A Computational Study on Emotional Responses via Amodal Propagation: Dimensional Vs. Discrete Emotions

José Á. Martínez-Huertas, Guillermo Jorge-Botana, Ricardo Olmos and Alejandro Martínez-Mingo

June 29, 2020

A computational study on emotional responses via amodal propagation: Dimensional vs. Discrete emotions

José Á. Martínez-Huertas[1]

Guillermo Jorge-Botana[2]

Ricardo Olmos[1]

Alejandro Martínez-Mingo[1]

[1] Psychology Department, Universidad Autónoma de Madrid

[2] Psychology Department, Universidad Nacional de Educación a Distancia

Author Note

The authors declare that there no conflicts of interest with respect to this preprint.

Correspondence should be addressed to José Á. Martínez-Huertas (email: josea.martinez@uam.es)

Abstract

We computationally emulated a link between symbolic and emotional representations of words using computational models and predictive models. We studied dimensional and discrete emotions using two different predictive models: linear regressions and neural networks. More than 13000 words were used to train the models and then they were tested in more than 4000 words. While important differences were observed between linear regressions and neural networks in dimensional emotions, no differences were observed in discrete emotions.

Keywords: emotional responses, neural network models, symbolic representations, embodiment, computational study

A computational study on emotional responses via amodal propagation: Dimensional vs. Discrete emotions

We will never be exposed to a *t-rex*. Nonetheless, we have an emotional representation and we can experiment sensations when we hear the word "*t-rex*". This effect is even more dramatic with abstract words like "*cruel*". Some authors have proposed that language acts as a bridge to propagate some part of the emotional and the sensorimotor information to understand our environment, including words without a previous emotional experience (see Pexman, 2019 for a review). In general, these theoretical proposals can be placed within the *dual-coding theory* (Paivio, 1971), the *language and situated simulation theory* (Barsalou et al., 2008) or the *symbol interdependency hypothesis* (Louwerse, 2011, 2018), that try to accommodate grounding processes and its consequences in the differential use of modal or amodal features when processing words. All these theories have inspired some formal and computational mechanisms to explain how the symbolic representations of words acquire emotional information by both direct exposure to emotional experiences and propagative processes through symbolic properties of words. In the second case, language would act as a bridge making possible that words without emotional experience activate emotional properties and its behavioral consequences (e.g., avoidance/scape, skin conductance). New proposals linking modal and amodal representations of meaning have recently emerged at a cognitive level of explanation. In this sense, the most featured theory about the comparison between symbolic space models and sensorimotor features is the *symbol interdependency hypothesis* (Louwerse, 2011, 2018). A central idea in Louwerse's proposal is that sensorimotor properties are also encoded in symbolic representations.

As the *symbol interdependency hypothesis* argues that amodal symbolic representations of meaning can bootstrap grounded properties from some words to other words through symbolic information (see Louwerse, 2011, 2018 for a review), we proposed here a link between emotional and symbolic representations of words. This link was formalized by two different linking mechanisms to predict emotional judgements (whether they are dimensional or discrete emotional categories) from word vectors (that are extracted from a semantic space):

1. **Backwards stepwise regressions** procedure proposed by Hollis et al. (2017). These authors analyzed how multiple linear regressions can explain the emotionality of words using amodal features as predictors. This study showed appropriate predictions from backwards stepwise regressions to predict modal features.
2. **Neural network models**. We trained a neural network model to predict emotional judgements. See the *Method* section for a complete explanation of this procedure.

If these predictive models are able to generate valid emotional predictions without requiring a specific-word training (a direct emotional experience), then there will be evidences that amodal or symbolic representation captures modal or emotional features of word meaning by means of a linking mechanism.

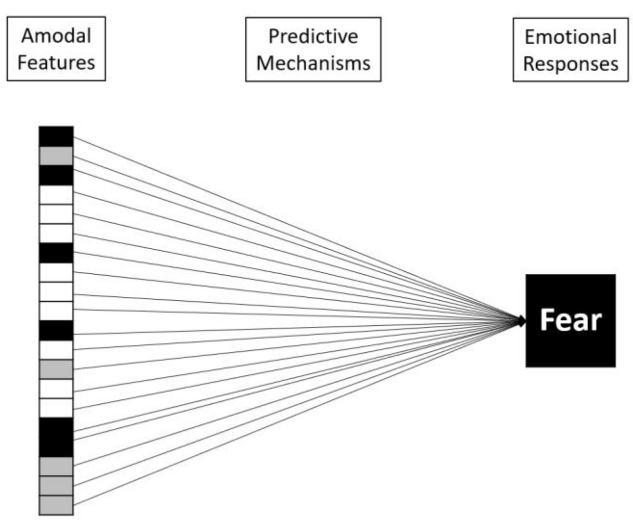
We tested both procedures in different emotional categories, namely: dimensional and discrete emotional categories. These categories are usually conceived as affect and emotion, respectively. But we maintained the names of the data sets used in this study (i.e., Fraga et al., 2018; Guasch et al., 2016; Hinojosa et al., 2016). In the present study, the first emotional categories are valence and arousal while the second emotional categories are happiness, anger, sadness, fear and disgust.

The aim of this study was to validate the scores of backwards stepwise regressions and neural network models comparing their performance with the agreement of individual human raters. Thus, our objective is different to the objectives of other studies based on testing the performance of these models according to general measures of normative data sets. Specifically, in this study, we are going to compare the performance of these predictive models proposed to link symbolic and emotional representations of words, comparing them with each individual human rater and her/his agreement with the mean score of her/his reference group.

Method

Figure 1 presents a general schematic diagram of the procedure of this study. As it can be observed, amodal features of words (in this case, vector representations from *Latent Semantic Analysis*; Landauer et al., 2007) were linked to different emotional features of words (in this case, human emotional judgements from *emoFinder* platform; Fraga et al., 2018). The main point is that we compared two different linking mechanisms, namely: backwards stepwise regressions and univariate neural network models.

Figure 1. *Schematic Diagram of the Procedure of the Present Study.*



As stated above, amodal features were extracted from vector representations of *Latent Semantic Analysis* (Landauer et al., 2007). We used a random sample of the Spanish Wikipedia composed by 455,969 documents (paragraphs) and 70,244 unique terms. We applied standard *Latent Semantic Analysis* procedures and 300 dimensions were imposed for the semantic space.

Emotional features were extracted from *emoFinder* platform (Fraga et al., 2018). This normative data set was split into a training data set and a test data set. The training data set was composed by 11,357 words for the dimensional categories and 2,266 words for the discrete categories. The test data set was composed by 4,167 words for the dimensional categories and 875 words for the discrete categories.

Then, different linking mechanisms were established between the amodal features and the emotional features using the training data set. This means that both backwards stepwise regressions and univariate neural network models were trained in this sample of words (training data set). In this step, a mapping function was established between both types of representations of words.

Once the predictive models learned how to predict emotional features from amodal features in the training data set, they were tested in the test data set. It is noteworthy that we used data sets from Hinojosa et al. (2016) and Guasch et al. (2016) as the test data set for this specific study. Here, we compared the performance of these models with the agreements of individual raters with the mean emotional judgements of her/his reference group. Specifically, we compared the Pearson correlation coefficients of the predictions of neural network models, backwards stepwise regressions, and individual raters with mean emotional judgements of the *emoFinder* platform.

Results

The neural network models and the backwards stepwise regressions reached a good performance when predicting mean emotional judgements of human raters for individual words. This is why we tested here if the performance of these predictive models can be compared to individual human raters from normative data sets. Our purpose here was to study where is placed the neural network as a rater (if it were another human rater) by correlating its emotional scores with the average rater scores. Data sets from Hinojosa et al. (2016) and Guasch et al. (2016) were used to obtain individual correlations for the different emotional categories (valence, arousal, happiness, anger, sadness, fear, and disgust).

Table 1 presents Pearson correlation coefficients as a measure of agreement between human raters and also between computational methods and human raters. Specifically, we report here the range of Pearson correlation coefficients between each human rater and the mean of its reference group for different words. Also, we report the percentiles of neural networks and backwards stepwise regression scores in comparison with individual human agreements with the mean score of her/his reference group. Percentiles act as a measure of the relative position of these models comparing to the agreement between individual human raters.

Table 1. *Range of Pearson correlation coefficients (r) for human raters (HR), neural networks (NN) and backwards stepwise regression (BSR) scores, and percentiles (Perc.) of their performance in comparison to HR.*

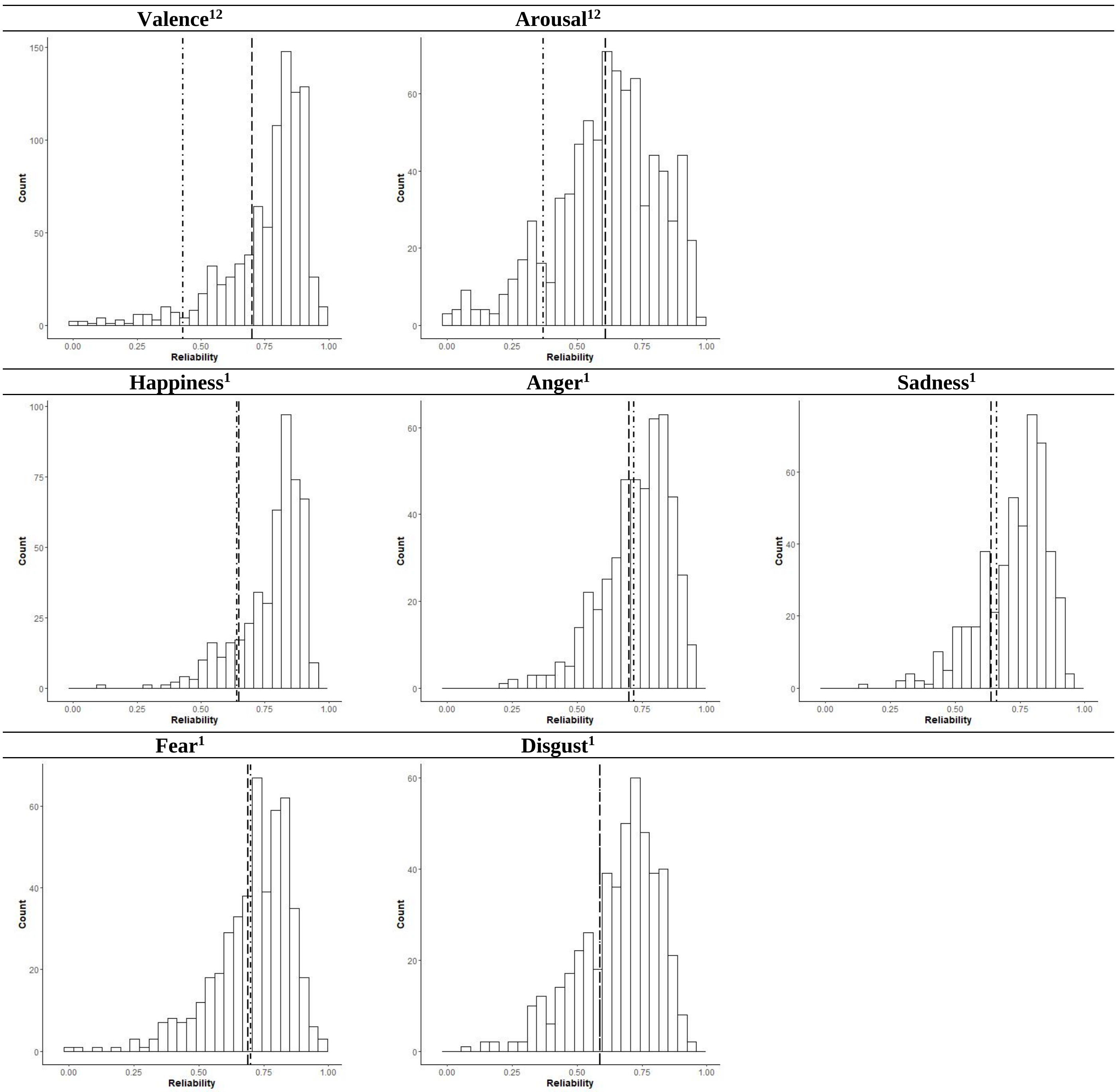
		Dimensional categories		**Discrete categories**				
		Valence[12]	Arousal[12]	Happiness[1]	Anger[1]	Sadness[1]	Fear[1]	Disgust[1]
r	HR	-.71–.97	-.65–.96	-.20–.95	-.13–.94	-.24–.94	-.25–.97	-.14–.94
	NN	.70	.61	.65	.70	.64	.69	.59
	BRS	.43	.37	.64	.72	.66	.70	.59
Perc.	NN	.27	.47	.16	.38	.27	.38	.28
	BRS	.06	.14	.15	.42	.29	.40	.28

Note: Percentiles (Perc.) represent the position of neural network reliabilities within inter-rater reliabilities. 1 = Data set from Hinojosa et al. (2016). 2 = Data set from Guasch et al. (2016).

Table 1 shows a considerable variability within human raters. Thus, we graphically represent the distribution of Pearson correlation coefficients of human raters in *Figure 2*, and we positioned the performance of neural networks and backwards stepwise regression scores to show their similarity with ordinary human raters. Moreover, *Figure 2* also shows the relative position (equivalent to the percentiles) of the scores of these computational methods in the range [.00-1.00] of human rater agreements (this means that negative correlations for the agreements of human raters were excluded for the graphical representation). These histograms present a considerable similarity of computational scores with human scores, but important differences between computational methods are observed for the dimensional and discrete emotional categories.

Figure 2

Distribution of Pearson Correlation Coefficients and Neural Networks Performance (vertical long-dash line) and Backwards Linear Regressions Performance (vertical dot-dash line).



Note: x axes represent human inter-rater agreements with range [.00-1.00]. y axes represent counts. 1 = Data set from Hinojosa et al. (2016). 2 = Data set from Guasch et al. (2016).

Discussion

Some words can be associated to their sensorimotor and emotional information by direct exposure, but other words can acquire their sensorimotor and emotional information through amodal propagation just because they are symbolically connected with those words that had a previous sensorimotor and emotional experience. We postulated a linking mechanism by means of neural network models and backwards stepwise regressions to predict emotional properties from symbolic representations. In this study, we validated such link using the agreements of individual human raters as a reference criterion (which is not very usual in computational science due to usually the mean scores of human raters is used as a validity criteria). In this case, we analyzed if neural networks and backwards stepwise regression scores can be used as an ordinary human rater.

All the scores of neural network models were constrained in ±1.5 standard deviations of the mean of human reliabilities (this information can be inferred from the percentiles). Thus, neural network models can be considered as a mean human rater. Backwards stepwise regressions showed a considerably worst performance in dimensional emotional categories (also called affect), that is, in valence and arousal. Thus, while no differences were obtained between both methods in discrete emotions, neural network models doubled the backward stepwise regressions performance in dimensional emotions. These results reinforce the necessity to study emotional and cognitive relations using an integrative perspective for dimensional and discrete models of emotion

These normative data sets are very useful tools for researchers and computational methods like the ones analyzed here are a great alternative to generate scores effortless. In this way, the main importance of normative data sets is the estimation of the mean for different groups of human raters. But we tested here how neural network models can learn to be a mean human rater and thus showing its relevance for different assessment tasks within experimental research. Computational measures are relevant to predict different psycholinguistic variables like the emotional valence.

Acknowledgments

The authors would like to thank Marc Guasch from Universitat Rovira i Virgili, and José A. Hinojosa from Universidad Complutense de Madrid, for their data to analyze human scores reliabilities.

References

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.

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.

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.

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.

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.

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.

Paivio, A. (1971). Imagery and Language. In S.J. Segal (Ed.), *Imagery: Current Cognitive Approaches* (pp.7-32). New York: Academic Press.

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.