Title

Predicting Conceptual Concreteness from Sensorimotor Information Using Artificial Neural Networks

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Ciria, Alejandra
Arias-Trejo, Natalia
Angulo-Chavira, Armando Q
et al.

Publication Date

2026

Predicting Conceptual Concreteness from Sensorimotor Information Using Artificial Neural Networks

Alejandra Ciria (alejandra.ciria@unam.mx)[1], Natalia Arias-Trejo[2],
Armando Quetzalcóatl Angulo-Chavira[2], J. Daniel Arzate-Mena[3] & Bruno Lara[3]
[1]Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México
[2]Facultad de Psicología, Universidad Nacional Autónoma de México
[3]Centro de Investigación en Ciencias, Universidad Autónoma del Estado de Morelos.

Abstract

The concreteness–abstractness distinction of concepts faces challenges from proposals that conceptual representations occupy a continuous, multidimensional sensorimotor space. Additionally, there are concerns regarding psycholinguistic norms, arguing that presenting words in isolation introduces ambiguity and that the tendency to have high variability in mid-range ratings complicates the interpretability of judgments. Here, an artificial neural network (ANN) was used to model the relationship between concreteness and sensorimotor information by incorporating mean sensorimotor strength and inter-rater variability across 11 perceptual and action dimensions to allow for nonlinear interactions. The ANN predicted graded concreteness, indicating that sensorimotor information encodes structured regularities irreducible to additive effects. Prediction error analysis revealed that inter-rater variability provides informative structure rather than noise, and that performance is not driven by lexical ambiguity. These findings suggest that conceptual representations rely on systematically organized sensorimotor systems, supporting the concreteness–abstractness distinction as a graded, probabilistic construct grounded in sensorimotor experience.

Keywords: Embodied cognition; grounded cognition; concreteness; sensorimotor norms; artificial neural networks

Introduction

Cognition is shaped by embodied experience and grounded in interactions with the physical and social environment (Barsalou, 2008; Foglia & Wilson, 2013; Varela et al., 2017; Wilson, 2002). This idea is captured by the notion of embodiment, according to which conceptual semantics and representations emerge from perceptual and motor systems (Fischer & Zwaan, 2008). While grounded cognition theories have provided strong accounts of concrete concepts, the grounding of abstract concepts, such as *freedom* or *truth*, remains a central theoretical debate, mainly because many abstract concepts lack direct perceptual or motor referents (Borghi et al., 2017; Dove, 2016, 2021). At the same time, it has been argued that the meanings of both concrete and abstract words are grounded in experiences of perception, action, affect, and social interaction (Borghi & Mazzuca, 2023; Gallese, 2025). However, the dichotomy between concrete and abstract concepts has been challenged, with concepts instead characterized as positions along a continuous, multidimensional sensorimotor space (Banks & Connell, 2023; Barsalou, 2020; Diveica et al., 2025). This raises the question of whether sensorimotor information can still reliably differentiate degrees of concreteness within such a continuous, multidimensional sensorimotor space.

The present study examines whether sensorimotor information provides a structured basis for predicting conceptual concreteness. To this end, the Lancaster sensorimotor norms (Lynott et al., 2020) and concreteness norms (Brysbaert et al., 2014), both based on ≈40,000 concepts, were coupled. This study extends prior work showing that sensorimotor mean ratings are related to concreteness (Banks & Connell, 2023; Diveica et al., 2025; Dolatabadi, 2023), by using an artificial neural network (ANN) to test whether graded concreteness can be predicted from the joint structure of sensorimotor mean ratings and their corresponding inter-rater variability, allowing for nonlinear interactions among these variables.

The present work also examines the structure of the ANN's prediction error as a function of key conceptual variables, including lexical ambiguity derived from WordNet (Miller et al., 1990), the dominant modality or effector, and other sources of sensorimotor conceptual grounding available in the Lancaster sensorimotor norms (Lynott et al., 2020). Systematic variation in prediction error provides a test of whether model performance reflects structured relationships within the sensorimotor space rather than simple or additive effects, and whether participants rely on systematically organized sensorimotor representations rather than isolated or ambiguous cues when making both sensorimotor and concreteness judgments. To further evaluate this, three permutation tests were implemented. PT-1 assesses the relevance of individual input features, PT-2 tests whether the ANN exploits multivariate dependencies by disrupting linear and nonlinear interdependencies, and PT-3 evaluates whether performance exceeds that expected under no input–output association. Together, these tests provide a principled framework for assessing whether concreteness can be accounted for by structured, multivariate sensorimotor representations.

Concreteness is a key construct for understanding language representation and processing, as concrete and abstract words differ systematically in performance across linguistic tasks (Brysbaert et al., 2014). On average, concrete words are processed faster and more accurately in tasks such as reading aloud, lexical decision, and recognition memory, a pattern known as the concreteness effect (Fliessbach et al., 2006; Pexman et al., 2017). However, concreteness alone does not fully account for these effects. In lexical decision and word-naming tasks, modality-specific experience, particularly the

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

maximum perceptual strength of a concept, has been shown to be a more reliable predictor of performance than concreteness or imageability (Connell & Lynott, 2012). More broadly, sensorimotor information has been shown to facilitate lexical processing speed and accuracy (Lynott et al., 2020).

Sensorimotor norms offer a powerful methodological tool for addressing this issue, as they capture phenomenological sensorimotor experiences for large sets of concepts. The Lancaster Sensorimotor Norms (Lynott et al., 2020) provide ratings across 11 dimensions for ≈40,000 concepts, comprising six perceptual modalities (auditory, gustatory, haptic, olfactory, visual, interoceptive) and five action effectors (head, arm/hand, mouth, leg/foot, torso). Participants rate the extent to which a concept has been experienced (from 0 *not experienced at all* to 5 *experienced greatly*) based on their sensorimotor experience. These norms provide a rich source of information about conceptual grounding, including exclusivity scores (the extent to which a particular concept is experienced through a single dimension), dominant modality or effector (maximum sensorimotor strength), and a composite strength measure across 11 dimensions (Minkowski 3 distance) (Lynott et al., 2020).

Following Raia (2023), rather than addressing the long-standing debate on whether conceptual representations are modal or amodal (Dove, 2009; Machery, 2007), the present study adopts an empirical approach that uses an ANN to examine the sensorimotor information associated with concepts to predict concreteness, including both mean ratings and inter-rater variability, while also analyzing the structure of the model's prediction error. From this perspective, sensorimotor norms inform how embodied experiential information is accessed and distributed across individuals during conceptual processing, without making claims about the ontological status of conceptual representations. This task-based perspective provides a principled framework for examining conceptual grounding.

Concreteness ratings

In their large-scale concreteness norms study, Brysbaert et al. (2014) collected behavioral ratings for ≈40,000 English concepts to provide researchers with reliable concreteness values for a wide lexical range. Participants were instructed to rate how concrete or abstract the meaning of each word was on a 5-point scale (1 *abstract language-based*, to 5 *concrete experience-based*), based on the extent to which it could be directly experienced through the senses or actions (e.g., seeing, touching, hearing), with lower ratings indicating meanings primarily understood through language. This distinction follows the traditional definition explicitly given in the instructions: concrete concepts referred to entities or actions grounded in physical experience (e.g., "couch" or "jump"), whereas abstract concepts could not be experienced directly and were best explained using other words to capture their meaning (e.g., "justice").

Pollock (2018) cautions against an uncritical use of concreteness and related semantic variables, highlighting that high standard deviations, particularly for concepts with mid-range concreteness ratings, pose a serious interpretative challenge. Brysbaert et al. (2014), ratings tend to be highly consistent: concepts such as *apple* (mean = 5, SD = 0) or *bike* (mean = 5, SD = 0) are reliably judged as concrete, whereas *belief* (mean = 1.19, SD = 0.68) or *justice* (mean = 1.45, SD = 0.74) are consistently rated as abstract. In contrast, concepts located near the center of the scale exhibit considerable disagreement among raters. For example, words such as *question* (mean = 3.36, SD = 1.70) and *drive* (mean = 3.86, SD = 1.21) yield mid-range averages that conceal large dispersion, resulting in distributions that closely mirror those observed for concreteness ratings more generally. A similar pattern was observed by Pollock (2018) in the Modality Exclusivity Norm (MEN; Lynott and Connell, 2013), despite its multi-dimensional sensory representation across vision, audition, touch, taste, and smell. The observed distributions indicate that participants tend to resist rating sensorimotor experience on a continuous scale, rarely using mid-scale values, thereby challenging the interpretability of mean-based measures of semantic psycholinguistic norms.

An additional source of difficulty in interpreting concreteness ratings stems from the fact that participants are typically presented with isolated, decontextualized lexical forms rather than with clearly specified concepts or meanings (Löhr, 2022). This difficulty also applies to sensorimotor ratings (Trott & Bergen, 2022). Lexical forms can be conventionally associated with multiple concepts that differ substantially in their degree of concreteness, as in cases of homonymy and polysemy (e.g., bank, chair, school) or, more generally, when a single expression applies to both physical and abstract entities (e.g., art, border) (Löhr, 2022). This ambiguity makes it difficult to determine which meaning participants have in mind when providing a rating and raises the possibility that judgments reflect an averaging across multiple senses. Such ambiguity is especially problematic for mid-range items and offers a plausible explanation for the high variability observed in concreteness judgments (Pollock, 2018), reinforcing concerns about relying exclusively on mean-based measures to characterize conceptual representations.

Sensorimotor ratings and concreteness

Beyond characterizing sensorimotor profiles, prior work has examined whether perceptual information can predict degrees of concreteness. Using perceptual mean ratings from Lynott and Connell (2009, 2013) and concreteness ratings from the MRC database (Coltheart, 1981), Connell and Lynott (2012) showed that perceptual modalities contribute inconsistently to concreteness judgments when abstract and concrete words are analyzed separately using stepwise multiple linear regression models. For abstract words, concreteness was weakly predicted by a subset of modalities, with visual strength showing a positive association and auditory and olfactory strength showing negative associations, whereas for concrete words, visual and olfactory strength yielded better model fit. Across analyses, only visual experience consistently predicted concreteness, while other modalities changed direction, including a reversal of the olfactory effect across the abstract–concrete divide. These findings suggest that participants relied on different decision criteria when judging abstract and concrete words, challenging the assumption that concreteness reflects a single perceptual continuum.

Recent studies show that both concrete and abstract concepts are grounded in multidimensional sensorimotor experience. Using the Lancaster sensorimotor norms (Lynott et al., 2020) and exploratory analyses of category-generated concepts, Banks and Connell (2023) found that abstract and concrete concepts engage overlapping dimensions but differ in dominant profiles: concrete concepts are more associated with vision, haptic information, and hand–arm actions, whereas abstract concepts rely more on interoception, audition, and mouth and head movements. Complementing this view, Diveica et al. (2025) used data-driven graph network analyses to show differences in the internal structure of dimensions, with haptic information central for concrete concepts and interoception and mouth movements central for abstract concepts, supporting multidimensional, non-categorical accounts of conceptual representation. Similarly, Dolatabadi (2023) showed that concreteness can be moderately predicted from sensorimotor mean ratings ($R^2 \approx 0.51$), although the use of linear models raises concerns about whether such approaches capture the interdependent structure of sensorimotor dimensions.

Existing studies range from exploratory and network-based analyses to mean-based linear models. While network approaches capture interdependencies among sensorimotor dimensions, they do not directly model how these dimensions jointly predict graded concreteness or incorporate inter-rater variability, and linear models impose additive assumptions. Moreover, standard deviations (SD) in linguistic norms are typically excluded, despite their potential to capture variability across individuals and inform the structure of conceptual representations.

Current Work

In the present study, an ANN was trained on sensorimotor ratings (Lynott et al., 2020), using both mean and SD values, to predict concreteness ratings (Brysbaert et al., 2014). An ANN framework was adopted because prior work has relied mainly on linear regression approaches (Connell & Lynott, 2012; Dolatabadi, 2023), which assume additive contributions of sensorimotor dimensions, whereas these dimensions are theoretically interdependent and may contribute to concreteness through nonlinear interactions. The goal was therefore not simply to improve prediction, but to test whether jointly modeling mean strength and inter-rater variability across dimensions recovers structured relationships that additive linear models may miss. While mean ratings capture central tendencies of sensorimotor grounding, SD values index variability across individuals in how that grounding is accessed during the rating task, providing complementary information about the stability and distribution of conceptual representations.

Based on this framework, it is hypothesized that concepts with higher concreteness ratings will exhibit more stable sensorimotor patterns due to their grounding in shared interactions with the environment, resulting in more accurate predictions, whereas concepts with lower concreteness ratings are expected to show greater variability in their sensorimotor patterns, as they are associated with perceptually heterogeneous actions and events (Reilly et al., 2025), which may affect prediction accuracy. In addition, interoception is expected to play a central role in predicting degrees of concreteness (Diveica et al., 2025; Mazzuca et al., 2025; Villani et al., 2021). To test these hypotheses, ANN performance was evaluated on a test set, and prediction error was analyzed. Lexical ambiguity was also examined as a predictor of model error to assess whether successful prediction reflects averaging across meanings or a stable sensorimotor structure associated with a concept.

Although high variability in concreteness ratings has been taken to challenge the usefulness of the concreteness–abstractness distinction (Pollock, 2018), and the dichotomy has been reconsidered in favor of a continuous, multidimensional sensorimotor space (Banks & Connell, 2023; Barsalou, 2020; Diveica et al., 2025), the present work suggests that sensorimotor information nonetheless supports reliable prediction of graded concreteness, indicating that this distinction remains useful not as a categorical divide, but as a continuous and probabilistic construct. Moreover, if an ANN trained on sensorimotor information can reliably predict concreteness, this suggests that, despite the ambiguity inherent in presenting words in isolation, participants are not responding arbitrarily but rely on systematically organized sensorimotor features when making sensorimotor and concreteness judgments.

Method

The R environment (Team et al., 2016) (v4.3.1) was used to run *Keras* (Arnold, 2017) (v2.15.0) and *TensorFlow* (v2.15.1) to train a feed-forward, fully connected ANN, and the statistical analysis. The ANN used as training data the 22 unit vectors, corresponding to the sensorimotor strength means and SD values for each concept.

The ANN uses rectified linear activation units (ReLU) in three hidden layers where $f(x) = max(0, x)$ and a logistic unit in the output layer as $f(x) = 1 + 4 \cdot \sigma(x) = 1 + \frac{4}{1+e^{-x}}$. The layers have 22-64-32-16-1 units each. The ANN is trained using the Root Mean Square Propagation algorithm (RMSProp) (Tieleman & Hinton, 2012) with a learning rate of 0.00001, $\rho = 0.9$, a momentum = 0.0, and a very small ϵ ($\epsilon = 1e - 7$) to avoid division by zero.

Training

The dataset used consists of 39,708 words with their respective sensorimotor strength (mean and SD ratings) (Lynott & Con-

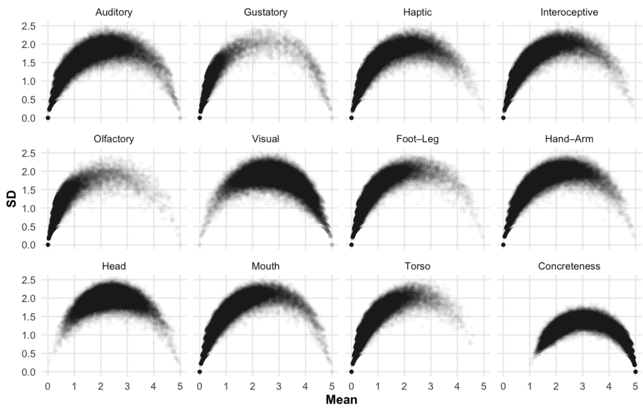

Figure 1: Dispersion plot of means and SD for each sensorimotor dimension (Lynott et al., 2020) and for concreteness ratings (Brysbaert et al., 2014). The scale of sensorimotor mean ratings goes from 0 to 5, and the concreteness mean scale goes from 1 to 5.

nell, 2009) and concreteness mean ratings (Brysbaert et al., 2014) (Fig. 1). Prioritizing a robust evaluation of the model, the dataset was divided into three different sets: a) training (70% of the data), which is used to modify the free parameters of the network while training, b) test (15% of the data), used for monitoring the training, propagating this dataset after every epoch, and calculating the error, and c) prediction set (15% of the data), these examples are not *seen* during training and are used to calculate the errors reported in statistical analysis (Table 1). The network was trained for 400 epochs.

Permutation tests

The statistical significance of the ANN performance and the relevance of the input features were assessed using 3 permutation tests (PT-1, PT-2, and PT-3) (Good, 2013; Ojala & Garriga, 2010). For each permutation test, a null distribution was generated by disrupting specific statistical dependencies in the data while preserving the overall structure of the learning pipeline. Depending on the PT, this involved breaking the association between inputs (sensorimotor means and SD ratings) and outputs (concreteness means ratings), disrupting interdependencies among input features, or permuting individual features. In all cases, any systematic relationship learned by the ANN was selectively removed while preserving marginal distributions where appropriate. For each permutation, the same training, hyperparameter selection, and evaluation procedure used for the original data was applied, and the corresponding performance of MSE and MAE was recorded.

PT-1 assesses the relevance of each input feature through a feature-wise permutation procedure. Specifically, one input feature was selected, and its values were randomly shuffled across samples, while all other features were kept fixed, after which the ANN was retrained. The resulting metric was then compared to the baseline metric obtained using the non-permuted data. This procedure was repeated independently 50 times for each feature. PT-1 therefore evaluates the contribution of individual input features to prediction error.

PT-2 tests whether the ANN exploits multivariate relationships among input variables to achieve its predictive performance. Under the null hypothesis that the ANN does not rely on multivariate dependencies among input features, PT-2 disrupts both linear and nonlinear interdependencies by independently permuting, across samples, the observed values of each feature while preserving their marginal distributions. This procedure was performed for 500 permutations of the input data, with the ANN trained for 400 epochs on each permuted dataset. Based on performance across permutations, an empirical p-value was computed following (Ojala & Garriga, 2010) as:

$$p = \frac{1 + \sum_{i=1}^{N_{\text{perm}}} \mathbb{I}(M_i \leq M_{\text{obs}})}{1 + N_{\text{perm}}} \tag{1}$$

where N_{perm} is the number of permutations, and $\mathbb{I}(M_i \leq M_{\text{obs}})$ returns the number of ANN models where the MSE was smaller than the MSE on the non-permuted data (M_{obs}).

Finally, the PT-3 evaluates whether the observed prediction performance is greater than expected under the null hypothesis of no input–output association by randomizing the order of the output vectors (concreteness) across samples. A total of 500 label permutations were performed.

Results

The training, test, and prediction errors can be seen in Table 1, reporting Mean Squared Error (MSE), which is the loss error used for training, and Mean Absolute Error (MAE).

Table 1: MSE and MAE values for each network after 400 training epochs

Dataset	MSE	MAE
Training	0.359	0.475
Test	0.375	0.482
Prediction	0.357	0.473

The error values in Table 1 measure the predictive performance of the ANN under standard test conditions. Further tests were performed to assess the extent to which that performance (prediction error) depends on the structure of the input data and its relation to the output data. The results for PT-1 can be seen in Fig. 2 and show the relative significance of each feature of the input data and its relative contribution to the resulting error. It can be seen that when not presented in the correct and structured order, the interoceptive mean values have the most impact on the network performance. Additionally, for permutation tests (PT-2, PT-3), regardless of the error metrics used (MSE or MAE), the observed performance on the original data was significantly better than chance (permutation test, $p = 0.001$).

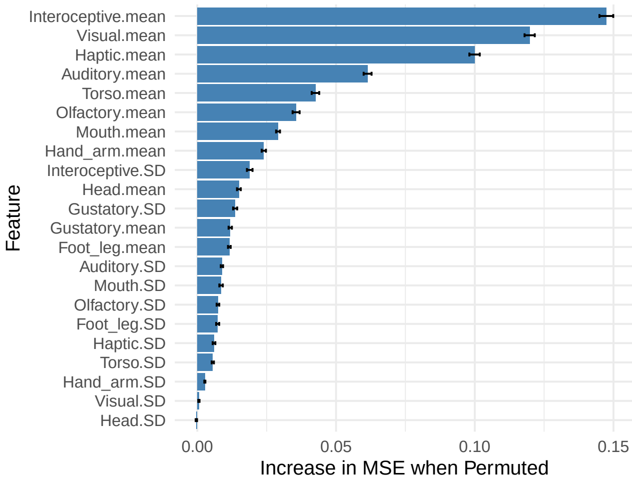

Figure 2: Effect of single input features on the MSE when permuted (PT-1) and forwarded on a trained ANN.

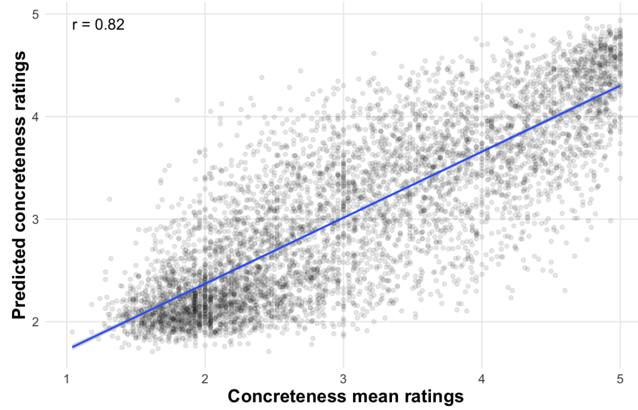

Figure 3: Relationship between actual concreteness mean ratings and concreteness values predicted by the ANN. Each point represents a concept. The solid line depicts the least-squares linear fit.

Model Performance

Model performance was assessed by examining the association between actual concreteness ratings and the values predicted by the ANN, using Pearson's correlation. A strong positive correlation was observed between actual and predicted concreteness ratings, $r = 0.82$, 95% CI $[0.81, 0.82]$, $t(5953) = 109.05$, $p < 0.001$. This result indicates that the ANN accurately captured the metric structure of concreteness ratings, accounting for approximately 67% of the variance in the observed data ($r^2 = 0.67$). Importantly, the magnitude of the correlation, together with the results of PT-2 and PT-3, suggests that the network did not simply reproduce ordinal relationships among concepts, but rather learned graded differences in concreteness. This finding supports the claim that sensorimotor information provides sufficient structure for predicting fine-grained conceptual properties such as degree of concreteness.

Concreteness and Prediction Error

To examine which aspects of sensorimotor grounding account for variation in the ANN's prediction error, we fitted a Gamma generalized linear model with a log link function. Visual inspection of the error distribution, together with Cullen–Frey diagnostics implemented in the *fitdistrplus* package (Delignette-Muller & Dutang, 2015), indicated that the data were well described by a Gamma distribution. The fitted model showed adequate dispersion (dispersion parameter = 0.58). In addition to sensorimotor predictors, the model incorporated a measure of lexical ambiguity derived from WordNet (Miller et al., 1990), operationalized as the number of senses associated with each word across all parts of speech. WordNet is a large lexical database in which word meanings are organized into sets of cognitive synonyms, providing a widely used resource for estimating semantic ambiguity and polysemy in psycholinguistic research.

The full model included actual concreteness ratings (Brysbaert et al., 2014), the coefficient of variation of concreteness judgments, sensorimotor exclusivity, global sensorimotor strength indexed by the Minkowski-3 distance, dominant sensorimotor dimension (Lynott et al., 2020), and WordNet-based lexical ambiguity. Dominant sensorimotor dimension was entered as a categorical factor, with *Interoceptive* serving as the reference category. Model diagnostics conducted using the *performance* package (Lüdecke et al., 2021) revealed no substantial violations of distributional assumptions, problematic collinearity, or influential observations. The model significantly reduced deviance relative to the null model (residual deviance = 4735.1 on 5935 degrees of freedom; null deviance = 4816.4 on 5950 degrees of freedom), yielding an AIC of 2634.7, indicative of an adequate overall fit.

The intercept ($\beta = -0.88$, SE $= 0.11$, $t = -7.66$, $p < 0.001$) corresponds to the expected log prediction error for concepts with interoceptive dominance at zero-centered predictor values. Actual concreteness ratings exhibited a robust positive association with prediction error ($\beta = 0.085$, SE $= 0.011$, $t = 8.04$, $p < 0.001$; standardized $\beta = 0.09$, 95% CI [0.07, 0.11]), corresponding to an estimated 8.8% increase in prediction error for each one-unit increase in concreteness (95% CI [6.7%, 11.0%]). In contrast, the coefficient of variation of concreteness ratings was negatively associated with prediction error ($\beta = -0.16$, SE $= 0.079$, $t = -2.09$, $p = 0.037$; standardized $\beta = -0.03$, 95% CI $[-0.05, 0.00]$), indicating an estimated 15.2% decrease in prediction error per unit increase in interindividual variability of concreteness judgments (95% CI [−27.0%, −1.4%]).

Crucially, WordNet-based lexical ambiguity did not significantly predict prediction error ($\beta = -2.78 \times 10^{-5}$, SE $= 0.003$, $t = -0.01$, $p = 0.99$; standardized $\beta \approx 0$, 95% CI $[-0.02, 0.02]$), indicating that the ANN's performance was not modulated by the number of lexical senses associated with a concept. Neither sensorimotor exclusivity ($\beta = 0.16$, SE

Table 2: Estimated marginal means of prediction error by dominant sensorimotor dimension (response scale). The reference category used in the model (*Interoceptive*) is shown in bold.

Dominant sensorimotor	Mean	SE	95% CI
Auditory	0.43	0.02	[0.40, 0.47]
Gustatory	0.42	0.03	[0.36, 0.49]
Haptic	0.43	0.03	[0.37, 0.51]
Interoceptive	**0.48**	0.02	[0.44, 0.53]
Olfactory	0.61	0.09	[0.46, 0.81]
Visual	0.48	0.01	[0.47, 0.50]
Foot–leg	0.45	0.05	[0.37, 0.55]
Hand–arm	0.47	0.02	[0.42, 0.52]
Head	0.46	0.01	[0.44, 0.48]
Mouth	0.44	0.03	[0.39, 0.49]
Torso	0.55	0.07	[0.43, 0.71]

= 0.16, $t = 0.99$, $p = 0.32$; standardized $\beta = 0.01$) nor global sensorimotor strength indexed by the Minkowski-3 distance ($\beta = -0.017$, SE = 0.013, $t = -1.31$, $p = 0.19$; standardized $\beta = -0.02$) significantly predicted prediction error.

Relative to interoceptive dominance, no dominant sensorimotor dimension differed significantly at the conventional $\alpha = 0.05$ level. However, auditory-dominant concepts showed a marginal tendency toward lower prediction error compared to interoceptive concepts ($\beta = -0.11$, SE = 0.062, $t = -1.75$, $p = 0.08$). None of the remaining perceptual (gustatory, haptic, visual, olfactory) or action-related (hand–arm, foot–leg, head, mouth, torso) dominance categories differed reliably from the interoceptive baseline (all $ps > 0.11$). Estimated marginal means on the response scale are reported in Table 2.

Discussion

This study extends prior evidence that sensorimotor information is related to concreteness by showing that graded concreteness can be recovered from multidimensional sensorimotor ratings using a nonlinear model. The main contribution is therefore not simply that an ANN predicts concreteness with good accuracy, but that jointly modeling mean strength and inter-rater variability (SD) across sensorimotor dimensions is sufficient to recover graded structure in concreteness judgments. In this sense, the findings suggest that the relation between sensorimotor grounding and concreteness is structured and multidimensional, and may not be fully captured by additive linear effects alone.

More specifically, the results indicate that sensorimotor information provides a structured basis for predicting concreteness despite lexical ambiguity and substantial interindividual variability. Notably, greater variability in concreteness ratings was associated with lower prediction error. This suggests that dispersion in ratings is not merely noise but reflects structured variability in how concepts are grounded and judged. One possible interpretation is that concepts with greater variability allow the model to rely on salient or prototypical sensorimotor features shared across partially overlapping instantiations. On this view, variability may index flexibility in how concepts are grounded across contexts and individuals, providing complementary information to mean ratings in psycholinguistic norms.

The results also speak to two issues that are often raised in work with decontextualized lexical items. First, lexical ambiguity did not predict model error, suggesting that the network's performance was not substantially driven by the number of senses associated with a word, and that participants converge on salient or representative sensorimotor features when rating both sensorimotor experience and concreteness. Second, interoceptive mean ratings had the strongest impact on performance in the feature permutation analysis, supporting the view that interoception plays an important role in the grounding of concepts, including along the concreteness–abstractness continuum (Banks & Connell, 2023; Diveica et al., 2025; Mazzuca et al., 2025). At the same time, concepts did not need to be dominantly interoceptive for interoception to matter, indicating that its contribution is distributed rather than confined to a single dominant profile.

These conclusions should be interpreted with caution. Because both predictors and outcomes are human rating norms, the model may reflect not only conceptual grounding, but also partial overlap in the criteria participants use to judge sensorimotor experience and concreteness. However, this overlap alone is unlikely to account for the present findings. The recovery of graded concreteness from multidimensional sensorimotor features suggests that the relationship is structured rather than trivial. Accordingly, the results do not imply that sensorimotor grounding causally determines concreteness, but show that sensorimotor ratings encode sufficient and systematically organized information to reconstruct graded concreteness judgments, even in the presence of lexical ambiguity.

Taken together, the present findings support the continued usefulness of the concreteness–abstractness distinction when treated as a graded and probabilistic construct. Although variability in concreteness ratings has been argued to challenge the interpretability of concreteness norms (Pollock, 2018), sensorimotor information captures systematic regularities across the scale. More broadly, these results are consistent with embodied and grounded accounts of conceptual representation and support a multidimensional, non-additive organization of concreteness judgments.

Future cross-linguistic work will be important for determining whether these patterns generalize beyond English, and whether concreteness ratings in one language can be predicted from sensorimotor norms in others. Establishing such generalization would provide insight into the extent to which conceptual representations are grounded in shared sensorimotor structures, and it would help clarify whether concreteness reflects universal properties of embodied experience or varies as a function of linguistic and cultural factors.

Acknowledgments

This work was funded by the National Autonomous University of Mexico (PAPIIT: IN304825, "Representación del conocimiento conceptual: análisis del contenido sensorimotor de los conceptos mediante herramientas de Inteligencia Artificial y experimentación con contextos oracionales").

The authors acknowledge the use of artificial intelligence tools to assist with drafting and revising portions of the manuscript, including paraphrasing and improving clarity and coherence. These tools were not used to generate data or make scientific decisions. All substantive contributions, interpretations, and conclusions remain the responsibility of the authors.

References

Arnold, T. B. (2017). Kerasr: R interface to the keras deep learning library. *Journal of Open Source Software*, *2*(14), 296. https://doi.org/10.21105/joss.00296

Banks, B., & Connell, L. (2023). Multi-dimensional sensorimotor grounding of concrete and abstract categories. *Philosophical Transactions of the Royal Society B*, *378*(1870), 20210366. https://doi.org/10.1098/rstb.2021.0366

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645. https://doi.org/0.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2020). Challenges and opportunities for grounding cognition. *Journal of Cognition*, *3*(1). https://doi.org/10.5334/joc.116

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

Borghi, A. M., & Mazzuca, C. (2023). Grounded cognition, linguistic relativity, and abstract concepts. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12663

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, *46*, 904–911. https://doi.org/10.3758/s13428-013-0403-5

Coltheart, M. (1981). The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*(3), 452–465. https://doi.org/10.1016/j.cognition.2012.07.010

Delignette-Muller, M. L., & Dutang, C. (2015). Fitdistrplus: An r package for fitting distributions. *Journal of statistical software*, *64*, 1–34. https://doi.org/10.18637/jss.v064.i04

Diveica, V., Muraki, E. J., Binney, R. J., & Pexman, P. M. (2025). Contrasting the organization of concrete and abstract word meanings. *Psychonomic Bulletin & Review*, 1–13. https://doi.org/10.3758/s13423-025-02671-z

Dolatabadi, M. (2023). Predicting english word concreteness through its multidimensional perceptual and action strength norms. *Australian Journal of Applied Linguistics*, *6*(3), 176–187. https://doi.org/10.29140/ajal.v6n3.1003

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412–431. https://doi.org/10.1016/j.cognition.2008.11.016

Dove, G. (2016). Three symbol ungrounding problems: Abstract concepts and the future of embodied cognition. *Psychonomic bulletin & review*, *23*, 1109–1121. https://doi.org/10.3758/s13423-015-0825-4

Dove, G. (2021). The challenges of abstract concepts. *Handbook of embodied psychology: Thinking, feeling, and acting*, 171–195. https://doi.org/10.1007/978-3-030-78471-3_8

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly journal of experimental psychology*, *61*(6), 825–850. https://doi.org/10.1080/17470210701623605

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage*, *32*(3), 1413–1421. https://doi.org/10.1016/j.neuroimage.2006.06.007

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 319–325. https://doi.org/10.1002/wcs.1226

Gallese, V. (2025). Language and bodily multimodality. the role of embodied simulation. *Sistemi intelligenti*, *37*(2), 165–182. https://doi.org/10.1422/117529

Good, P. (2013). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.

Löhr, G. (2022). What are abstract concepts? on lexical ambiguity and concreteness ratings. *Review of Philosophy and Psychology*, *13*(3), 549–566. https://doi.org/10.1007/s13164-021-00542-9

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An r package for assessment, comparison and testing of statistical models. *Journal of open source software*, *6*(60). 10.21105/joss.03139

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior research methods*, *41*(2), 558–564. https://doi.org/10.3758/BRM.41.2.558

Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior research methods*, *45*, 516–526. https://doi.org/10.3758/s13428-012-0267-0

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, *52*, 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Machery, E. (2007). Concept empiricism: A methodological critique. *Cognition*, *104*(1), 19–46. https://doi.org/10.1016/j.cognition.2006.05.002

Mazzuca, C., Fini, C., De Livio, C., Falcinelli, I., Maggio, F., Tummolini, L., & Borghi, A. M. (2025). Words as social tools (wat): A reprise. *Physics of Life Reviews*, *52*, 109–128. https://doi.org/10.1016/j.plrev.2024.12.011

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, *3*(4), 235–244. https://doi.org/10.1093/ijl/3.4.235

Ojala, M., & Garriga, G. C. (2010). Permutation tests for studying classifier performance. *Journal of machine learning research*, *11*(6). https://doi.org/10.3758/s13428-016-0720-6

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The calgary semantic decision project: Concrete/abstract decision data for 10,000 english words. *Behavior research methods*, *49*(2), 407–417. https://doi.org/10.3758/s13428-016-0720-6

Pollock, L. (2018). Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior Research Methods*, *50*(3), 1198–1216. https://doi.org/10.3758/s13428-017-0938-y

Raia, R. (2023). An analysis of conceptual ambiguities in the debate on the format of concepts. *Phenomenology and the Cognitive Sciences*, 1–26. https://doi.org/10.1007/s11097-023-09938-7

Reilly, J., Shain, C., Borghesani, V., Kuhnke, P., Vigliocco, G., Peelle, J. E., Mahon, B. Z., Buxbaum, L. J., Majid, A., Brysbaert, M., et al. (2025). What we mean when we say semantic: Toward a multidisciplinary semantic glossary. *Psychonomic bulletin & review*, *32*(1), 243–280. https://doi.org/10.3758/s13423-024-02556-7

Team, R. C., et al. (2016). R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. *http://www. R-project. org/*.

Tieleman, T., & Hinton, G. (2012). Lecture 6.5 - RMSProp: Divide the gradient by a running average of its recent magnitude [URL: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf].

Trott, S., & Bergen, B. (2022). Contextualized sensorimotor norms: Multi-dimensional measures of sensorimotor strength for ambiguous english words, in context. *arXiv preprint arXiv:2203.05648*. https://doi.org/10.48550/arXiv.2203.05648

Varela, F. J., Thompson, E., Rosch, E., & Kabat-Zinn, J. (2017). *The embodied mind*. The MIT Press.

Villani, C., Lugli, L., Liuzza, M. T., Nicoletti, R., & Borghi, A. M. (2021). Sensorimotor and interoceptive dimensions in concrete and abstract concepts. *Journal of memory and language*, *116*, 104173. https://doi.org/10.1016/j.jml.2020.104173

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*, 625–636. https://doi.org/10.3758/BF03196322