Redundancy, Isomorphism and Propagative Mechanisms between Emotional and Amodal Representations of Words: A Computational Study

Memory & Cognition

https://doi.org/10.3758/s13421-020-01086-6

José Á. Martínez-Huertas (josea.martinez@uam.es). Universidad Autónoma de Madrid

Guillermo Jorge-Botana* (gdejorge@psi.uned.es). Universidad Nacional de Educación a Distancia

José María Luzón (jmluzon@psi.uned.es). Universidad Nacional de Educación a Distancia

Ricardo Olmos (ricardo.olmos@uam.es). Universidad Autónoma de Madrid

*** Corresponding author:**

Guillermo Jorge-Botana

Dpt. of Psicología Evolutiva y de la Educación
Facultad de Psicología
Calle de Juan del Rosal, 10
Universidad Nacional de Educación a Distancia
28049-Madrid
Email address: gdejorge@psi.uned.es
Telephone number: +34 914978450

Martínez-Huertas, J.A., Jorge-Botana, G., Luzón, J.M., & Olmos, R. (2020). Redundancy, Isomorphism and Propagative Mechanisms between Emotional and Amodal Representations of Words: A Computational Study. *Memory & Cognition*. https://doi.org/10.3758/s13421-020-01086-6.

ABSTRACT

Some proposals claim that language acts as a link to propagate emotional and other modal information. Thus, there is an eminently amodal path of emotional propagation in the mental lexicon. Following these proposals, we present a computational model that emulates a linking mechanism (mapping function) between emotional and amodal representations of words using vector space models, emotional feature-based models, and neural networks. We analyzed three central concepts within the embodiment debate (redundancy, isomorphism, and propagative mechanisms) comparing two alternative hypotheses: *semantic neighborhood hypothesis* vs. *specific dimensionality hypothesis*. Univariate and multivariate neural networks were trained for dimensional (N=11,357) and discrete emotions (N=2,266), and later we analyzed its predictions in a test set (N=4,167 and N=875, respectively). We showed how this computational model could propagate emotional responses to words without a direct emotional experience via amodal propagation, but no direct relations were found between emotional rates and amodal distances. Thereby, we found that there were clear redundancy and propagative mechanisms, but no isomorphism should be assumed. Results suggested that it was necessary to establish complex links to go beyond amodal distances of vector spaces. In this way, although the emotional rates of semantic neighborhoods could predict the emotional rates of target words, the mapping function of specific amodal features seemed to simulate emotional responses better. Thus, both hypotheses would not be mutually exclusive. We also showed that discrete emotions could have simpler relations between modal and amodal representations than dimensional emotions. All these results and their theoretical implications are discussed.

Keywords: neural networks; emotional words; vector space models; mental lexicon; grounded cognition

INTRODUCTION

Language is a sophisticated symbolic tool acquired throughout thousands of years of evolution. As modern alphabet symbols, language has no direct analogic relation with the real referents (with some exceptions like onomatopoeias and calligrams). Language representations of reality encode several kinds of information such as propositional and conceptual, but also encodes sensorimotor or emotional information (Louwerse, 2011, 2018). According to Louwerse (2011, 2018), amodal symbols can inform not only about conceptual constructions (such as relations with other amodal symbols in propositional frames), but also about sensorimotor or emotional information extracted from language statistics (e.g., “pain” shares an amodal frame with “blood” by means of common amodal contexts). In computational studies, the amodal format of language representation is usually emulated by vector space models (see Günther, Rinaldi, & Marelli, 2019; Jones, Gruenenfelder, & Recchia, 2018; or Jorge-Botana, Olmos, & Luzón, 2020, for a recent review on vector-space models) as the source of these models is purely linguistic (usually, cooccurrences in paragraphs, sentences, or textual windows). These paragraphs, sentences, or textual windows are processed in order to generate a semantic space where words are spatially represented by vectors. That vector space is useful to extract a distance metric that represents semantic relations within words. In this vector space, words with similar meanings would compose a semantic neighborhood, which is a group of closely related words in this amodal vector space (these words are usually called *semantic neighbors* because they share spatial positions or coordinates in the semantic space).

In contrast, modal representations can be conceived as the first stages of perception that encode and simulate sensorimotor and emotional information in a purer state (e.g., Barsalou, 2008; Barsalou, Santos, Simmons & Wilson, 2008). Emotional information also encodes sympathetic nervous system responses and motor responses, for example, avoidance/escape preparation or facial responses in some motor areas (Barrett, 2006). For this reason, authors like Havas, Glenberg and Rinck (2007) found emotion-sentence compatibility effects and they explained their findings as a result of emotional simulation processes. In empirical studies, the modal format of these representations is usually emulated by feature-based models (Andrews, Frank & Vigliocco, 2014; Riordan & Jones, 2011) in which people are asked to identify features of words like shapes, movement, size, etc. In the case of emotional information, these feature-based models can be generated using data sets that contain emotional features such as happiness, anger, sadness, fear, disgust, or affective features such as valence or arousal (e.g., Fraga et al., 2018; Stevenson, Mikels, & James, 2007).

As can be suspected after reading the above description, sensorimotor and emotional information are encoded twice in modal and amodal representations. This *redundancy* in the organization of the representations of the cognitive system has been observed modeling vector space models and human feature-based studies jointly. Riordan and Jones (2011) studied such redundancy comparing six different corpus-based models (amodal representations) with three feature-based models (modal representations). They found that the predictions of amodal models were comparable to human judgements within feature-based models in different clustering tasks. This redundancy has been formally conceptualized in different theoretical frameworks like the *language and situated simulation theory* (Barsalou, et al., 2008), but specially and explicitly in the *symbol interdependency hypothesis* (Louwerse, 2011, 2018). According to Barsalou's and Louwerse's models, word processing also relies on sensorimotor and emotional information. Both models identify what kind of representation (modal or amodal) is required to give an efficient answer and the different timings and tasks dependencies of each one.

The *language and situated simulation theory* assumes that conceptual processing is the result of the interaction between linguistic processing (amodal representations) and a situated simulation (modal representations) by means of statistical underpinnings (Barsalou et al., 2008). For example, processing “the yellow bus is near” would generate an interacting mirroring between simulations in perception for the yellow bus and language and propositional interpretations. But this theory proposes that language activation peaks before that perceptual simulation activation (for empirical evidence see Ianì, Foiadelli, & Bucciarelli, 2019). That is, the system recruits amodal representations before modal ones. These amodal representations, usually identified as linguistic inferences, can be used for tasks that only need some superficial sensorimotor and emotional information, but they are very useful in reasoning (e.g., conceptual tasks with “bus”, “near”, etc., as generic concepts). In a second phase, when the language activation peaks and its related inferences are generated, a perceptual (sensorimotor/emotional) simulation would start to take place in interaction with those inferences. These modal simulations can be used for tasks in which deeper sensorimotor and emotional information are needed (e.g., “bus” as a situated shape that comes fast to me).

The *symbol interdependency hypothesis* shares part of Barsalou´s theoretical perspective but emphasizes language employing the encoding of emotional, perceptual, or sensorimotor information in it. Louwerse (2011, 2018) proposes, as we explained at the beginning of this section, that sensorimotor and emotional information are also encoded in amodal representations of language. For this reason, Louwerse’s proposal genuinely claims for redundancy of sensorimotor information. When amodal representations interact in a “closed world” (like words in a language), they also reach sensorimotor and emotional meanings through linguistic inferences. While Barsalou´s model proposes two kinds of modal and amodal representations to univocally represent sensorimotor and conceptual information, Louwerse’s proposal explicitly claims that this relation is not univocal because amodal representations encode both conceptual and sensorimotor information. Therefore, sensorimotor features could be judged in a time-efficient way just with amodal representations without recruiting modal representations. Modal representations would be recruited only for featured-oriented tasks.

The proposals that claim redundancy have been supported by some empirical evidence from behavioral and brain studies (Louwerse, Hutchinson, Tillman, & Recchia, 2015; Louwerse & Jeuniaux, 2010; Louwerse & Hutchinson, 2012; Günther, Dudschig & Kaup, 2018) and computational studies (Bestgen & Vincze, 2012; Kuhlmann, Hofmann & Jacobs, 2017; Louwerse & Benesh, 2012; Louwerse & Zwaan, 2009; Riordan & Jones, 2011; Recchia & Louwerse, 2015).

But beyond this redundancy, there is another important concept: *Isomorphism*. If modal and amodal representations were isomorphic, the relations within emotional rates in modal representations of words would be equivalent to their relations in amodal representations. In other words, if a perfect isomorphism exists, it would be possible to exhaustively predict the emotional valence of a word from its amodal representation. And even more, the emotional features of a word could be predicted by the emotional features of its semantic neighborhood because the semantic neighbors of a word would share similar amodal features (positions or coordinates in the amodal space). Some computational studies have demonstrated that such predictions are reasonable. For example, amodal distances between words in vector space models predict emotional properties accurately (Bestgen & Vincze, 2012; Kuhlmann et al., 2017; Hofmann, et al, 2018; Recchia & Louwerse, 2015). In the present study, we analyzed this proposal and we called it the *semantic neighborhood hypothesis*.

Nonetheless, some studies suggested that modal and amodal representations do not need to be purely isomorphic because they could distribute information differently with some kind of bias. For example, Riordan and Jones (2011) showed that sensorimotor information that is represented in modal and amodal representations is biased. They found redundant information in both representation formats, but modal representations (emulated by feature-based models) were biased to perceptive properties like touch or smell, while amodal representations (emulated with vector space models) were biased to function, action, and situation relations. Some studies have also found that only a part of amodal representations from vector space models is important to predict the emotional features of words (Hollis, Westbury, & Lefsrud, 2017). This implicitly nuances the *semantic neighborhood hypothesis* and its implicit isomorphism. Tentatively, it could suggest that a link between both formats of representations is needed to recruit modal representations by the amodal ones, as suggested by some studies. In fact, a biological structure for such a link has been suggested: the supramodal hub at the anterior temporal lobe may interface amodal representations with sensorimotor machinery, especially if sensorimotor features are demanded by retrieval tasks (for a review, see Nastase & Haxby, 2016). This proposal is in accordance with studies that showed that, in some circumstances, modal representations are activated in the presence of words (Binder, Westbury, McKiernan, Possing & Medler, 2005; Borghi et al., 2017; Hauk, Johnsrude & Pulvermüller, 2004; Meteyard, Rodríguez-Cuadrado, Bahrami & Vigliocco, 2012; Zwaan, 2014, 2016; Zwaan & Yaxley, 2003). Functionally, this link could be considered as a mapping function similar to transformations of visual or sound signal representations on abstract conceptual representations as suggested by some authors (Gärdenfors, 1996; Gärdenfors, 2000; Balkenius & Gärdenfors, 2016). In the present study, we analyzed this mapping function and we called it the *specific dimensionality hypothesis*.

Another research question about the emotional properties of words is how grounding processing affects words with direct emotional experience and words without direct emotional experience. Emotional features can be propagated to words without any direct emotional experience as it has been suggested by some studies on verbal propagation of emotional properties in language acquisition like Field & Schorah, (2007), García-Palacios et al. (2018), or Grégoire & Greening (2020). However, it is unclear if we need to expose all words to emotional information to produce adequate emotional responses. Recently, Hoffman, McClelland and Lambon-Ralph (2018) have formalized the concept of "*acquired embodiment*". They proposed a propagation mechanism of sensorimotor and emotional features to words that have not been previously grounded. These authors implemented a small model based on neural networks, a model that is partially symbolic and partially sensorimotor and emotional, and they achieved this sensorimotor and emotional propagation. The formalization of *acquired embodiment* acts as a bridge between theories that proposed that symbols need to be grounded in the physical experience (e.g. Glenberg, 1997) and theories that proposed that knowledge is eminently grounded in a linguistic format (e.g., Landauer & Dumais, 1997).

The Present Study

We presented three central concepts within the embodiment debate: redundancy, isomorphism, and propagative mechanisms. These concepts guided our hypothesis, our design, and our analyses. Specifically, in the present study, we analyzed the pertinence of a link between emotional and amodal representations of words in a large-scale normative data set of words. We modeled said link by means of a neural network that was trained to predict emotional judgements using amodal vector representations of some words in a vector space model. Then, we studied the underlying redundancy, isomorphism and propagative mechanisms that produce emotional responses using amodal representations of words without direct emotional experience.

The whole computational model is designed to learn and take advantage of the redundancy of modal and amodal representations of words following the theoretical framework of the *symbol interdependency hypothesis* (Louwerse, 2011, 2018) and the *language and situated simulation theory* (Barsalou, et al., 2008). We tested the underlying mechanism of the redundancy and the isomorphism of modal and amodal representations of words, comparing the performance of the *semantic neighborhood hypothesis* vs. the *specific dimensionality hypothesis*. For the *semantic neighborhood hypothesis*, we analyzed the capacity of the amodal distances in the vector space model to predict the emotional judgements of words. For the *specific dimensionality hypothesis*, we tested the relative importance of amodal features of the vector space model to predict the emotionality of words. In this case, the superiority of the *specific dimensionality hypothesis* would mean that there is not a perfect isomorphism and, probably, a link is needed to join amodal and modal representations when the task is biased to emotional features. This hypothetical superiority of the *specific dimensionality hypothesis* would not invalidate the *semantic neighborhood hypothesis* if the performance of semantic neighborhood predictions remains reliable. Thus, the *specific dimensionality hypothesis* and the *semantic neighborhood hypothesis* would not be mutually exclusive.

Moreover, inspired by Hoffman et al. (2018), we also studied the underlying mechanism that produces the propagation of emotions within amodal representations of words in a large-scale normative data set. In our computational model, some of the amodal representations of words were grounded by means of the neural network model training. In this training, some dimensions gain relevance and act as a mapping between emotional and amodal representations of words. Then, we tested the underlying propagative mechanisms in a second group of words without direct emotional experience. If the neural network model generates valid emotional predictions in this second group of words, there would be evidence of a mechanism that would propagate the effect of the link via amodal representations. That is, in addition to the encoding of emotional features in amodal representations that arise through the interaction of words in a language, pure amodal representations could also need a linking mechanism to modal simulations. This is close to verbal conditioning of emotions (Field & Schorah, 2007; García-Palacios et al., 2018; Grégoire & Greening, 2020).

Given that the present study is focused on emotional representations of words, we explored the potential differences between dimensional and discrete emotional categories within this link for modal and amodal representations of words. The dimensional categories, also known as affection, are conceptualized as a reduced number of subjacent dimensions like valence (unpleasant-pleasant) and arousal (calming-arousing) that are transversal to all emotions (Russell, 1979). The discrete categories, also known as basic emotions, are conceptualized as a limited number of discrete emotions that have specific characteristics, physiological and behavioral correlates like happiness, anger, sadness, fear and disgust (Ekman, 1992). This distinction is very useful for researchers of emotion (e.g., Fraga et al., 2018; Stevenson et al., 2007), but the potential differences between affection and emotion within emotional and amodal relations remain unexplored.

METHOD

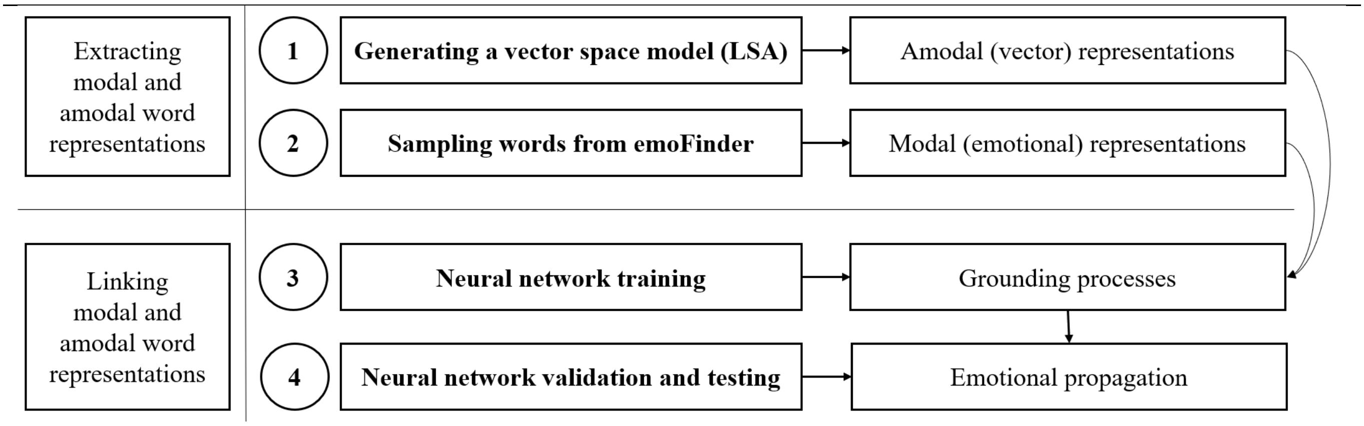
In order to test the assumptions of redundancy, isomorphism and propagation mechanisms, we generated a computational model by means of a neural network architecture to emulate a link between emotional and amodal representations. It proposes an explanatory mechanism to emotional simulations that goes beyond the amodal codification of emotional features (even in words with no modal experiences through propagative processes). The procedure of this study can be summarized in four consecutive phases (see *Figure 1*). Firstly, a general domain corpus was processed to extract amodal word vector representations for the input of the neural network models. We used the vector space model *Latent Semantic Analysis* (LSA; Landauer & Dumais, 1997) to generate vector representations as amodal representations of words (see Günther et al., 2019; Jones et al., 2018; or Jorge-Botana et al., 2020, for a recent review on vector space models).

Secondly, the emotional ratings of words were sampled from different data sets of emoFinder (Fraga et al., 2018), which is the most suitable source of emotional normative data sets for Spanish words (the language used in this study) to obtain modal (emotional) representations of words. A subset of these emotional ratings was used to set the emotional grounding and represent the output for the training phase of the neural network models. Specifically, we used both dimensional (valence, arousal) and discrete (happiness, anger, sadness, fear, disgust) emotional categories for the modal representation of words.

Thirdly, the neural network architecture was trained with both amodal and emotional representations. Here, some LSA amodal word representations were grounded by means of the emotional ratings of emoFinder. That is, we modeled a set of grounding events that joined the amodal representations with the emotional ones. In this training phase, the neural network architecture uses the representation of words derived from the space model (LSA[1]) as input, and the emotional ratings from emoFinder (emotional experience) as output.

Fourthly, the neural network architecture was validated and tested using independent samples of words (i.e., validation and testing data sets) to test the propagation of the effect of the link within purely amodal representations of words (with no modal experience). It is worth mentioning that this validity strategy not only uses an independent sample of words, but also different normative data sets.

Figure 1. Procedure of the present study: Extraction of modal and amodal word representations. Generation and testing of a link between modal and amodal word representations by means of a neural network architecture.



The details of the procedure are described below. *Gallito Studio* software (Jorge-Botana, Olmos, & Barroso, 2013) was used to generate the latent semantic space and to train and validate neural network models[2]. Then, neural network models were tested

[1] LSA provides a formal representation of meaning in orthogonal dimensions/vectors (i.e., its latent dimensions correlate zero). In this way, orthogonal vectors from LSA can be considered an ideal input due to its dimensionality reduction (Mandl, 1999).

[2] This software uses the *encog* library from https://www.heatonresearch.com/encog/ to train the neural network models.

using *GallitoAPI* (e.g., www.gallitoapi.net), which is an API that exposes the functionalities of *Gallito Studio* online. Both the API requests and the statistical analyses were performed using R software (Spanish word lists, R code, and instructions are available at: https://osf.io/m8wux/; DOI:10.17605/OSF.IO/M8WUX).

Amodal word vector representations

A textual corpus was processed with the LSA technique in order to use its vectors as the amodal input for the different neural network models. 455,969 documents (paragraphs) and 70,244 unique terms from a random sample of the Spanish Wikipedia were used as the corpus to generate the semantic space. It was generated using standard procedures (Landauer, McNamara, Dennis, & Kintsch, 2007). After *Singular Value Decomposition* (SVD), 300 dimensions were imposed for the semantic space. The log-entropy weighted function was used as the preprocessing (Nakov, Popova, & Mateev, 2001).

Selecting a training set of words for neural network training

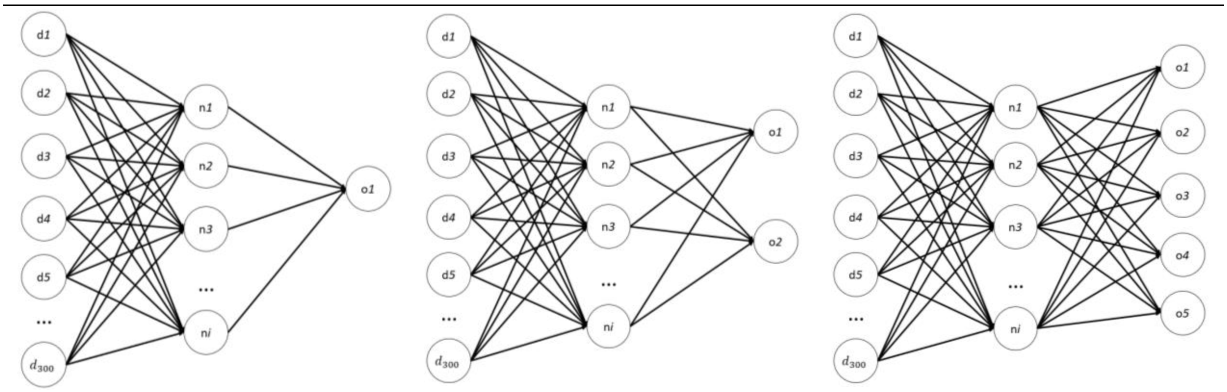
A training set of words was selected for neural network training. The larger normative data sets of emoFinder (Fraga et al., 2018) were selected to train the neural network models in both dimensional (Ferré et al., 2017) and discrete (Stadthagen-González et al., 2017) emotional categories. Words of the training data set were filtered by deleting the common words between the training and the test sets to ensure that words from the test set were not trained at all. *Table 1* shows the data sets and the number of words used in each step of the neural network models. Emotional values of each word were standardized on a 0-1 scale to make all data sets equivalent and to have a suitable output for logistic functions of neural networks.

Training and validating neural network models

Different neural network models were trained for each of the emotional categories considered in the present study. 11,357 words were used to train the neural network for the dimensional categories (valence, arousal) and 2,266 words were used for the discrete ones (happiness, anger, sadness, fear, disgust) (see *Table 1*). Different versions of neural networks were trained to learn and to link in the training set. Namely: univariate and multivariate models (see *Figure 2*). These neural networks use RPROP+ (resilient propagation with a bias node) with one hidden layer in order to estimate the regression weights of the model. A logistic transformation in each node is used to produce a response.

Figure 2. Graphical representation of univariate (2.A) and multivariate (2.B and 2.C) neural networks.

A) Univariate neural networks	B) Multivariate neural networks for dimensional categories	C) Multivariate neural networks for discrete categories



Note. d_k = From 1 to 300 LSA dimensions as the input. n_i = From *1* to *i* hidden nodes (see *Table 2*). o_1-o_5 = Different affective/emotional predictions as the output.

While a univariate neural network was generated for each emotional category (with only one modal scalar prediction for each feature as output), the multivariate one jointly propagates different categories (which generates a modal vector of values as output). In this training phase, different iterative processes based on backpropagation modified the weights that interrelated LSA amodal word vector representations with human emotional judgements (acting as modal representations). The data set was divided into two parts: a sample of 85% of the words were used as the training set and 15% of the words were used as the validation set. This was done to select the most appropriate number of hidden nodes in each affective/emotional category (see *Results* section), but this 15% validation data set is also the first evidence of emotional propagation.

Testing neural network models

To analyze the propagation of emotional responses, the performance of neural network models was tested using words from new normative data sets (different words, different human evaluators, and different researchers). 4,167 words were used to test the dimensional categories (valence, arousal) and 875 words were used to test the discrete ones (happiness, anger, sadness, fear, disgust) (see *Table 1*). 93 words of dimensional categories and 25 words of discrete ones were excluded from the analyses because they did not have a vector representation in the LSA. As was pointed out before, neural network models estimated modal values (a modal vector) for the amodal representation of each word. Then different statistical analyses were performed to test these network scores using the human scores from the data set as the dependent variable. Once the univariate and the multivariate neural network models were tested, additional analyses on the network scores were performed.

Table 1. Specification of the normative data sets used in each step of the neural network estimations.

Variable	Dimensional emotions		Discrete emotions				
	Valence	Arousal	Happiness	Anger	Sadness	Fear	Disgust
Training and validation data set	Stadthagen-González et al. 2017 (N=11,357)		Ferré et al. 2017 (N=2,266)				
Test data set	Redondo et al. 2005 (N=478) Redondo et al. 2007 (N=1,034) Ferré et al. 2012 (N=380) Guasch et al. 2016 (N=1,400) Hinojosa et al. 2016 (N=875)		Hinojosa et al. 2016 (N=875)				

Note: N = Number of selected words in each data set.

Analyzing the semantic neighborhood hypothesis

In order to test the *semantic neighborhood hypothesis*, we estimated the emotional scores of the words from test data sets (i.e., target words) by means of their semantic neighbors in the amodal vector space (see similar procedures in Bestgen & Vincze, 2012; Kuhlmann et al., 2017; Hofmann, et al, 2018; Recchia & Louwerse, 2015). That is, we tested the capacity of amodal semantic distances to predict emotional properties of words. We computed the cosine measure to determine all the results of the *semantic neighborhood hypothesis*, in other words, to select the semantic neighbors and to compute the amodal distances. These target word scores were computed as the mean of the human emotional rates of its 30 closest semantic neighbors. For each target word, we extracted the 30 closest semantic neighbors and then retrieved the human rate of each semantic neighbor from the normative data set (each target word contained from 2 to 26 semantic neighbors with valid values). We used a multiple linear regression comparing the mean of the human rates of semantic neighbors and the neural network scores to identify whether the neural networks explained something that goes beyond and above amodal neighbors. Moreover, we also analyzed the relation between emotional ratings and amodal distances (cosines) of semantic neighbors of target words through the mean Pearson correlation coefficients in the sample of words.

Analyzing the specific dimensionality hypothesis

In order to test the *specific dimensionality hypothesis*, we obtained different evidence of the relative importance of the LSA amodal dimensions in the neural network architecture to predict the emotionality of words. First, we analyzed the relative importance of the LSA amodal dimensions in the neural network architecture employing Olden (Olden, Joy, & Death, 2004) and Garson (Garson, 1991; Goh, 1995) algorithms using R's *NeuralNetTools* package (Beck, 2018). As expected, the absolute values of Olden algorithm's results were equivalent to Garson's ones. Thus, we reported here the results of Olden algorithm to represent the relative importance of the LSA amodal dimensions. We also tested the *specific dimensionality hypothesis* using backward stepwise linear regressions using the AIC index replicating the procedure of Hollis et al. (2017). As a complement, we compared the performance of the scores of neural networks and backward stepwise linear regressions using a test for dependent correlations (Hittner, May, & Silver, 2003) and estimating the exact confidence interval for those differences (Zou, 2007) using R's *cocor* package (Diedenhofen & Musch, 2015).

Analyzing the overlapping between the dimensions of the link and the amodal dimensions of a tentative core of emotions

The *specific dimensionality hypothesis* assumes that only some relevant dimensions are used to bypass emotional responses from amodal representations. As it was stated above, the Olden algorithm (Olden et al., 2004) is a way to identify such relevant LSA amodal dimensions. On the contrary, the *semantic neighborhood hypothesis* requires the generation of semantic neighborhoods and thus assumes that all the dimensions are relevant in the process. But there is room for differential participation of dimensions even in the amodal vector space. The emotional bias of tasks could cause some amodal dimensions to be more activated to generate judgments and neighbors. This is the case where the amodal emotional concept of words is reached by activating or inhibiting different parts of their amodal representations. For example, the emotional meaning of "bank" can be reached amodally by promoting part of its amodal representation, such as the one that is related to "crisis" and "unemployment" or "fishing" and "family". Therefore, we isolate those amodal dimensions that are potentially coherent with the amodal concept of each emotion to compare them with the dimensions that participate in the link emulated by the neural network models.

We computed the importance of the LSA dimensions in the amodal vector space by means of two complementary methods. The first is simple. We identified the words from the data set with high human rates (at least, more than one standard deviation above the mean) in each emotion. In each emotional pool, we randomly generated 100 sets of six words, such as {"fear", "terror", "terrorism", "monster", "horror", "bank"}, and computed the vector sum of each, such as:

$$v_{centroid} = v_{fear} + v_{terror} + v_{terrorism} + v_{monster} + v_{horror} + v_{bank}$$

The result was a vector that represented the centroid of the six words (Landauer & Dumais, 1997). As words had a high emotional score in each specific emotion, it was inferred that the resulting vector dimensions would preserve the amodal representation of that emotion. We computed the mean of these 100 resulting vectors in each amodal dimension. Finally, we had a unique vector that represented the dimensions of an amodal concept of the emotion. We called this method centroid.

The second method is more sophisticated, but it is well known and has been widely applied in the literature. One of the algorithms that take into account the context of the words is the *Construction-Integration* (C-I) algorithm for vector spaces (Kintsch, 1998; Kintsch, 2000; Kintsch, 2001; Kintsch & Bowles, 2002; Millis & Larson, 2008; Jorge-Botana, Olmos, & Barroso, 2012). The C-I algorithm has inhibition and activation mechanisms that promote context-agree components and inhibit context-disagree ones (see Kintsch & Welsch, 1991 for details of the original conception). This mechanism can magnify those dimensions that are consistent with the amodal concept of emotion.

For example, the set of words {“fear”, “terror”, “terrorism”, “monster”, “horror”, “bank”} activates a part of the amodal representation and inhibits another part. As each word of the set has a high emotional score and acts as a context for the other words, then, for example, only the emotionalized meaning of “bank” would be activated. The C-I works as follows: First, in the construction phase, a net is constructed with the first *n* neighbors of each word in the set. Second, in the integration phase, inhibition/activation mechanisms are run until a stable state is reached. Third, the vectors of the *k* most activated words from that net are summed. The resulting vector would be the amodal representation of the emotion. For the parameters *n* and *k*, we used n={15, 30, 60, 90, 120} and k={6, 12}. We called this method the C-I method.

The dimensions with the highest scores (in absolute values) of the resulting vectors are those dimensions that participate in a potentially biased selection of the amodal context of emotions in the amodal vector space. In this way, we can compare those dimensions that participate in the neural network model using the Olden algorithm (Olden et al., 2004) with those LSA amodal dimensions that define the context of specific emotions using both the centroids and the C-I method resulting vectors. If the relative importance of both groups of dimensions does not coincide, it would support the necessity of a mechanism that acts as a link to produce emotional simulations. We analyzed the Kendall rank correlation coefficients to avoid potential metric bias when analyzing the relations between the relative importance of both measures.

RESULTS

Aligned with the objectives of this study, results have the following structure: (1) We tested the performance of univariate and multivariate neural network models in the validation and the test sets to show the grounding and propagative mechanisms of the link between emotional and amodal representations; (2) We present the analyses that were performed around the *semantic neighborhood hypothesis* to test the capacity of amodal semantic distances to predict emotional properties of words; (3) We present the analyses that were performed around the *specific dimensionality hypothesis* to test the relative importance of the LSA amodal dimensions to predict the emotionality of words; (4) We present the analyses that were performed to compare the *semantic neighborhood hypothesis* and the *specific dimensionality hypothesis*.

1. Validating and testing neural network emotional predictions

As it was mentioned above, neural network models were trained using 11,357 words for dimensional categories and 2,266 words for discrete ones as the training data set. 15% of these words (1,704 and 340 words, respectively) were randomly selected as a validation data set (thus, they were not included in the training). Results were robust (i.e., the number of hidden nodes did not show a great impact on the performance) and presented a considerable propagation to new words in terms of its similarity with normative data sets (see *Table 2*). We selected the number of hidden nodes with higher performance in the validation data set for the following analyses.

Table 2. Performance of the neural network models in the validation data set (Pearson correlation coefficients between predicted scores from neural networks and human rates)

	Hidden nodes	Dimensional emotions*		Discrete emotions**				
		Valence	Arousal	Happiness	Anger	Sadness	Fear	Disgust
Univariate	5	.519	.504	.626	.641	.579	.597	.537
	Epochs	#502	#70	#116	#38	#56	#67	#41
	10	.548	.522	.647	.626	.611	.636	.525
	Epochs	#164	#54	#70	#37	#39	#65	#43
	15	.568	.521	.616	.616	.618	.646	.516
	Epochs	#137	#51	#63	#36	#45	#55	#45
	20	.559	.534	.643	.631	.590	.587	.533
	Epochs	#121	#50	#65	#35	#39	#54	#40
	25	.562	.512	.655	.599	.602	.626	.476
	Epochs	#137	#44	#81	#39	#39	#55	#44
	30	.562	.514	.691	.610	.582	.635	.518
	Epochs	#121	#60	#74	#38	#45	#58	#49
	Hidden nodes	Dimensional emotions*		Discrete emotions**				
		Valence	Arousal	Happiness	Anger	Sadness	Fear	Disgust
Multivariate	10	.538	.498	.592	.558	.551	.631	.536
	Epochs	#221				#125		
	20	.567	.474	.620	.605	.573	.611	.496
	Epochs	#116				#87		
	30	.591	.504	.632	.613	.588	.636	.514
	Epochs	#139				#74		
	40	.574	.510	.673	.633	.603	.647	.547
	Epochs	#130				#68		
	50	.572	.521	.680	.593	.564	.614	.514
	Epochs	#127				#65		
	60	.576	.511	.650	.577	.584	.611	.515
	Epochs	#98				#73		

Note: All Pearson correlation coefficients were statistically significant (p<.01). Grey cells show the highest Pearson correlation coefficients. # = number of epochs per training. * = 1,704 words were used as the validation set. ** = 340 words were used as the validation set.

The following results were obtained in the test data set (different words, different human evaluators and different researchers). Specifically, dimensional emotions were tested using 4,074 words, while discrete emotions were tested using 850 words. The performance of neural network models was robust and coherent with the previous results (see *Table 3*). Results showed the propagation of the emotionality via amodal representations, that is, the capacity to accurately predict the emotional properties of the words of the test set in both univariate and multivariate neural network models.

Table 3. Performance of the neural network models in the test set (Pearson correlation coefficients).

		Dimensional emotions		Discrete emotions				
		Valence	Arousal	Happiness	Anger	Sadness	Fear	Disgust
Univariate	Hidden nodes	15	20	30	5	15	15	5
	Performance	.702	.609	.646	.702	.636	.692	.590
Multivariate	Hidden nodes	40		40				
	Performance	.700	.598	.645	.676	.626	.709	.562

Note: All Pearson correlation coefficients were statistically significant (p<.01).

2. *Analyzing the semantic neighborhood hypothesis*

Different studies showed that the surrounding semantic neighbors of a target word in the amodal vector space were good proxies to predict human scores of this target word (e.g., Bestgen & Vincze, 2012; Hollis et al., 2017; Kuhlmann et al., 2017; Lenci et al., 2018; Recchia & Louwerse, 2015; Turney & Littman, 2003). In this text, we have called this the *semantic neighborhood hypothesis* because implicitly it is assumed that the amodal representation as a whole is enough to capture the emotionality of a word and, consequently, all semantic dimensions would have the same relevance to predict emotionality. To put to the test the prediction based on this assumption, 30 semantic neighbors were selected for each target word based on the cosine in the amodal vector space. Then, the mean of the human emotional rates of the semantic neighbors was used to predict the emotional score of the target words of the test data set. As it can be observed in *Table 4*, the mean emotional score of the semantic neighbors was capable of predicting human rates (R^2 ranged from .28 to .41) as it was found in previous studies. Thus, this procedure can be useful as a strategy to impute missing values in data sets using semantic neighbors as a proxy of the emotional scores of target words. Also, we included the neural network scores in the same linear regression model. Results showed that the predictions of the neural network model were more important when predicting the scores of target words, whose standardized β coefficient almost doubled the semantic neighbors (except for disgust). This means that neural networks explained something that goes beyond and above the semantic neighbors suggesting that an ideal isomorphism does not exist.

Table 4. Standardized β coefficients from multiple linear regressions to predict human rates using average human rate of the 30 closest neighbors of the predicted word, and neural network scores.

	Dimensional emotions		Discrete emotions				
	Valence	Arousal	Happiness	Anger	Sadness	Fear	Disgust
Semantic neighbors*	.597	.542	.537	.637	.572	.646	.578
R^2	.36	.29	.28	.41	.32	.41	.33
Neural networks	.707	.612	.680	.735	.656	.714	.609
R^2	.50	.37	.46	.46	.43	.51	.37
Semantic neighbors	.209	.229	.217	.266	.279	.294	.358
Neural networks	.523	.453	.555	.558	.488	.512	.423
R^2	.52	.40	.49	.58	.48	.56	.46

Note: All the standardized β coefficients were significant at $p<.01$. * = Each target word contained from 2 to 26 (out of 30) semantic neighbors.

Another way to test the *semantic neighborhood hypothesis* is to correlate amodal distances and emotions of semantic neighbors of target words. That is, if the word “tomb” predicts a high human score in fear, it could be expected that its semantic neighbors are also associated with a high prediction of fear. To do this, the covariation of human scores and amodal distances (cosines; less amodal distances -higher cosines- would imply more similar emotional predictions) were analyzed to test if semantic neighbors represented the emotionality of target words. *Table 5* shows that the relation between both distributions was not different to zero in any of the emotional categories, that is, human emotional scores do not present a similar distribution to distances in the amodal vector space. As it will be discussed below, these results support that there is not an ideal isomorphism and, thus, a mapping link between amodal and emotional representations of words is probably needed to simulate specific emotional features.

Table 5. Mean Pearson correlation coefficient [95%CI] between human emotional scores and distances (cosines) in the amodal vector space of semantic neighbors of target words.

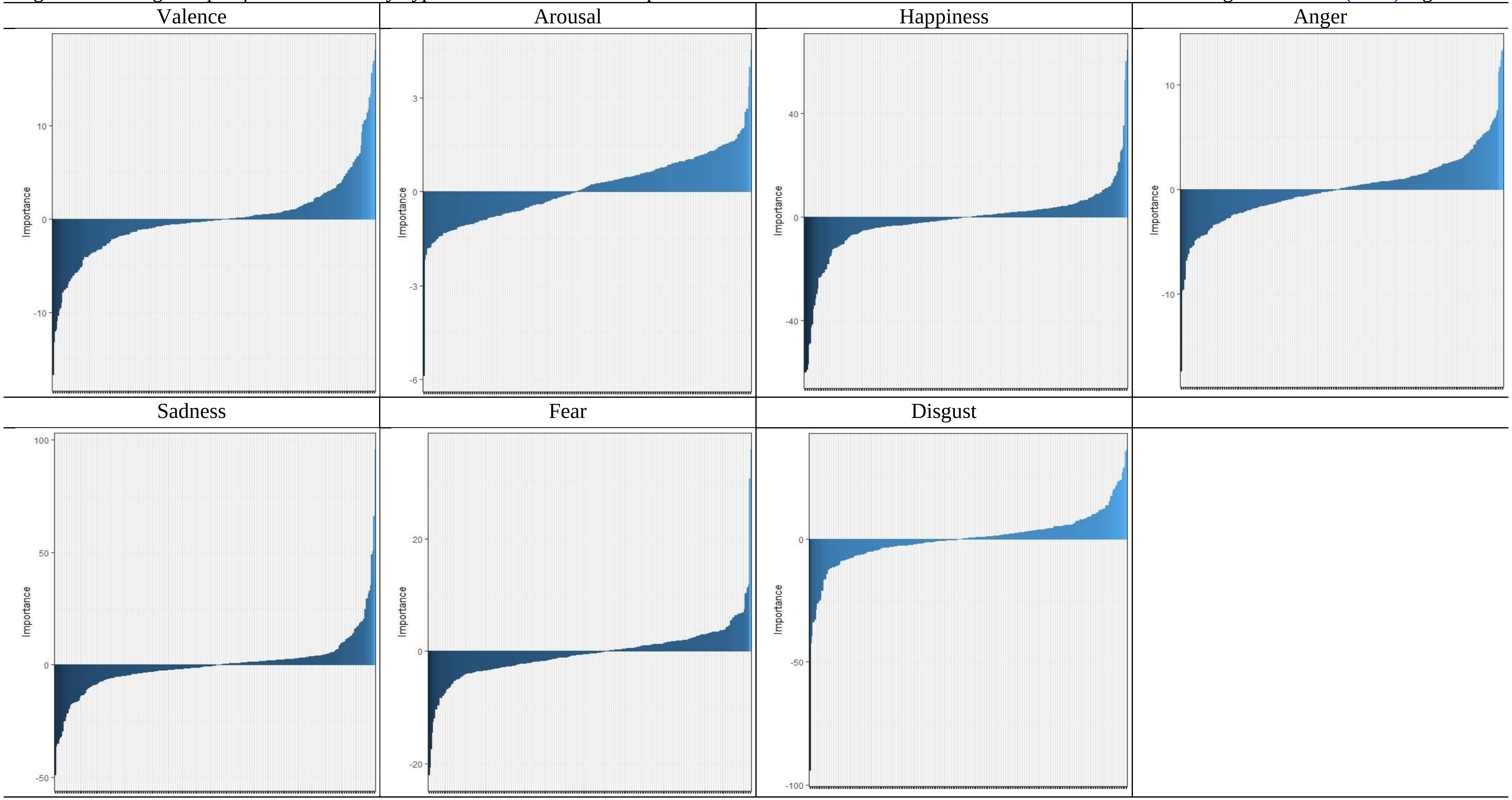
Dimensional emotions		Discrete emotions				
Valence	Arousal	Happiness	Anger	Sadness	Fear	Disgust
.008	-.002	.043	.025	-.011	-.032	-.001
[-.013—.029]	[-.023—.019]	[-.028—.114]	[-.051—.103]	[-.084—.063]	[-.110—.046]	[-.079—.063]

Note: Pearson correlation coefficients were estimated only for those target words that had more than ten semantic neighbors for dimensional categories and more than five for discrete categories.

3. *Analyzing the specific dimensionality hypothesis*

Neural networks and multiple linear regression models use specific dimensions (with different bias or weight combinations) to capture emotional properties. These models assume that modal representations are differentially concentrated in the amodal dimensions and, consequently, that the emotionality of a word cannot be reduced to the whole semantic representation. A direct way to test the *specific dimensionality hypothesis* is to analyze the relative relevance of features of amodal representations of words (in this case, 300 LSA amodal dimensions) to predict emotional scores in the neural network models. Here, we computed the relative importance of the LSA amodal dimensions in the neural network architecture using Olden algorithm (Olden et al., 2004). *Figure 3* shows that only a few dimensions of the amodal vector space were relevant to predict the emotionality of words in the emotional category (please note that some dimensions were relevant to activate and others to inhibit emotional responses).

Figure 3. Testing the *specific dimensionality hypothesis* as the relative importance of LSA amodal dimensions in the neural architecture using Olden et al. (2004) algorithm.



Note: 300 features of amodal representations of words (LSA dimensions) are ordered in the x axis by their relative importance. Relative importance of these amodal features to predict emotional scores is plotted in the y axis (Olden et al. 2004 algorithm was used to compute this measure).

Moreover, we compared the performance of univariate neural networks with multiple linear regressions, which is an alternative procedure to capture the relation between amodal and emotional representations of words (Hollis et al., 2017). Linear regression models have a simpler structure than neural networks regarding the sensibility of variable interactions. Thus, the sensitivity to more complex emotional constructs may be favored by a non-linear architecture as the neural network. We replicated Hollis et al. (2017) procedure where human rates were predicted by amodal vector representations using backward stepwise regressions through the AIC index. Concretely, different multiple linear regression models were estimated using the training set and were tested later in the test set. Results showed that backward stepwise regressions can equal neural network performances in discrete emotions, while neural networks nearly double backward stepwise regression performances in dimensional emotions (see *Table 6*). It seems that simple linear relations between emotional and amodal representations of words were sufficient to explain discrete categories. That is, discrete emotional constructs could present simpler emotional and cognitive relations.

Table 6. Performance of backward stepwise regressions in the test set as Pearson correlation coefficients, and comparison of the performance of neural networks and backward stepwise regressions.

		Dimensional emotions		Discrete emotions				
		Valence	Arousal	Happiness	Anger	Sadness	Fear	Disgust
Backward stepwise regressions	N. relevant dimensions	146	141	94	104	100	93	83
	Performance	.439	.367	.640	.721	.668	.700	.598
Neural networks vs. backward stepwise regressions	z (*p*)	18.24 (*p*<.01)	14.92 (*p*<.01)	.21 (*p*=.42)	-.81 (*p*=.79)	-1.15 (*p*=.88)	-.32 (*p*=.63)	-.26 (*p*=.60)
	95%CI	[.21—.29]	[.17—.27]	[-.05—.06]	[-.07—.03]	[-.09—.02]	[-.06—.04]	[-.07—.05]

Note: All Pearson correlation coefficients were statistically significant (*p*<.01). N. relevant dimensions = Number of LSA amodal dimensions in the final linear regression model using Hollis et al. (2017) procedure. z = One-sided backtransformed average Fisher's Z to compare dependent correlation coefficients using Hittner, May, & Silver (2003) modification. 95%CI = 95% confidence interval for differences between neural networks and backward stepwise regressions using Zou (2007) estimation.

4. Analyzing the overlapping between the dimensions of the link and the amodal dimensions of a tentative core of emotions

In order to study the overlapping between the dimensions of the link and the amodal dimensions of a tentative core of emotions, we analyzed the relations between the relative importance of LSA amodal dimensions in neural network models and the amodal vector space. Kendall rank correlation coefficients showed that no relevant relations were obtained for any of these measures (see *Table 7*). Some exceptions could be seen in the statistical significance of C-I results for sadness or fear, but the effect size was very small. This means that there was no overlapping between the relative importance of amodal dimensions for the neural network models and amodal dimensions of a tentative core of emotions in the vector space.

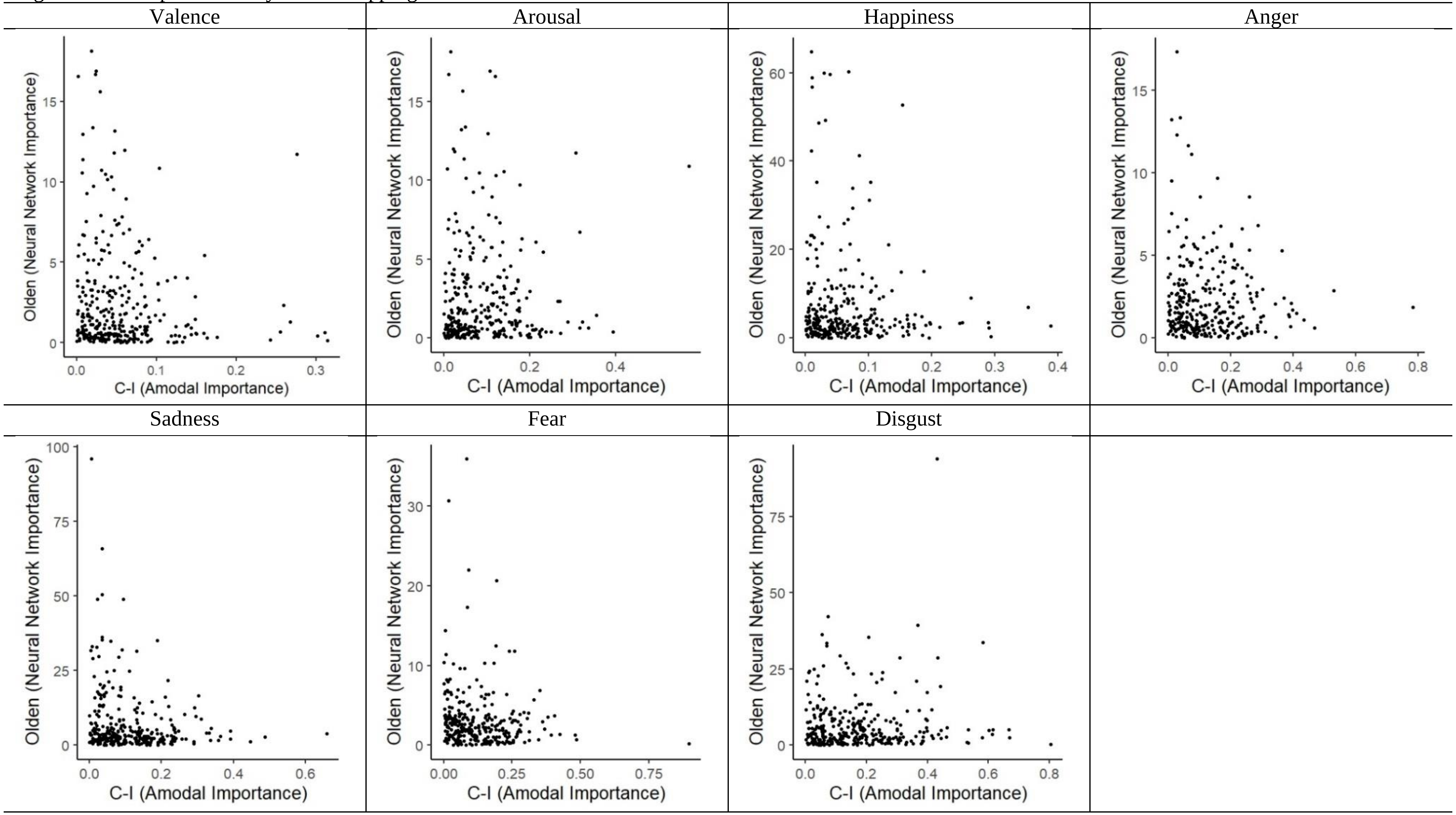
Table 7. Kendall rank correlation coefficients to compare the relative importance of LSA amodal dimensions in neural network models and in the amodal tentative core of emotions.

	Initial words	Dimensional categories		Discrete categories				
		Valence	Arousal	Happiness	Anger	Sadness	Fear	Disgust
Centroids		.01 (p=.81)	-.04 (p=.29)	-.02 (p=.68)	.00 (p=.93)	-.01 (p=.71)	-.07 (p=.06)	.00 (p=.81)
C-I	15	-.08 (p=.04)	.04 (p=.29)	-.02 (p=.53)	-.03 (p=.42)	-.10 (p<.01)	-.13 (p<.01)	-.03 (p=.47)
	30	-.06 (p=.14)	.08 (p=.07)	.00 (p=.81)	-.03 (p=.38)	-.11 (p<.01)	-.11 (p<.01)	.00 (p=.90)
	60	-.03 (p=.42)	.07 (p=.06)	.00 (p=.96)	-.03 (p=.49)	-.09 (p=.02)	-.10 (p=.01)	-.03 (p=.49)
	90	-.03 (p=.47)	.05 (p=.20)	.01 (p=.76)	-.03 (p=.37)	-.09 (p=.03)	-.10 (p<.01)	-.03 (p=.50)
	120	.02 (p=.72)	.06 (p=.13)	.00 (p=.86)	-.02 (p=.53)	-.08 (p=.04)	-.10 (p<.01)	-.02 (p=.58)

Note: N = 300 dimensions. All Kendall rank correlation coefficients were computed between the absolute value of Olden algorithm result (relative importance of LSA amodal dimensions in neural network models), and different estimations of the relative importance of LSA dimensions in the amodal vector space (centroids and different parametrizations of C-I).

Figure 4 represents the relations between the relative importance of amodal dimensions for the neural network models and amodal dimensions of a tentative core of emotions in the vector space. These graphs show that the most important dimensions for the neural network models were irrelevant for the tentative core of emotions in the vector space (and vice versa). This means that there was no relation between the relative importance of the LSA amodal dimensions for the neural network models and the tentative core of emotions in the vector space. There were some minimal agreements for valence, arousal, happiness, and disgust, but they are more anecdotal than robust. Thus, it can be concluded that there was no agreement between both measures about the most relevant LSA amodal dimensions.

Figure 4. Scatterplots to study the overlapping between the dimensions of the link and the amodal dimensions of a tentative core of emotions.



Note: Absolute values were imposed for both measures. In this case, Olden algorithm results are equivalent to Garson results. Six final semantic neighbors from 15 initial semantic neighbors were used for these C-I results.

DISCUSSION

The aim of this study was to explore and analyze three central concepts within the embodiment debate: redundancy, isomorphism, and propagative mechanisms, by means of a computational model. This model was designed according to the theoretical frameworks of the *symbol interdependency hypothesis* (Louwerse, 2011, 2018) and the *language and situated simulation theory* (Barsalou, et al., 2008), but we tested some underlying mechanisms comparing two different but compatible hypotheses: the *semantic neighborhood hypothesis* and the *specific dimensionality hypothesis*. This computational model is also in accordance with recent proposals like *acquired embodiment* and its propagative mechanisms (Hoffman et al., 2018). As was explained before, we first trained a neural network model as a mapping function to formalize emotional simulation as a bypass capable to activate emotional representations from amodal ones. In our computational architecture, amodal representations from a LSA vector space were used as inputs to predict human rates on different emotional categories (as a feature-based emotional model).

A first remark is that this computational model properly predicted human emotional rates of words. In this process, some dimensions gained relevance and acted as a mapping between emotional and amodal representations. In other words, some amodal dimensions were qualified to simulate emotional responses. In addition, words without a direct emotional experience were also beneficiated by this process via amodal representations thanks to said link. This was only possible because the words without a direct experience present similar feature values in those qualified amodal dimensions. This mechanism could be understood as verbal conditioning in which words elicit emotional responses (Field & Schorah, 2007; García-Palacios et al., 2018; Grégoire & Greening, 2020). A second remark is that results did not show a differential performance of univariate and multivariate neural network models in dimensional nor discrete emotional categories (probably, this result can be explained because the emotional categories are relatively independent and thus no benefit is gained with a multivariate model). But relevant differences were obtained between dimensional and discrete emotions when neural networks and backward stepwise regressions were compared. While no differences were obtained between both methods in discrete emotions, neural network models doubled the backward stepwise regressions performance in dimensional emotions. Consequently, it seems to be much more complex to model the relation between emotional and amodal representations of words for dimensional emotions than for discrete emotions. These results reinforce the necessity to study emotional and cognitive relations using an integrative perspective for dimensional and discrete models of emotion as it has been proposed by Harmon-Jones (2019). In this way, neural networks seem to be appropriate models for different emotional categories, supporting previous results about the complexity of models of emotional processing (e.g., Berrios, Totterdell & Kellett, 2015; Hamann, 2012).

Regarding the *semantic neighborhood hypothesis* and the *specific dimensionality hypothesis*, we found evidence in favor of the *specific dimensionality hypothesis*, although as in other studies, we also found that the *semantic neighborhood hypothesis* is plausible. The specificity of the neural network model seems to be more plausible but it is also possible to predict the emotionality of target words using the mean of the emotional rates of their semantic neighbors (see also some previous studies: Bestgen & Vincze, 2012; Hollis et al., 2017; Kuhlmann et al., 2017; Lenci et al., 2018; Recchia & Louwerse, 2015; Turney & Littman, 2003). Additional explorations showed that while the mean emotional value of the amodal semantic neighbors could predict the emotional properties of target words, the emotional values of the neighbors did not correlate with the distances in the amodal vector space. That is, closer neighbors were not necessarily scored as the target. These results suggest that there is redundancy but no isomorphism between emotional and amodal representations. In other words, although the *semantic neighborhood hypothesis* is plausible as words from the semantic neighborhoods tend to present similar amodal features, the mapping function of some specific amodal features seems to better simulate emotional responses. Thus, the *specific dimensionality hypothesis* and the *semantic neighborhood hypothesis* would not be mutually exclusive (as pointed out by some authors like Goldstone & Rogosky, 2002). These findings could indicate that decisions based on amodal representations could be efficiently used unless the task demands modal simulations as it was suggested by the *symbol interdependency hypothesis* (Louwerse, 2011, 2018), where a linking mechanism would be needed.

Furthermore, we conducted a direct comparison between the importance of the LSA amodal dimensions of the vector space for the neural network models and the amodal context of emotional words in the vector space. The first one is directly related to the *specific dimensionality hypothesis*, while the second is more related to the *semantic neighborhood hypothesis*. The latter hypothesis assumes that all the dimensions are relevant when computing emotional estimation because the semantic neighbors are estimated by means of cosines (cosine measures take into account all the dimensions without a weighting process guided by emotional context). But it could be the case that, even in the amodal representations, not all the dimensions participate equally in the emotional responses of words. Our purpose was to isolate the amodal dimensions potentially promoted by the concept of each emotion and to compare them with the dimensions that participate in the link emulated by the neural networks. That is, we compared the dimensions that are relevant for an amodal concept of emotion with the dimensions that are relevant for the link of emotional simulations. We did not find an overlapping between both kinds of dimensions, that is, the linking mechanism uses dimensions to simulate emotional responses that are not relevant to the amodal context of the emotional categories (exceptionally, only a few dimensions showed a relative agreement). Again, this fact goes beyond isomorphism as some dimensions would be important for emotional simulations while other dimensions would be important for amodal judgments.

The theoretical implications of this study are related to different theoretical predictions about language processing using the results of computational emotional estimations, linking mechanisms and semantic neighbors. Our computational model and its predictions could suggest some plausible mechanisms to deal with relative redundant information that spreads in both modal (emotional) and amodal representations, as some authors have suggested (Barsalou, et al., 2008; Louwerse, 2011, 2018), but using a mapping function (that is, without assuming isomorphism). Moreover, the propagative mechanisms via amodal properties ensure that even words without emotional experience benefit from that mechanism. Thus, these results support proposals claiming that some words can be associated to their sensorimotor and emotional information by direct exposure, while other words would acquire their sensorimotor and emotional information through amodal propagation just because they are symbolically connected with those words that had sensorimotor and emotional experience (extending previous research such as Hoffman et al., 2018 study). It can be said that complex links between emotional and amodal representations of words are needed to explain a large proportion of the emotionality of words against simple co-occurrence statistics (such as the semantic similarity in the amodal semantic space). The results of this study showed that the advantages of these linking mechanisms are more evident for dimensional emotions. It is noteworthy that these propagative mechanisms can be the key to understanding abstract words since some studies claim a propagation of emotionality to reach mature meanings (see Pexman, 2019 for a review on abstract words maturation).

The formalization of this linking mechanism between emotional and amodal representations of words can be manipulated in future research in order to test task-dependence effects. Different mathematical functions could control the relevance of the mapping function to model grounded and non-grounded tasks (tasks with or without emotional simulation), by amplifying or attenuating emotional or amodal representations of words, depending on task demands. While some tasks could require an active link, other tasks would not need it (or gradually use it), such as tasks whose processes rest in amodal representations (e.g., pure linguistic processing). The applied implications of these studies are clear: Valid models of emotional language processing can generate automatic ratings of emotional properties of words that are useful for experimental research. Although we are not there yet, we are moving in a direction where automatic raters of emotional properties of words could be useful for effortlessly testing theoretical predictions.

Conclusions

This computational study proposed a linking mechanism between emotional and amodal representations of words following different theoretical proposals, such as the *symbol interdependency hypothesis* (Louwerse, 2011, 2018) and the *language and situated simulation theory* (Barsalou, et al., 2008), and new concepts like *acquired embodiment* and its propagative mechanisms (Hoffman et al., 2018). As a computational model, it should be considered as a reduction of the complexity of the real neurocognitive system. Nevertheless, this study generated some interesting results that could be modeling the interfacing of amodal representations with sensorimotor machinery in the supramodal hub (for a review, see Nastase & Haxby, 2016; Meteyard et al., 2012) in the algorithm Marr's level. Although this model was static (non-dynamic), it enables the study of some underlying assumptions of these proposals like the redundancy, the isomorphism, and the propagative mechanisms. We showed how this computational model was able to propagate emotional responses to words without a direct emotional experience via amodal propagation. Thus, we found that there are clear redundancy and propagative mechanisms, but no isomorphism should be assumed. However, it is necessary to establish complex links to go beyond amodal distances of vector space models. Moreover, we showed that some emotional categories, such as discrete emotions, could have simpler relations between modal and amodal representations of words than dimensional emotions. Finally, we made different tentative proposals to test theoretical predictions using these computational models.

ACKNOWLEDGMENTS

The authors would like to thank Kerry McGreevy for comments and suggestions of a previous version of our article. We also would like to thank Marc Guasch from Universitat Rovira i Virgili, and José A. Hinojosa from Universidad Complutense de Madrid, for sharing their unpublished data with us.

OPEN PRACTICES STATEMENT

Both the API requests and the statistical analyses were performed using R software. Program code for neural network estimations and semantic spaces can be found in the following OSF project: https://osf.io/m8wux/; DOI:10.17605/OSF.IO/M8WUX. The data sets for all experiments are publicly available in their corresponding articles.

REFERENCES

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science, 6*(3), 359-370. https://doi.org/10.1111/tops.12096.

Balkenius, C., & Gärdenfors, P. (2016). Spaces in the brain: From neurons to meanings. *Frontiers in Psychology, 7*, 1-12. http://dx.doi.org/10.3389/fpsyg.2016.01820.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science, 1*(1), 28-58. https://doi.org/10.1111/j.1745-6916.2006.00003.x.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. de Vega, A. Glenberg and A. C. Graesser (Eds), *Symbols, Embodiment, and Meaning* (pp. 245-283). Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199217274.003.0013.

Beck, M. W. (2018). NeuralNetTools: Visualization and analysis tools for neural networks. *Journal of Statistical Software, 85*(11), 1-20. http://dx.doi.org/10.18637/jss.v085.i11.

Berrios, R., Totterdell, P., & Kellett, S. (2015). Eliciting mixed emotions: a meta-analysis comparing models, types, and measures. *Frontiers in Psychology, 6*, 428, 1-15. https://doi.org/10.3389/fpsyg.2015.00428.

Bestgen, Y., & Vincze, N. (2012). Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavior Research Methods, 44*(4), 998-1006. https://doi.org/10.3758/s13428-012-0195-z.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience, 17*(6), 905–917. https://doi.org/10.1162/0898929054021102.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin, 143*(3), 263-292. https://dx.doi.org/10.1037/bul0000089.

Diedenhofen, B. & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS ONE, 10*(4): e0121945. https://dx.doi.org/10.1371/journal.pone.0121945.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169–200. https://doi.org/10.1080/02699939208411068.

Ferré, P., Guasch, M., Martínez-García, N., Fraga, I., & Hinojosa, J. A. (2017). Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories. *Behavior Research Methods, 49*(3), 1082-1094. https://doi.org/10.3758/s13428-016-0768-3.

Ferré, P., Guasch, M., Moldovan, C., & Sánchez-Casas, R. (2012). Affective norms for 380 Spanish words belonging to three different semantic categories. *Behavior Research Methods, 44*(2), 395–403. https://doi.org/10.3758/s13428-011-0165-x.

Fraga, I., Guasch, M., Haro, J., Padrón, I., & Ferré, P. (2018). EmoFinder: The meeting point for Spanish emotional words. *Behavior Research Methods, 50*(1), 84-93. https://doi.org/10.3758/s13428-017-1006-3.

Field, A. P., & Schorah, H. (2007). The verbal information pathway to fear and heart rate changes in children. *Journal of Child Psychology and Psychiatry, 48*(11), 1088-1093. https://doi.org/10.1111/j.1469-7610.2007.01772.x.

García-Palacios, A., Costa, A., Castilla, D., del Río, E., Casaponsa, A., & Duñabeitia, J. A. (2018). The effect of foreign language in fear acquisition. *Scientific Reports, 8*(1157), 1-8. https://doi.org/10.1038/s41598-018-19352-8.

Gärdenfors, P. (1996). Conceptual spaces as a basis for cognitive semantics. In A. Clark, J. Ezquerro, & J. M. Larrazabal (Eds.), *Philosophy and Cognitive Science: Categories, Consciousness, and Reasoning* (pp. 159-180). Springer Netherlands.

Gärdenfors P. (2000). *Conceptual Spaces: On the Geometry of Thought*. MIT Press: Cambridge, MA.

Garson, G. D. (1991). Interpreting neural network connection weights. *Artificial Intelligence Expert, 6*(4), 46-51.

Glenberg, A. M. (1997). Mental models, space, and embodied cognition. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes* (p. 495–522). American Psychological Association. https://doi.org/10.1037/10227-018.

Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering, 9*(3), 143-151. https://doi.org/10.1016/0954-1810(94)00011-S.

Goldstone, R. L., & Rogosky, B. J. (2002). Using relations within conceptual systems to translate across conceptual systems. *Cognition, 84*(3), 295-320. https://doi.org/10.1016/S0010-0277(02)00053-7.

Grégoire, L., & Greening, S. G. (2020). Fear of the known: semantic generalisation of fear conditioning across languages in bilinguals. *Cognition and Emotion, 34*(2), 352-358. https://doi.org/10.1080/02699931.2019.1604319.

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods, 48*(4), 1358–1369. https://doi.org/10.3758/s13428-015-0684-y.

Günther, F., Dudschig, C., & Kaup, B. (2018). Symbol Grounding Without Direct Experience: Do Words Inherit Sensorimotor Activation From Purely Linguistic Context? *Cognitive Science, 42*(2), 336-374. https://doi.org/10.1111/cogs.12549.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science, 14*(6), 1006-1033. https://doi.org/10.1177/1745691619861372.

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in Cognitive Sciences*, *16*(9), 458-466. https://doi.org/10.1016/j.tics.2012.07.006.

Harmon-Jones, E. (2019). On motivational influences, moving beyond valence, and integrating dimensional and discrete views of emotion. *Cognition and Emotion, 33*(1), 101-108. https://doi.org/10.1080/02699931.2018.1514293.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*(2), 301-307. https://doi.org/10.1016/S0896-6273(03)00838-9.

Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review*, *14*(3), 436-441. https://doi.org/10.3758/BF03194085

Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U., Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2016). Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions. *Behavior Research Methods, 48*(1), 272–284. https://doi.org/10.3758/s13428-015-0572-5.

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of General Psychology, 130*(2), 149-168. https://doi.org/10.1080/00221300309601282.

Hoffman, P., McClelland, J. L., & Lambon-Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review, 125*(3), 293-328. http://dx.doi.org/10.1037/rev0000094.

Hollis, G., Westbury, C., & Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *The Quarterly Journal of Experimental Psychology, 70*(8), 1603-1619. https://doi.org/10.1080/17470218.2016.1195417.

Ianì, F., Foiadelli, A., & Bucciarelli, M. (2019). Mnemonic effects of action simulation from pictures and phrases. *Acta Psychologica, 194*, 37-50. https://doi.org/10.1016/j.actpsy.2019.01.012.

Jones, M. N., Gruenenfelder, T. M., & Recchia, G. (2018). In defense of spatial models of semantic representation. *New Ideas in Psychology, 50,* 54-60. https://doi.org/10.1016/j.newideapsych.2017.08.001.

Jorge-Botana, G., Olmos, R., & Barroso, A. (2013, July). *Gallito 2.0: A Natural Language Processing tool to support Research on Discourse*. Proceedings of the Twenty-third Annual Meeting of the Society for Text and Discourse, Valencia.

Jorge-Botana, G., Olmos, R., & Luzón, J. M. (2020). Bridging the theoretical gap between semantic representation models without the pressure of a ranking: some lessons learnt from LSA. *Cognitive Processing, 21*(1), 1–21. https://doi.org/10.1007/s10339-019-00934-x.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge University Press: Cambridge.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review, 7*(2), 257-266. https://doi.org/10.3758/BF03212981.

Kintsch, W. (2001). Predication. *Cognitive Science, 25*(2), 173-202. https://doi.org/10.1207/s15516709cog2502_1.

Kintsch, W., & Bowles, A. R. (2002). Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and Symbol, 17*(4), 249-262. https://doi.org/10.1207/S15327868MS1704_1.

Kintsch, W, & Welsch, D. (1991). The construction-integration model: a framework for studying memory for text. In W. E. Hockley & S. Lewandowsky S (Eds), *Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock* (pp. 367-385). Erlbaum, Hillsdale.

Kuhlmann, M., Hofmann, M. J., & Jacobs, A. M. (2017). If You Don't Have Valence, Ask Your Neighbor: Evaluation of Neutral Words as a Function of Affective Semantic Associates. *Frontiers in Psychology, 8*, 1-7. https://doi.org/10.3389/fpsyg.2017.00343.

Landauer, T. K. & Dumais, S. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–40. https://dx.doi.org/10.1.1.184.4759.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *The Handbook of Latent Semantic Analysis*. New Jersey: Routledge. https://doi/10.4324/9780203936399.

Lenci, A., Lebani, G. E., & Passaro, L. C. (2018). The emotions of abstract words: A distributional semantic analysis. *Topics in Cognitive Science, 10*(3), 550-572. https://doi.org/10.1111/tops.12335.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3*(2), 273-302. https://dx.doi.org/10.1111/j.1756-8765.2010.01106.x.

Louwerse, M. M. (2018). Knowing the Meaning of a Word by the Linguistic and Perceptual Company It Keeps. *Topics in Cognitive Science, 10*(3), 573-589. https://doi.org/10.1111/tops.12349.

Louwerse, M. M., & Benesh, N. (2012). Representing spatial structure through maps and language: Lord of the Rings encodes the spatial structure of Middle Earth. *Cognitive Science, 36*(8), 1556-1569. https://doi.org/10.1111/cogs.12000.

Louwerse, M. M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in Psychology, 3*, 1-11. https://doi.org/10.3389/fpsyg.2012.00385.

Louwerse, M. M., Hutchinson, S., Tillman, R., & Recchia, G. (2015). Effect size matters: The role of language statistics and perceptual simulation in conceptual processing. *Language, Cognition and Neuroscience, 30*(4), 430-447. https://doi.org/10.1080/23273798.2014.981552.

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition, 114*(1), 96-104. https://doi.org/10.1016/j.cognition.2009.09.002.

Louwerse, M. M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science, 33*(1), 51-73. https://doi.org/10.1111/j.1551-6709.2008.01003.x.

Mandl, T. (1999, September). *Efficient preprocessing for information retrieval with neural networks*. Presented at the 7th European Congress on Intelligent Techniques and Soft Computing (pp.13-16). Aachen, Germany.

Meteyard, L., Rodríguez-Cuadrado, S., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788-804. https://doi.org/10.1016/j.cortex.2010.11.002.

Millis, K., & Larson, M. (2008). Applying the construction-integration framework to aesthetic responses to representational artworks. *Discourse Processes, 45*(3), 263-287. https://doi.org/10.1080/01638530802070015.

Nakov, P., Popova, A., & Mateev, P. (2001, September). *Weight functions impact on LSA performance*. Presented at the EuroConference Recent Advances in Natural Language Processing (RANLP'01) (pp.187-193). Sophia, Bulgary.

Nastase, S. A., & Haxby, J. V. (2016). Structural basis of semantic memory. In J. H. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 133–151). Cambridge, MA: Academic Press.

Olden, J. D., Joy, M. K., & Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling, 178*(3-4), 389-397. https://doi.org/10.1016/j.ecolmodel.2004.03.013.

Pexman, P. M. (2019). The role of embodiment in conceptual development. *Language, Cognition and Neuroscience, 34*(10), 1274-1283. https://doi.org/10.1080/23273798.2017.1303522.

Recchia, G., & Louwerse, M. M. (2015). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *The Quarterly Journal of Experimental Psychology, 68*(8), 1584-1598. https://doi.org/10.1080/17470218.2014.941296.

Redondo, J., Fraga, I., Comesaña, M., & Perea, M. (2005). Estudio normativo del valor afectivo de 478 palabras españolas. *Psicológica, 26*, 317–326.

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior Research Methods, 39*(3), 600-605. https://doi.org/10.3758/bf03193031.

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science, 3*(2), 303-345. https://doi.org/10.1111/j.1756-8765.2010.01111.x.

Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology, 37*(3), 345–356. https://doi.org/10.1037/0022-3514.37.3.345.

Stadthagen-González, H., Imbault, C., Pérez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods, 49*(1), 111–123. https://doi.org/10.3758/s13428-015-0700-2.

Stevenson, R. A., Mikels, J. A., & James, T. W. (2007). Characterization of the affective norms for English words by discrete emotional categories. *Behavior Research Methods, 39*(4), 1020-1024. https://doi.org/10.3758/BF03192999.

Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS), 21*(4), 315-346. https://doi.org/10.1145/944012.944013.

Zou, G. Y. (2007). Exact confidence interval for Cohen's effect size is readily available. *Statistics in Medicine, 26*(15), 3054-3056. https://doi.org/10.1002/sim.2830.

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences, 18*(5), 229-234. https://doi.org/10.1016/j.tics.2014.02.008.

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review, 23*(4), 1028-1034. https://doi.org/10.3758/s13423-015-0864-x.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review, 10*(4), 954-958. https://doi.org/10.3758/BF03196557.