an emotion". In the Pérez-Sánchez et al.'s dataset, prototypicality ratings are distributed across the entire range of the scale, with 43% of the words classified as highly prototypical emotion words (i.e., they have a prototypicality score $\geqq 3$). Also, words in this dataset tend to show mid-to-high arousal scores and, according to valence ratings, there are more unpleasant than pleasant words. Pérez-Sánchez et al. also computed emotionality scores. Emotionality indicates the degree of affective charge of a word, regardless its positive-negative polarity. It is obtained by subtracting, in absolute value, the valence rating of each word from 5 (5 is the midpoint of the 9-points Likert scale used to collect valence scores). This index indicated that most words included in the dataset have mid-emotionality levels. Finally, regarding the other characteristics of the stimuli, the dataset is mostly composed of mid-frequency (i.e., scores in Zipf scale; Van Heuven et al., 2014) and late-acquired words (i.e., AoA was rated as the age at which a word was first learned, in a 1-to-11 scale, where 1 = lower than 2-years-old and 11 = equal or higher than 11-years-old). Thirty-five percent of the words are nouns, 43% are adjectives, and the remainder are verbs.

Procedure

We elaborated questionnaires to collect ratings for all 1,286 words for action, body expression, evaluation, feeling, interoception, morality, social interaction, and thought. Two previous studies (Díez-Álamo et al., 2019; Hinojosa et al., 2016) provided sensory experience ratings for a number of the words in our set, so it was only necessary to collect new ratings for this variable for 844 words. Apart from the critical words to be rated, 22 additional words were included, which were repeated across all the questionnaire versions for each variable. These words were considered as "controls" to assess the reliability of the ratings. For sensory experience, the control words were selected from previous studies (Díez-Álamo et al., 2019; Hinojosa et al., 2016). Finally, all questionnaires included seven calibrator words, which were presented at the beginning of each questionnaire, before the critical items. The purpose was to allow participants to see exemplars spanning the full range of the scale to be used for the ratings. Regarding sensory experience, we selected the calibrators from the mentioned studies, taking two words with low scores (i.e., *denominar* (to denominate) and *elemento* (element)), three words with medium scores (i.e., *cómico* (comical), *fortuna* (fortune), and *celebrar* (to celebrate)), and two words with high scores (i.e., *madre* (mother) and *verano* (summer)). With respect to the other variables, since there were no prior published data, we selected as calibrators a new set of words[2] that we estimated could cover the full range of the corresponding scale in the assessed variable.

The words in the dataset were randomly divided and included in questionnaires with 238-245 words each (including also controls and calibrators). There were four

[2] The calibrators for action, body expression, evaluation, feeling, morality were (in alphabetic order): *contexto* (context), *crítica* (critique), *honestidad* (honesty), *igualdad* (equality), *tamaño* (size), *tenis* (tennis), and *silencio* (silence). The calibrators for interoception were (same order): *acusación* (accusation), *anarquía* (anarchy), *decidir* (to decide), *destructivo* (destructive), *estresante* (stressing), *hambriento* (hungry), and *imaginación* (imagination). The calibrators for social interaction were (same order): *cantidad* (quantity), *capacidad* (capacity), *grupo* (group), *liderazgo* (leadership), *matrimonio* (marriage), *mensaje* (message), and *párrafo* (paragraph). The calibrators for thought were (same order): *conocimiento* (knowledge), *idea* (idea), *origen* (origin), *paisaje* (landscape), *preferencia* (preference), *tenis* (tennis), and *victima* (victim).

questionnaires for sensory experience and six questionnaires for the remaining variables. The order of presentation of words within a questionnaire was randomized individually for each participant, except the calibrator words, which were always presented first.

Participants accessed the questionnaires online through Qualtrics (http://www.qualtrics.com). They first read and accepted an informed consent form, then completed a few demographic questions, and finally received written instructions for the target variable of that questionnaire. A summary of the instructions and the labels of the scale for each variable are shown in Table 1 (note that the instructions and scales for the variables not collected in this study, but taken from Pérez-Sánchez et al., 2021, are also displayed in Table 1 because they were used in the analyses). The exact wordings in Spanish, as well as an English translation of the instructions are provided in https://osf.io/a2hcu. Words were displayed five at a time on each screen, with the rating scale below each word. After rating a set of five words, participants clicked on the 'Next' button to display another set of five words. The name of the variable to be rated (e.g., feeling) was displayed at the top and bottom of each screen. Participants could indicate that they did not know the word if that was the case ("*No conozco la palabra*"). Participants took about 20 minutes to complete a questionnaire.

Table 1. *Abbreviated instructions and rating scale for each variable.*

Variable	Instructions	Scale
Action	Rate whether the word is related to acting, doing, and influencing.	1 = fully disagree, 4 = mid-point, 7 = fully agree
Body expression	Rate whether the word is related to changes in parts of your body, such	1 = fully disagree, 4 = mid-point, 7 = fully agree

	as your voice, posture, gestures, or face expression.	
Evaluation	Rate whether the word is related to situations that are important for your well-being and/or your survival (either because they favor it or because they make it difficult).	1 = fully disagree, 4 = mid-point, 7 = fully agree
Feeling	Rate whether the word is related to feelings.	1 = fully disagree, 4 = mid-point, 7 = fully agree
Interoception	Rate whether the word is related to internal body sensations.	1 = fully disagree, 4 = mid-point, 7 = fully agree
Morality	Rate whether the word is related to morality, rules, or any other principle that guides your behavior.	1 = fully disagree, 4 = mid-point, 7 = fully agree
Sensory experience	Rate the degree of sensory experience (i.e., a real sensation of taste, touch, sight, sound, or smell) that you experience when reading the word.	1 = no sensory experience, 4 = moderate sensory experience, 7 = intense sensory experience
Social interaction	Rate whether the word is related to relationships between people.	1 = fully disagree, 4 = mid-point, 7 = fully agree
Thought	Rate whether the word is related to mental activity, ideas, opinions, and judgments.	1 = fully disagree, 4 = mid-point, 7 = fully agree

Valence *	Rate how you feel when reading the word.	1 = unhappy, 5 = mid-point, 9 = happy
Arousal *	Rate how you feel when reading the word.	1 = calm, 5 = mid-point, 9 = excited
Prototypicality *	Rate to what extent the word refers to an emotion.	1 = this word does not refer to an emotion, 5 = this word clearly refers to an emotion

* Ratings taken from Pérez-Sánchez et al. (2021).

Transparency and openness

We report how we determined our sample size, all data exclusions, and all measures in the study. All data and research materials are available at https://osf.io/a2hcu. This study's design and its analysis were not pre-registered.

Results

Data trimming

Each questionnaire was completed by at least 20 participants (M = 22, max = 24)[3]. Following the common procedure in this field (e.g., Pérez-Sánchez et al., 2021; Stadthagen-Gonzalez et al., 2018), individual ratings that correlated less than .10 with the average ratings of the participants in the same questionnaire were removed and replaced with ratings from a new participant who completed the respective questionnaire (1% of the collected data). Correlations close to zero are assumed to

[3] Ratings for 11 words were missed for one or more variables due to technical issues, so these words were rated in another data collection round with at least 30 responses for each word. The exact number of ratings collected for each word and variable can be found at https://osf.io/a2hcu. Data for these 11 words were not used in the reliability analyses.

reflect idiosyncratic response patterns, whereas negative correlations indicate that the participant understood the rating scale in the reverse order. We also checked that there were no questionnaires with the same rating score for more than 95% of the words. Most words were known by all the participants. There were, however, a subset of words that were not known by most raters: for 74 words (5.8% of the total set) we could only collect 10 valid responses (i.e., the other raters indicated that they did not know the word) in at least one variable (44 words in action, 55 in evaluation, 65 in body expression, 32 in interoception, 76 in feeling, 44 in morality, 27 in sensory experience, 53 in social interaction, and 40 in thought). Furthermore, 29 words (2.3% of the total set) yielded less than five valid responses in at least one variable (11 words in action, 10 in evaluation, 12 in body expression, 8 in interoception, 15 in feeling, 11 in morality, 5 in sensory experience, 6 in social interaction, and 9 in thought). Finally, three words were not known (and therefore not rated) by any participant in at least one variable: *paroxístico* (paroxysmal) in action and body expression, *contrito* (contrite) in evaluation, and *timorato* (timorous) in morality, so they were not included in any further analyses. The entire dataset, as well as the descriptive statistics and density plots for all the collected variables, are available at https://osf.io/a2hcu.

Reliability and validity of the ratings collected

We computed the standard error of the mean and the coefficient of variation of each variable as accuracy measures (see Table 2). The standard error of the mean ranged from 0.02 to 0.03 across variables, and the coefficient of variation ranged from 1.7% to 3.1%. These values indicate that the ratings had a good precision. We also explored the inter-rater reliability of the ratings by calculating the intraclass correlation coefficients (ICCs) for all the questionnaires using the *psych* package version 1.8.12 (function *ICC,* default arguments except for missing = F; Revelle, 2019) in R version

4.0.4 (R Core Team, 2021), via RStudio version 1.4.1106 (RStudio Team, 2021). Since there were several questionnaire versions for each variable, Table 2 shows the average ICC across questionnaires and some variability statistics for each variable.

Table 2. Accuracy and reliability statistics for the collected variables

Variable	***SDM***	***CV***	***M-ICC***	***SD-ICC***	**Min – Max ICC**
Action	0.02	1.6%	.69	0.06	.61 - .78
Body expression	0.02	1.8%	.69	0.06	.60 - .75
Evaluation	0.03	3.1%	.90	0.02	.86 - .91
Feeling	0.03	2.0%	.81	0.02	.79 - .83
Interoception	0.02	1.8%	.80	0.05	.74 - .88
Morality	0.03	2.7%	.83	0.05	.75 - .90
Sensory experience	0.02	1.9%	.72	0.05	.66 - .77
Social interaction	0.03	2.5%	.84	0.02	.81 - .86
Thought	0.02	1.7%	.71	0.03	.68 - .75

Note: SDM, standard deviation of the mean. CV, coefficient of variation. M-ICC, mean of the ICC values for all questionnaires of a variable. SD-ICC standard deviation for the ICC values for all questionnaires of a variable. Min-Max-ICC, maximum and minimum ICC values for all questionnaires of a variable. ICC values reported are those of two-way random absolute agreement for ICC(2, K).

We also assessed the reliability of our data through the comparison of the ratings of control words across questionnaires, by calculating the mean rating given by the participants to each control word in each questionnaire of a particular variable. Then, we correlated (Pearson's *r*) those means between pairs of questionnaires using JASP version 0.16.0.0 (JASP Team, 2021). All the resulting correlations were very high (Bonferroni adjusted alpha level of .005 per test (.05/9), with $p < .001$ in all correlations), where the lowest coefficients were $r(22) = .900$ for action, $r(22) = .824$ for body expression, $r(22) = .948$ for evaluation, $r(22) = .931$ for feeling, $r(22) = .654$ for interoception, $r(22) = .917$ for morality, $r(22) = .955$ for sensory experience, $r(22) = .880$ for social interaction, and $r(22) = .894$ for thought. In addition, we assessed the validity of our sensory experience ratings by comparing the ratings of the control words in our study to their ratings in previous studies. Significant and extremely high correlations were found for both comparisons (Bonferroni adjusted alpha level of .025 per test (.05/2), with all $p < .001$): $r(22) = .948$ and $r(22) = .946$ with the ratings from Díez-Álamo et al. (2019) and Hinojosa et al. (2016), respectively. These results validate the use of ratings from those studies together with the ones collected here in the data analyses.

Finally, considering the disproportionate number of women and men included in the study, it was important to examine whether the ratings generalize across genders. To this end, we carried out two types of analyses: (1) correlations between the mean ratings by women and by men to each word in each variable; and (2) a series of LMM analysis, in which we included gender as the only fixed-effect term. The bivariate correlations between the mean ratings given by men and by women were $r(1046) = .375$ for action, $r(1196) = .400$ for body expression, $r(1269) = .755$ for evaluation, $r(1009) = .398$ for feeling, $r(1272) = .463$ for interoception, $r(1239) = .564$ for morality, $r(816) = .327$ for

sensory experience, $r(1260) = .602$ for social interaction, and $r(1230) = .307$ for thought (Bonferroni adjusted alpha level of .005 per test (.05/9), with $p < .001$ in all correlations). Although all these correlations can be considered relatively high for a normative study (see Gignac & Szodorai, 2016), many of the mean ratings from men showed high standard error rates because they were obtained from very few individuals and therefore, these comparisons should be viewed with caution. As a more adequate and powerful analysis for unbalanced designs, we also compared the ratings from men and women with LMM analyses. The model for each collected variable was [ratings ~ gender + (1 I participants) + (1 I items)], where "ratings" refer to the scores for the corresponding variable (i.e., action, body expression, evaluation, feeling, interoception, morality, sensory experience, social interaction and thought). The models were conducted with the *lmer* function from lme4 package in R via RStudio. All models correctly converged and showed proper goodness-of-fit (the lowest ICC was .393). Gender was clearly non-significant in every one of the analyses, with *t*-values < -1.3 and *p* values > .22 in all cases, indicating that there are no significant differences between male and female ratings.

Correlations

Table 3 presents the matrix of partial correlations between each pair of variables controlling for the rest. Considering the strongest correlations, prototypicality was positively correlated with feeling, interoception, and thought. Regarding the relationship between the new variables assessed in this study, there were several low-to-moderate correlation coefficients. To name just a few, the correlation between variables mostly related to the physical component of emotions (i.e., interoception and body expression), although there were also significant correlations between variables related to the distinct

components (e.g., between feeling and interoception, or between body expression and action). Finally, some of the newly assessed variables were also correlated with the dimensions of valence and arousal. The most remarkable correlation was the one between valence and evaluation. Regarding arousal, the highest correlation was with action.

Table 3. Partial-correlation matrix of the analysed variables.

Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1. Prototypicality	1.000											
2. Action	-.076*	1.000										
3. Arousal	.044	.250**	1.000									
4. Body expression	-.034	.212**	.018	1.000								
5. Emotionality	-.001	-.095*	-.014	.018	1.000							
6. Evaluation	-.044	.049	-.068*	.092*	.212**	1.000						
7. Feeling	.408**	.075*	-.110**	.102**	.187**	.003	1.000					
8. Interoception	.340**	-.038	.017	.252**	-.070*	.039	.153**	1.000				
9. Morality	-.083*	.158**	-.028	-.036	-.073*	.369**	.190**	-.053	1.000			
10. Sensory experience	.058*	.064*	.216**	.281**	.266**	.004	.075*	.267**	-.053	1.000		
11. Social interaction	-.042	.143**	-.027	-.122**	.089*	.064*	.196**	-.060*	.153**	.144**	1.000	
12. Thought	.140**	.287**	.019	.018	.038	.072*	.199**	.028	.215**	-.179**	.049	1.000
13. Valence	-.039	-.053	-.086*	-.011	-.308**	.712**	-.043	-.028	.135**	.115**	.061**	-.184**

**Bonferroni adjusted alpha level $p < .0006$ (.05/78), *$p < .05$

Prediction of prototypicality

The contribution of each variable to prototypicality was examined through several multiple linear regression analyses using JASP (JASP Team, 2021). First, we performed the analysis introducing all the predictor variables (i.e., the 9 variables collected in this study as well as valence, arousal, and emotionality). The resulting model equation was significant, $F(7, 1270) = 148.49$, $p < .001$, with an R^2 of .583 (adjusted R^2 of .580), and with five predictor variables with significant standardized Beta-weights at $p < .05$, namely from the highest to lowest: feeling, interoception, morality, thought, and action. However, this model showed inadequate multicollinearity coefficients for evaluation (tolerance = .145 and variance inflation factor, VIF = 6.89) and valence (tolerance = .184 and VIF = 5.43), as both variables are highly correlated. Therefore, we performed a second analysis following the stepwise method (with entry at p-value < .05, and removal at p-value > .1 as steeping criteria), which reduces the regression model to those factors that slightly contribute to the explained variance and therefore may avoid excessive multicollinearity. Acceptable multicollinearity coefficients were obtained, with a minimum tolerance of .264 and a maximum VIF of 3.78 for morality. The resulting model equation was significant, $F(7, 1275) = 253.30$, $p < .001$, with an R^2 of .582 (adjusted R^2 of .579), and there were seven predictor variables with significant standardized Beta-weights (β) at $p < .05$. They were, from highest to lowest β: feeling, interoception, morality, thought, evaluation, action, and arousal (see Table 4).

Table 4. Linear regression analysis on prototypicality

Predictors	R^2 increment	β	t value
Interoception	.417	.363	15.643***

Feeling	.105	.494	17.752***
Evaluation	.047	-.119	-3.703***
Thought	.004	.123	-4.915***
Morality	.007	-.137	-3.899***
Action	.001	-.064	-2.552*
Arousal	.001	.041	2.033*

Note: Stepwise method. Variables are presented in the same order in which they were entered into the model. * $p < .05$, ** $p \leq .01$, *** $p \leq .001$

Semantic space of emotion words

To explore the inner structure underlying all the variables considered in the study and to find out the locus of emotional prototypicality, we conducted a series of additional analyses from a multidimensional perspective. First, we conducted an exploratory factor analysis (EFA) taking together the ratings of the 1,283 words on the 13 variables, using FACTOR (Lorenzo-Seva & Ferrando, 2006). A previous Parallel Analysis suggested that there were only two underlying factors. So, we fitted a bidimensional solution. Furthermore, as no preliminary information regarding the relation between the factors was available, an oblique solution was specified. The two-factor initial solution was fitted with ULS estimation and the rotation criterion was normalized direct oblimin. As a preliminary check, the Kaiser-Meyer-Olkin (KMO) index value was .846, which indicates that the correlation matrix is appropriate to carry out a factor analysis. The goodness of fit obtained with the bidimensional solution can be considered acceptable: the root mean square of the residuals (RMSR) was 0.062 and the Goodness of Fit Index estimate (GFI) was .984 (see McDonald & Mok, 1995). The

proportion of variance explained by the solution obtained was .599. The final rotated loading matrix (loadings below .30 are omitted), can be seen in Table 5.

Table 5. Rotated loading matrix resulting from the EFA

Variables	Factor 1	Factor 2
Evaluation	.938	
Valence	.886	
Morality	.872	
Social interaction	.578	
Action	.409	.455
Interoception		.768
Feeling		.762
Prototypicality		.760
Sensory experience		.679
Body expression		.639
Emotionality		.516
Thought		.499
Arousal		.342

The first rotated factor was defined by the indicators Evaluation, Valence, and Morality with very high loadings and, to a lesser extent by Social interaction and Action. The pattern on this factor exhibited a positive manifold (all the variables had a positive loading). This factor can be labelled “Socio-Emotional Polarity” because most of its indicators refer somehow to the positive/negative nature of the emotional experience or of the situation that elicits the emotion.

The second factor was defined by Interoception, Feeling, and Prototypicality as the main contributors, followed by Sensory Experience, Body expression, Emotionality, Thought, and Action. Arousal also contributed to this factor but with a modest loading. Again, the factor exhibited positive manifold. This factor may be labelled "Emotional Experience" because all the variables related to the different components of the emotional experience construct load significantly on it. It is noted that Action was the only factorially complex variable, having non-negligibly loadings in both factors. This result is probably because, apart from being part of the emotional response, the tendency to act (i.e., to approach or withdraw) is also related to the positive/negative nature of the situation.

To visualize how the words are distributed across the two factors, we used the words' standard factor score estimations on each factor as coordinates of a two-dimensional space (see Figure 1).

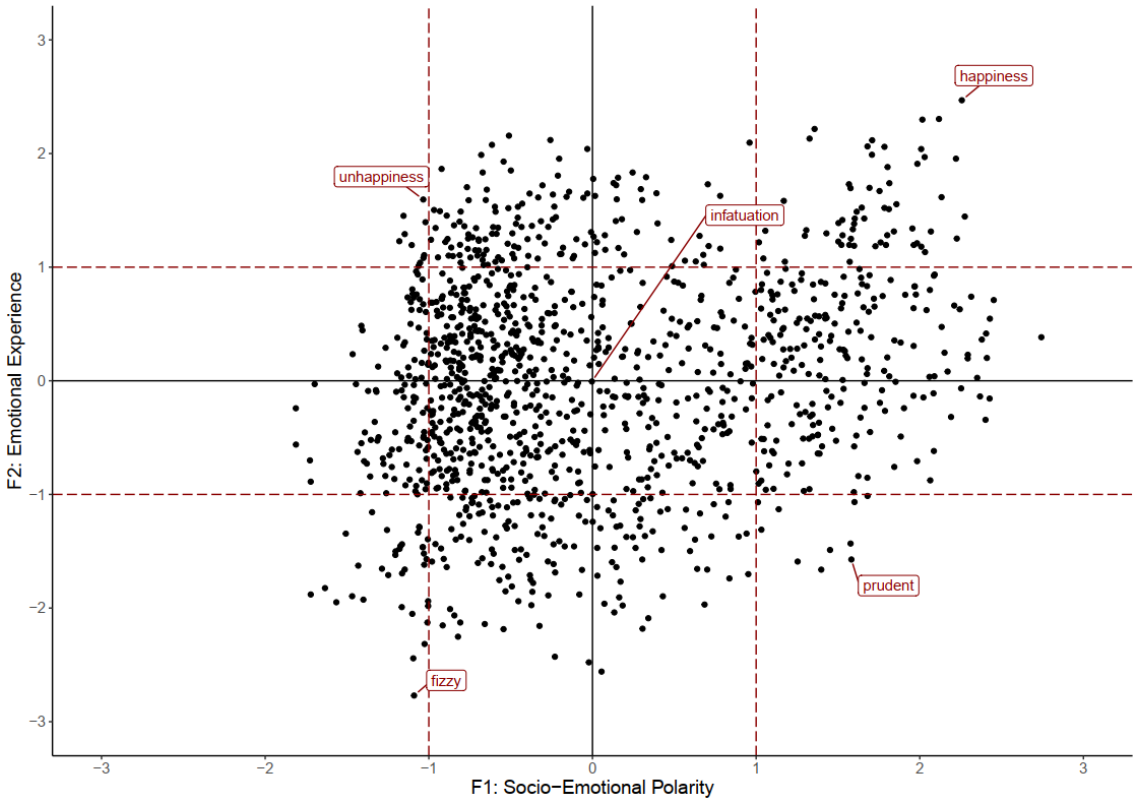

Figure 1. Scatterplot of the words in the space of the factor score estimates. Solid black lines around zero on both axes indicate the mean of the score estimates on each factor. Dashed red lines indicate one standard deviation thresholds above and below the mean on both axes.

Taking into consideration that these estimates are standard scores (with a mean value of zero and a standard deviation value of one), Figure 1 shows that the datapoints are distributed continuously around both axes, and that the overall shape tends to be circular. This continuous distribution seems to indicate that there are no clearly differentiable word clusters, and the circular shape of the scatterplot (as opposed to an elliptical shape) indicates that both factors are essentially independent. The estimated inter-factor correlation was .186, and was found to be statistically significant, but interpreted as an effect size estimate, this value would be classified as low, which implies that the two fitted dimensions are relatively independent of each other. The plot can also be considered as a semantic space of words, in which those words occupying central positions have low loadings (i.e., medium standard scores) in both factors (e.g., the word *encaprichamiento,* infatuation, which occupies the most central position). More interesting are the words that occupy the most extreme positions: those with high positive loadings in Emotional Experience and high negative loadings in Socio-Emotional Polarity (top left quadrant), high positive loadings in both factors (top right quadrant), high negative loadings in both factors (bottom left quadrant), and high positive loadings in Socio-Emotional Polarity and high negative loadings in Emotional Experience (bottom right quadrant).

To provide a more detailed analysis of the semantic space, we calculated the Euclidean distance of each point on the graph to the center point of the coordinate axes

and chose the most extreme word in each quadrant (i.e., the word that was farthest away in that quadrant). We then calculated the Euclidean distance between these extreme words and the other words in the space and identified the five closest words to each extreme word (see Table 6). Coordinates of each of the 1,283 words in this semantic space are available at https://osf.io/a2hcu.

Table 6. The most extreme words (in bold) and their five nearest words by quadrants of the semantic space

Top left quadrant			**Top right quadrant**		
Spanish	English	Distance	*Spanish*	English	Distance
infelicidad	**unhappiness**	-	***felicidad***	**happpiness**	-
enfurecido	enraged	0.11	*emoción*	emotion	0.22
furioso	furious	0.14	*feliz*	happy	0.30
pánico	panic	0.19	*alegre*	joyful	0.50
doloroso	painful	0.20	*amor*	love	0.52
humillado	humiliated	0.20	*amar*	to love	0.55
Bottom left quadrant			**Bottom right quadrant**		
Spanish	English	Distance	*Spanish*	English	Distance
efervescente	**fizzy**	-	***prudente***	**prudent**	-
arrojo	courage	0.33	*firmeza*	firmness	0.14
quieto	still	0.46	*paciente*	patient	0.15
antagónico	antagonistic	0.58	*aceptar*	to accept	0.20
hechizar	to bewitch	0.64	*audacia*	boldness	0.33
ultraje	outrage	0.65	*autosuficiente*	self-sufficient	0.51

Finally, the visual inspection of the points in Figure 1 suggests that the four quadrants are not equally populated. To better examine this, we represented the word distribution in a density plot (see Figure 2). The chart shows that many words are concentrated around moderate negative values in Socio-Emotional Polarity, and over a wide range of medium values in Emotional Experience, both positive and negative. There is also a second, smaller but notable group of words, with high positive scores in Socio-Emotional Polarity and positive intermediate scores in Emotional Experience.

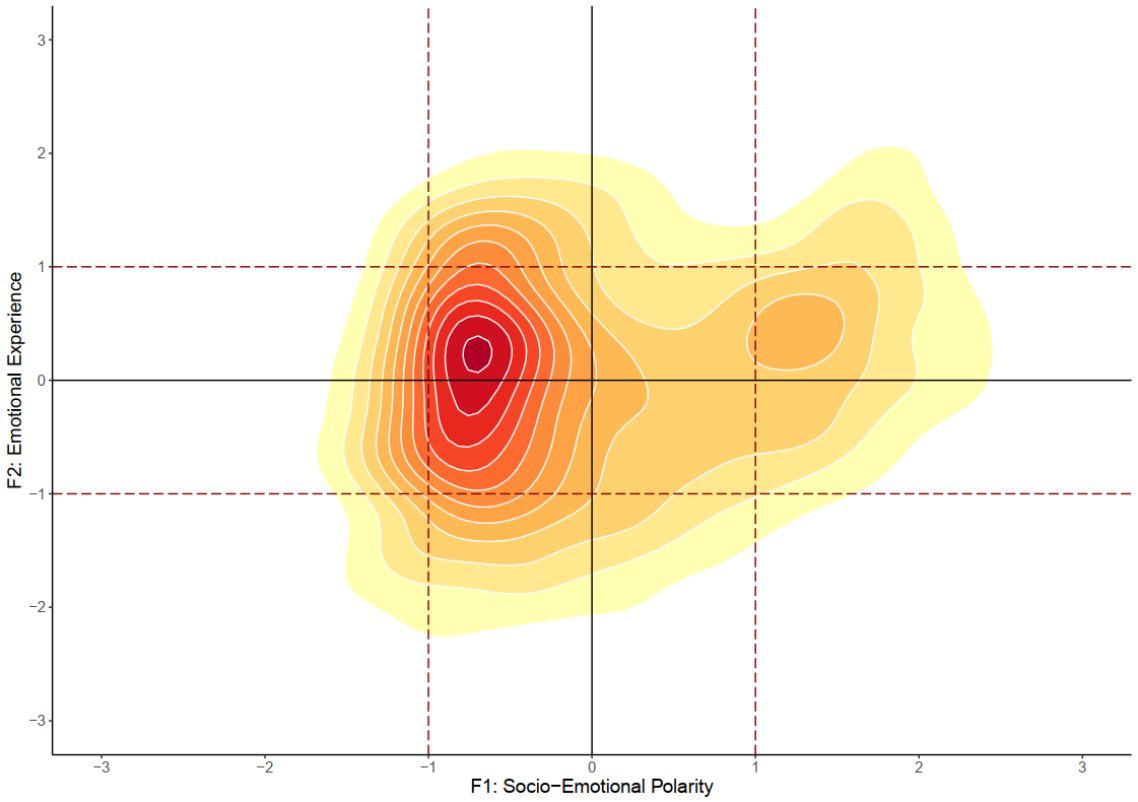

Figure 2. 2D Kernel Density plot depicting the density of the words in the semantic space. Solid black lines around zero on both axes indicate the mean of the scores on each factor. Dashed red lines indicate one standard deviation thresholds above and below the mean on both axes.

Discussion

The present study aimed to identify the crucial features that define emotion concepts, by examining the contribution of a number of variables to the emotion prototypicality of words. It also aimed to describe the organizational structure of those words in the lexicon. To that end, we collected ratings for several variables related to the different components of the emotional experience (e.g., Scherer & Moors, 2019, van Berkum, 2022): action, body expression, evaluation, feeling, and interoception, as well as for several variables related to the concrete/abstract distinction (Crutch et al., 2013; Troche et al., 2017): morality, sensory experience, social interaction and thought. Valence and arousal, as well as emotionality (i.e., the affective charge of a word, regardless of its positive or negative polarity), were also included in the analyses.

Regarding the first objective, there is consensus between the most influential theories that emotions are multi-componential phenomena, involving the following components: appraisal mechanisms, behavioral (body) expression, physiological reaction, action tendency and feeling (Sander et al., 2018, van Berkum, 2022). Previous studies have identified emotional charge and intensity as the most relevant variables to define emotion prototypicality (Niedenthal et al., 2004; Pérez-Sánchez et al., 2021; Zammuner, 1998). In these studies, participants were asked to rate the words based on their subjective experience, but the nature of the information participants relied on to make their judgments remained elusive. They may had relied on a general feeling (i.e., the particular subjective experience) or they may had paid more attention to other aspects, like bodily sensations. The only way to know which component is most relevant for a word to be considered an emotion term is to ask participants to focus on each component separately, as we did in this study. Our results are clear-cut: both the correlational and the regression analyses revealed that feeling and interoception are the most relevant components. That is, those words which are more associated with feelings

and with internal body sensations are considered by the speakers to be the best representatives of the category "emotion". Good examples of this relationship are the words *alegría* (happiness) and *tristeza* (sadness), which are among the most prototypical words in the dataset, and which have high scores in the interoception and feeling variables. The role of the other components is clearly minor; that is, words which have high scores in evaluation (e.g., *respeto*, respect), action (e.g., *impulso*, impulse) and body expression (e.g., *pereza*, laziness) are not necessarily perceived as good representatives of the category "emotion" by the speakers.

The term *feeling* refers to the phenomenological experience of the individual during an emotion episode. This experiential quality is different from that associated to other, non-affective, phenomenological states (e.g., hunger); it differs between at least some types of emotions and can be instantiated in varying degrees of intensities (see Reisenzein & Döring, 2009, for a historical overview of this concept). *Feeling* is conceptualized in most theories as a combination of valence and arousal (e.g., Russell & Barrett, 2009); that is, the phenomenological experience of emotion involves a pleasantness-unpleasantness quality and a certain degree of activation. Consequently, it may be argued that this dimension might be reducible to these two variables. Our results suggest that this is not the case. First, we did not find a significant correlation between feeling and valence. Second, the correlations between feeling and arousal (r = -.11) and between feeling and emotionality (r = .19), which are the equivalents of the variables identified as relevant in previous prototypicality studies (i.e., intensity and emotional charge; Niedenthal et al., 2004; Pérez-Sánchez et al., 2021; Zammuner, 1998), were small and medium correlations, respectively, from a normative perspective (Gignac & Szodorai, 2016). Surely, other emotion components contribute to subjective feelings

(see, for instance, Sander et al., 2018), although their exact contribution cannot be known from the data collected here. Future studies can shed light on this issue.

The second most important predictor of emotion prototypicality was *interoception*, indicating that bodily sensations are fundamental to the representation of emotion concepts. Several studies have reported a major role of interoceptive and somatosensory processes in the generation of the emotional experience (e.g., Kreibig, 2010; Nummenmaa et al., 2014). Such role is acknowledged in some of the most influential models of emotional processing (e.g., Barrett et al., 2007; Damasio & Carvalho, 2013; Lindquist, 2013). For instance, constructionist theories posit that it is the combination of the information about the body's internal state with other kinds of information, like information about the current context and the conceptual knowledge about emotion categories, what provokes the emergence of an emotion (see MacCormack & Lindquist, 2017, for a review).

In fact, the capacity of people to associate body sensations to emotions has been evidenced in a series of studies conducted recently by Nummenmaa and co-workers (Nummenmaa et al., 2014, 2018, Volynets et al., 2020). The authors have developed an innovative approach to elaborate bodily maps of emotions, based on a topographical self-report method: Participants are presented with human silhouettes and are asked to color the areas where they feel that the activation increases or decreases when they are presented with emotional stimuli (i.e., words, facial expressions, movies, and stories). These studies have reported a distinctive pattern of body sensations for each of the investigated emotions. They have also found that the body sensations associated with different emotions are highly similar across genders and across a wide range of cultures (Volynets et al., 2020). In fact, other authors have proposed that interoception may be a more adequate measure of the universality of emotions than other components, such as

facial expressions. The reason is that facial expressions are often visible to others, which provides opportunities for learning cultural rules. In addition, facial expressions can be used in some cases for affiliative purposes, rather than for communicating the emotional state (Manninen et al., 2017). By contrast, bodily states and the corresponding subjective bodily sensations are less likely to be influenced by culture, mostly because they cannot be observed (and copied) from others (Volynets et al., 2020).

The afore-mentioned work highlights the relevance of body sensations in the emotional experience. Convergent evidence has been found in studies about interoceptive ability, that is, the ability of individuals to recognize body signals. Individual differences in this ability have been related with differences in the intensity of the emotional experience (Pollatos et al., 2007; see also Garfinkel et al., 2015, and MacCormack & Lindquist, 2017, for an overview). In addition, some disorders associated with difficulties in identifying and describing others and own emotions, like autism spectrum disorders, have been linked to abnormal interoceptive processing (Garfinkel et al., 2016). These findings lend support to the idea that bodily sensations shape the emotional experience.

Interestingly, from a different line of research, interoception was also identified as the most relevant variable in distinguishing emotion words (e.g., shame) from non-emotion words (e.g., legality) in a study about the perceptual grounding of abstract concepts (Connell et al., 2018). Regarding emotion words, ratings for interoception were higher than for the other sensory modalities assessed (hearing, taste, touch, smell and vision), suggesting that interoception is the most important perceptual modality in the emotion experience. The pattern was the opposite for concrete words. These results agree with the idea that emotion words represent a particular set of words that differs

from concrete as well as from purely (non-emotion) abstract words (Altarriba & Bauer, 2004; Borghi et al., 2018; Mazzuca et al., 2017). Considering the preponderance of abstract words in our dataset, our findings align with those of Connell et al. (2018), suggesting that interoception is a crucial variable to distinguish emotion words from non-emotion abstract words. Morality and thought, two variables related to the concreteness/abstractness distinction (Troche et al., 2014), also contributed to some extent to emotion prototypicality, which suggests that it is more probable that an abstract word refers to an emotion when it has to do with mental activities, judgments, and moral rules.

Apart from identifying the variables that contribute the most to emotion prototypicality, the analyses performed here also provided relevant information about the relationship between the new variables included in this study. There were significant correlations between some of the variables related to the different components of the emotional experience (evaluation, action, body expression, feeling, and interoception). However, the correlation coefficients were not very high, and not all components were related to each other. Just as an example, although interoception and body expression were related, and both were related to feeling, only body expression was related to action, and none of the above variables was related to evaluation. Therefore, high ratings in one component do not necessarily imply high ratings in another component. For instance, the word *miedo* (fear) has high ratings in interoception and feeling, but not so high in action and body expression. By contrast, the word *rebeldía* (rebelliousness) has a high score in action, but not so high in interoception.

Regarding the relationship with valence and arousal (and emotionality, computed from valence), the most remarkable correlation was found between evaluation and valence. Evaluation was also correlated with emotionality. This means that words

referring to situations that people consider as most important for their well-being are those more emotionally (positively) charged. Two clear examples of this relationship are the words *amor* (love) and *felicidad* (happiness), which are among those with the highest ratings in the three variables mentioned (i.e., valence, arousal and emotionality). With respect to arousal, the highest correlation was observed with action, which means that the words related to acting, like *pasión* (passion), also produce a high activation.

Finally, concerning the variables related to the concreteness/abstractness distinction (morality, sensory experience, social interaction, and thought), the correlation coefficients were also moderate. The variable most correlated with morality was evaluation, which indicates that words related to situations that are important for people's well-being are also related to rules that guide behavior (e.g., *respeto*, respect, and *felicidad*, happiness). The highest correlation for sensory experience was the one obtained for body expression, indicating that words related to changes in parts of the human body, like *deseo* (desire), also evoke a high degree of sensory experience (i.e., taste, touch, vision, etc.). Regarding social interaction, the highest correlation was with feeling, suggesting that words highly related to feelings are also highly related to relationships between people (e.g., *afecto*, affection). Finally, the relationship between thought and action indicates that when a word is related to mental activities it also tends to be related to acting (e.g., *empatía,* empathy).

The second aim of this study was to examine the organizational structure of emotion words. Two factors emerged from the factorial analysis. We called the first factor Socio-Emotional Polarity, because most of the variables with high loadings (evaluation, valence, morality, social interaction, and action) refer somehow to the positive/negative nature of the emotional experience or the situation that provokes the emotion. Valence is in fact a direct estimation of emotional polarity. Furthermore,

words denoting positive emotions, like *felicidad* (happiness), are frequently associated with situations that involve social relationships and with acting according to moral rules. In addition, they tend to trigger action (approach tendencies). Of note, emotional prototypicality does not have a relevant load in this factor but in the second factor, which we called Emotional Experience because it is mostly composed of variables related to the components of the emotional experience (interoception, feeling, prototypicality, sensory experience, body expression, emotionality and thought). Arousal and action also had a modest load in this factor, probably indicating that intense affective experiences (e.g., *odiar*, hate) are associated with a high degree of activation and with a strong tendency to act.

The scores for each word in both factors were used to create a two-dimensional semantic space. The most extreme words and their nearest words are good exemplars of the type of words included in each quadrant. Words in the top left quadrant of the space (e.g., unhappiness) have high ratings on emotional prototypicality, interoception, and feeling. They also score high on some other components of emotion. Congruently, they are perceived as emotionally charged by the raters, being mostly negative words. The main difference with respect to the words in the top right quadrant (e.g., happiness) is their polarity, which in this case is positive. Words located at the bottom quadrants of the space have low prototypicality ratings. That is, they are not perceived by the speakers as good exemplars of the category "emotion" and therefore have low scores in the variables related to emotion components. The main difference is that words located in the right quadrant (i.e., prudent) have high scores for variables like valence, evaluation, and morality, being perceived as positive words by the speakers (although not denoting emotions). By contrast, words located in the left quadrant (e.g., vibrant)

have low scores in those variables, so some of them are perceived as negative words (e.g., *ultraje*, outrage), and others as devoid of emotional charge (e.g., *quieto*, quiet).

The density plot provides additional information regarding the distribution of the words in the semantic space. Two aspects are worth mentioning: first, there are more negative words than positive words. This result agrees with previous studies showing a preponderance of negative words in the emotion lexicon (e.g., Schrauf & Sánchez, 2004; Zammuner, 1998). A possible explanation is the different kinds of information processing associated with negative and positive events. Negative emotional experiences signal that there is a problem with potential consequences for survival, so they trigger a detailed cognitive analysis of the experience to find a solution. This processing results in more differentiated labels for those experiences. By contrast, positive emotional experiences signal that the environment is safe, so there is no need for a detailed information analysis and so they trigger a more general, less differentiated, cognitive processing (Schwarz, 1990). These differences would yield a more varied set of labels for negative emotions than for positive emotions. In line with this idea, it has been shown that people perceive positive stimuli as more similar than negative stimuli (Koch et al., 2016). A second interesting piece of information provided by the density plot is that, while the ratings for negative words are more distributed in the emotional experience dimension, positive words seem to be more concentrated on the positive side of such dimension. This indicates that, although there are less positive words than negative words in the dataset, the former are more homogeneous regarding their emotional charge (i.e., ratings in the variables related to the different components of emotion) than the latter.

This is not the first study examining the organizational structure of emotion terms, but most previous studies have relied on similarity judgments between a limited

set of words and have applied cluster analyses to see how they are grouped together (Alonso-Arbiol et al., 2006; Galati et al., 2008; Shaver et al., 1987, 2001). All those studies have identified valence as the main organizational variable. A few studies have used a procedure similar to ours, relying on participants' ratings in a number of variables with the aim of identifying the underlying dimensions with a principal component analysis (of note, these studies also included a small set of words). With such approach, Fontaine et al. (2007) concluded that four dimensions were needed to satisfactorily define the semantic space of emotion words: valence, power (i.e., control over the situation/stimulus), arousal and novelty. Similarly, Tamir et al. (2016) asked participants to classify a set of mental states in different variables and reported three dimensions: valence; rationality, which reflects whether a state is more cognitive or more emotional; and social impact, which reflects whether a state is highly socially directed or not. This last dimension combined the concepts of arousal and sociality, which are positively associated. This proposal, which was called the 3d Mind model, has been validated with behavioral and neural data (see Thornton & Tamir, 2020).

In conclusion, the good (positive)-bad (negative) distinction appears consistently in all the studies, suggesting that valence is the most decisive dimension in the characterization of emotion terms. Emotion words are grouped according to the positive/negative experience they denote (i.e., words denoting positive experiences are closer in the space and, consequently, perceived as more similar to each other than to words denoting negative experiences). Our first factor, "socio-emotional polarity", corresponds to this dimension. There is less agreement with regards to the number and the identity of the other dimensions reported. The second factor identified in our study, "emotional experience", may be considered to share some aspects with the "rationality" dimension identified by Tamir et al. (2016), because both basically distinguish

emotional from non-emotional states. Our main contribution is to have identified the variables with a major role in such distinction, which are interoception and feeling.

Finally, in this study we also present normative data for a large set of potential emotion words. The words assessed in this study were the same as those in Pérez-Sánchez et al. (2021), who included ratings for prototypicality as well as for other relevant affective variables, such as valence, arousal, emotionality, and for the relationship of the words with five discrete dimensions (i.e., happiness, sadness, fear, anger and disgust). Ratings for some lexico-semantic variables were also included. In particular, concreteness, lexical frequency, and age of acquisition. The present study complements that work by providing, for the first time, ratings for several affective variables related to the components of emotion, as well as for some variables related to the concreteness/abstractness distinction. In combination, both studies provide values for 21 relevant variables for 1,286 words. To our knowledge, this is the most comprehensive set of potential emotion terms (i.e., words denoting emotions) in the literature. Indeed, several studies conducted from the prototypicality approach have provided prototypicality as well as intensity and valence ratings for smaller sets of words (e.g., 153 words in Zammuner, 1998; 237 in Niedenthal et al., 2004). On the other hand, normative affective studies commonly include arousal and valence ratings for large sets of words (Guasch et al., 2016; Hinojosa et al., 2016; Stadthagen-Gonzalez et al., 2017; Warriner et al., 2013). However, most stimuli in these datasets are neutral words and, among the affectively charged words, there is not commonly a distinction between emotion words and emotion laden words (i.e., words that do not refer to an emotion but can provoke it, like *party* or *death*). Other normative studies, focused on the distinction between concrete and abstract words, have included ratings for some of the dimensions used here, but the number of words related to emotions was very low

there. For instance, only 21% of the words in the study of Troche et al. (2017) had scores equal or above the middle point of the emotionality scale. Therefore, the present dataset may be of great interest to researchers studying the interplay between language and emotion. However, a note of caution is needed: our sample was not balanced by gender and consisted of many more women than men. In addition, most participants were university students. Although the analyses performed revealed no gender effects in ratings, future studies involving more representative samples are highly desirable to further validate the present data.

To conclude, this multidimensional ratings study has pointed out the relevance of interoception and feeling in the experience of emotion. Words with high scores in these dimensions are perceived as the best representatives of the "emotion" category. These words are close in the semantic space elaborated here, where they are organized in terms of their emotional polarity. This study has theoretical and practical implications, by contributing to our understanding of emotion concepts. This knowledge can be used in the study of emotion processing, language processing, emotion knowledge and emotion perception in clinical as well as in healthy populations.

Acknowledgements

This study was supported by the Ministerio de Ciencia e Innovación (PID2019-107206GB-I00 /AEI/10.13039/501100011033) of Spain, the Government of the Comunidad de Madrid (H2019/HUM-5705), the Xunta de Galicia (ED431B 2022/19) and the Universitat Rovira i Virgili (2022PFR-URV-47).

References

Alonso-Arbiol, I., Shaver, P. R., Fraley, R. C., Oronoz, B., Unzurrunzaga, E., & Urizar, R. (2006). Structure of the Basque emotion lexicon. *Cognition and Emotion, 20*, 836-865. https://doi.org/10.1080/02699930500405469

Altarriba, J., & Bauer, L. M. (2004). The Distinctiveness of Emotion Concepts: A Comparison between Emotion, Abstract, and Concrete Words. *The American Journal of Psychology, 117*, 389-410. https://doi.org/10.2307/4149007

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*, 327–332. https://doi.org/10.1016/j.tics.2007.06.003

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences,* 22, 637-660. https://doi.org/10.1017/S0140525X99532147

Borghi, A. M., Barca, L., Binkofski, F., & Tummolini, L. (2018). Varieties of abstract concepts: development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*, 20170121. https://doi.org/10.1098/rstb.2017.0121

Borghi, A. M., Fini, C., & Mazzuca, C. (2022). Abstract Concepts, Social Interaction, and Beliefs. *Frontiers in Psychology, 13*:919808. https://doi.org/10.3389/fpsyg.2022.919808

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. In R. D. Lane & L. Nadel (Eds.), *Series in affective science. Cognitive neuroscience of emotion* (p. 242-276). Oxford University Press.

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behavior, 2,* 581–591. https://10.1038/s41562-018-0376-6

Connell, L., Lynott, D., & Banks, B. (2018). Interoception: The forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*, 20170143. https://doi.org/10.1098/rstb.2017.0143

Crutch, S. J., Troche, J., Reilly, J., & Ridgway, G. R. (2013). Abstract conceptual feature ratings: the role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Frontiers in Human Neuroscience, 7*:186. https://doi.org/10.3389/fnhum.2013.00186

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: Evolutionary and neurobiological origins. *Nature Reviews Neuroscience, 14*, 143-152. https://doi.org/10.1038/nrn3403

Desai, R. H., Reilly, M., & van Dam, W. (2018). The multifaceted abstract brain. *Philosophical Transactions Royal Society B, 373*, 1752. https://doi.org/10.1098/rstb.2017.0122

Díez-Álamo, A. M., Díez, E., Wojcik, D. Z., Alonso, M. A., & Fernandez, A. (2019). Sensory experience ratings for 5,500 Spanish words. *Behavior research methods*, *51*, 1205-1215. https://doi.org/10.3758/s13428-018-1057-0

Ekman, P. (2018). How emotions might work. In Fox, A., Lapate, R., Shackman, A. & Davidson, R. (Eds.), *The nature of emotion: Fundamental questions*, (XXVII-XXXV). Oxford University Press.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General, 113*, 464-486. https://doi.org/10.1037/0096-3445.113.3.464

Ferré, P., Guasch, M., Stadthagen-Gonzalez, H., Hinojosa, J. A., Fraga, I., Marín, J., & Pérez-Sánchez, M. A. (2023, July 2). What makes a word a good representative of the category of "emotion"? The role of feelings and interoception. Retrieved from https://osf.io/a2hcu

Fingerhut, J., & Prinz, J. J. (2018). Grounding evaluative concepts. *Philosophical Transactions Royal Society B, 373,* 20170142. https://doi.org/10.1098/rstb.2017.0142

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18*, 1050-1057. https://doi.org/10.1111/j.1467-9280.2007.02024.x

Galati, D., Sini, B., Tinti, C., & Testa, S. (2008). The lexicon of emotion in the neo-Latin languages. *Social Science Information, 47*, 205-220. https://doi.org/10.1177/0539018408089079

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology, 104,* 65-74. https://10.1016/j.biopsycho.2014.11.004

Garfinkel, S. N, Tiley, C., O'Keeffe, S., Harrison, N. A, Seth, A. K., & Critchley, H. D (2016). Discrepancies between dimensions of interoception in autism: implications for emotion and anxiety. *Biological Psychology, 114*, 117-126. https://10.1016/j.biopsycho.2015.12.003

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and individual differences, 102*, 74-78. https://doi.org/10.1016/j.paid.2016.06.069

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods, 48*, 1358-1369. https://doi.org/10.3758/s13428-015-0684-y

Haro, J., Calvillo, R., Poch, C., Hinojosa, J. A., & Ferré, P. (2022). Your words went straight to my heart: the role of emotional prototypicality in the recognition of emotion-label words. *Psychological Research. Advance online publication.* https://doi.org/10.1007/s00426-022-01723-6

Hinojosa, J. A., Rincón-Pérez, I., Romero-Ferreiro, M. V., Martínez-García, N., Villalba-García, C., Montoro, P. R., & Pozo, M. A. (2016). The Madrid Affective Database for Spanish (MADS): Ratings of Dominance, Familiarity, Subjective Age of Acquisition and Sensory Experience. *PLoS ONE, 11,* e0155866. https://doi.org/10.1371/journal.pone.0155866

JASP Team (2021). *JASP* (Version 0.16.0.0) [Computer software]. https://jasp-stats.org/

Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42 (8),* 1171-1192. https://doi.org/10.1037/xlm0000243

Kousta, S.-T., Vigliocco, G., Vinson, D.P., Andrews, G., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140 (1),* 14-34. https://doi.org/10.1037/a0021446

Kreibig, S. D. (2010). Autonomic Nervous System Activity in Emotion: A Review. *Biological Psychology, 84*, 394-421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Li, B. J., Bailenson, J. N., Pines, A., Greenleaf, W. J., & Williams, L. M. (2017). A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.02116

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review, 5(4)*, 356–368. https://doi.org/10.1177/1754073913489750

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion: The experience of fear as a conceptual act. *Psychological Science, 19,* 898–903. https://doi.org/10.1111/j.1467-9280.2008.02174.x

Lorenzo-Seva, U. & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior research methods, 38*, 88-91. https://doi.org/10.3758/BF03192753

MacCormack, J. K., & Lindquist, K. A. (2017). Bodily Contributions to Emotion: Schachter's Legacy for a Psychological Constructionist View on Emotion. *Emotion Review, 9(1)*, 36-45. https://doi.org/10.1177/1754073916639664

Manninen, S., Tuominen, L., Dunbar, R. I., Karjalainen, T., Hirvonen, J., Arponen, E., Hari, R., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2017). Social Laughter Triggers Endogenous Opioid Release in Humans. *The Journal of Neuroscience, 37,* 6125-6131. https://doi.org/10.1523/JNEUROSCI.0688-16.2017

McDonald, R. P., & Mok, M. M. C. (1995). Goodness of fit in item response models. *Multivariate Behavioral Research*, *30*, 23-40. https://doi.org/10.1207/s15327906mbr3001_2

Mazzuca, C., Barca, L., & Borghi, A. M. (2017). The Particularity of Emotional Words: A Grounded Approach. *Rivista Internazionale di Filosofia e Psicologia, 8*, 124-133. https://doi.org/10.4453/rifp.2017.0010

Niedenthal, P. M., Auxiette, C., Nugier, A., Dalle, N., Bonin, P., & Fayol, M. (2004). A prototype analysis of the French category "émotion". *Cognition and Emotion, 18*, 289-312. https://doi.org/10.1080/02699930341000086

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences, 111,* 646-651. https://doi.org/10.1073/pnas.1321664111

Nummenmaa, L., Hari, R., Hietanen, J. K., & Glerean, E. (2018). Maps of subjective feelings. *Proceedings of the National Academy of Sciences, 115,* 9198-9203. https://doi.org/10.1073/pnas.1807390115

Ortony, A., Clore, G. L., & Foss, M. A. (1987). The referential structure of the affective lexicon. *Cognitive Science, 11*, 341-364. https://doi.org/10.1016/S0364-0213(87)80010-1

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition, 11*, 147-164. https://doi.org/10.1017/S1366728908003283

Pérez-Sánchez, M. Á., Stadthagen-Gonzalez, H., Guasch, M., Hinojosa, J. A., Fraga, I., Marín, J., & Ferré, P. (2021). EmoPro–Emotional prototypicality for 1286 Spanish words: Relationships with affective and psycholinguistic variables. *Behavior Research Methods, 53*, 1857-1875. https://doi.org/10.3758/s13428-020-01519-9

Pollatos, O., Traut-Mattausch, E., Schroeder, H., & Schandry, R. (2007). Interoceptive awareness mediates the relationship between anxiety and the intensity of unpleasant feelings. *Journal of Anxiety Disorders, 21*(7), 931-943. https://doi.org/10.1016/j.janxdis.2006.12.004

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Reisenzein, R., & Döring, S. A. (2009). Ten Perspectives on Emotional Experience: Introduction to the Special Issue. *Emotion Review*, 1(3), 195–205. https://doi.org/10.1177/1754073909103587

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 1.8.12, https://CRAN.R-project.org/package=psych

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

RStudio Team (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA. http://www.rstudio.com

Russell, J. A. & Barrett, L. F. (2009). Core affect. In David Sander & Klaus R. Scherer (eds.), *The Oxford companion to emotion and the affective sciences*, 104. Oxford: Oxford University Press.

Sander, D., Grandjean, D., & Scherer, K. R. (2018). An Appraisal-Driven Componential Approach to the Emotional Brain. *Emotion Review, 10(3),* 219–231. https://doi.org/10.1177/1754073918765653

Scherer , K. R.(2022) Theory convergence in emotion science is timely and realistic, *Cognition & Emotion, 36(2*), 154-170. https://doi.org/10.1080/02699931.2021.1973378

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology, 70*, 719-745. https://doi.org/10.1146/annurev-psych-122216-011854

Schrauf, R. W., & Sánchez, J. (2004). The preponderance of negative emotion words in the emotion lexicon: A cross-generational and cross-linguistic study. *Journal of Multilingual and Multicultural Development, 25(2-3),* 266-284. https://doi.org/10.1080/01434630408666532

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 527-561). New York: Guilford Press.

Shaver, P. R., Murdaya, U., & Fraley, R. C. (2001). Structure of the Indonesian emotion lexicon. *Asian Journal of Social Psychology, 4*, 201-224. https://doi.org/10.1111/1467-839X.00086

Shaver, P. R., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*, 1061-1086. https://doi.org/10.1037//0022-3514.52.6.1061

Stadthagen-González, H., Ferré, P., Pérez-Sánchez, M. A., Imbault, C., & Hinojosa, J. A. (2018). Norms for 10,491 Spanish words for five discrete emotions: Happiness, disgust, anger, fear, and sadness. *Behavior Research Methods, 50, 1943-1952.* https://doi.org/10.3758/s13428-017-0962-y

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods, 49*, 111-123. https://doi.org/10.3758/s13428-015-0700-2

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences, 113*(1), 194-199. https://doi.org/10.1073/pnas.1511905112

Thornton, M.A., & Tamir, D.I. (2020). People represent mental states in terms of rationality, social impact, and valence: Validating the 3d Mind Model. *Cortex, 125,* 44-59. https://doi.org/10.1016/j.cortex.2019.12.012

Troche, J. (2017). Towards a unified model of semantic memory: validation and theoretical implications of the conceptual feature rating space. *Language, Cognition, and Neuroscience, 33*, 1-12. https://doi.org/10.1080/23273798.2017.1408852

Troche, J., Crutch, S., & Reilly, J. (2014). Clustering, hierarchical organization, and the topography of abstract and concrete nouns. *Frontiers in Psychology, 5*:360. https://doi.org/10.3389/fpsyg.2014.00360

Troche, J., Crutch, S. J., & Reilly, J. (2017). Defining a Conceptual Topography of Word Concreteness: Clustering Properties of Emotion, Sensation, and Magnitude among 750 English Words. *Frontiers in Psychology, 8*:1787. https://doi.org/10.3389/fpsyg.2017.01787

van Berkum, J. J. A. (2022). A survey of emotion theories and their relevance to language research. In Gesine Lenore Schiewer, Jeanette Altarriba, & Bee Chin Ng. *Language and Emotion. An international Handbook*. Walter de Gruyter. https://doi.org/10.1515/9783110347524-001

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly*

Journal of Experimental Psychology, 67(6), 1176-1190. https://doi.org/10.1080/17470218.2013.850521

Volynets, S., Glerean, E., Hietanen, J. K., Hari, R., & Nummenmaa, L. (2020). Bodily maps of emotions are culturally universal. *Emotion, 20,* 1127-1136. https://doi.org/10.1037/emo0000624

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 1191-1207. https://doi.org/10.3758/s13428-012-0314-x

Xu, X., Li, J., & Chen, H. (2022). Valence and arousal ratings for 11,310 simplified Chinese words. *Behavior Research Methods, 54*(1), 26-41. https://doi.org/10.3758/s13428-021-01607-4

Zammuner, V. L. (1998). Concepts of emotion:"Emotionness", and dimensional ratings of Italian emotion words. *Cognition and Emotion, 12*, 243-272. https://doi.org/10.1080/026999398379745

Zou, G. Y. (2012). Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Statistics in medicine, 31*(29), 3972-3981. https://doi.org/10.1002/sim.5466