



# Using agreement probability to study differences in types of concepts and conceptualizers

Enrique Canessa<sup>1,3</sup> · Sergio E. Chaigneau<sup>1,2</sup> · Sebastián Moreno<sup>3</sup>

Accepted: 18 November 2022 / Published online: 5 December 2022  
© The Psychonomic Society, Inc. 2022

## Abstract

Agreement probability  $p(a)$  is a homogeneity measure of lists of properties produced by participants in a Property Listing Task (PLT) for a concept. Agreement probability's mathematical properties allow a rich analysis of property-based descriptions. To illustrate, we use  $p(a)$  to delve into the differences between concrete and abstract concepts in sighted and blind populations. Results show that concrete concepts are more homogeneous within sighted and blind groups than abstract ones (i.e., exhibit a higher  $p(a)$  than abstract ones) and that concrete concepts in the blind group are less homogeneous than in the sighted sample. This supports the idea that listed properties for concrete concepts should be more similar across subjects due to the influence of visual/perceptual information on the learning process. In contrast, abstract concepts are learned based mainly on social and linguistic information, which exhibit more variability among people, thus, making the listed properties more dissimilar across subjects. Relative to abstract concepts, the difference in  $p(a)$  between sighted and blind is not statistically significant. Though this is a null result, and should be considered with care, it is expected because abstract concepts should be learned by paying attention to the same social and linguistic input in both, blind and sighted, and thus, there is no reason to expect that the respective lists of properties should differ. Finally, we used  $p(a)$  to classify concrete and abstract concepts with a good level of certainty. All these analyses suggest that  $p(a)$  can be fruitfully used to study data obtained in a PLT.

**Keywords** Property listing task · Agreement probability · Concrete/abstract concepts · Sighted/blind populations

## Introduction

In the current work, we discuss a measure of conceptual homogeneity and illustrate its potential by using it to analyze differences between two sets of concepts and two populations. Our data was collected using a semantic Property Listing Task (PLT, Lenci et al., 2013), where people freely produce featural descriptions of a given concept. Consequently,

our measure of homogeneity quantifies, across participants, the average correspondence between the descriptions (i.e., properties/features) that were produced for a given concept. The more similar across participants the descriptions are, the greater the homogeneity.

Concepts are probably variable and not homogeneous across a population, and differences may exist even when people conceptualize a situation similarly. Philosophers have pointed out the theoretical difficulties in asserting that different people share strictly the same concepts (Frege, 1893; Glock, 2009; Russell, 1997). Empirical evidence also suggests that the situation is that there is non-homogeneity in how people instantiate a given concept, and that there is even non-homogeneity in how a single person instantiates the same concept in two different occasions (Barsalou, 1987, 1993). In the current work, we take this non-homogeneity to be a fundamental characteristic of naturally occurring concepts. A simple source of non-homogeneity is learning. When concepts are learned in natural environments (in contrast to experimental environments), the most likely situation is that people will

✉ Enrique Canessa  
ecanessa@uai.cl

<sup>1</sup> Center for Cognition Research (CINCO), School of Psychology, Universidad Adolfo Ibáñez, Av. Presidente Errázuriz 3328, Las Condes, Santiago, Chile

<sup>2</sup> Center for Social and Cognitive Neuroscience, School of Psychology, Universidad Adolfo Ibáñez, Av. Presidente Errázuriz 3328, Las Condes, Santiago, Chile

<sup>3</sup> Faculty of Engineering and Science, Universidad Adolfo Ibáñez, Av. P. Hurtado 750, Lote H, Viña del Mar, Chile

be exposed to different training sets, and so they will develop different versions of putatively the same concept.

In the current work, we will use the idea that concepts have variable instantiations as a guiding principle and offer a quantitative probabilistic measure of this non-homogeneity (to be explained shortly). To show one example of our measure's usefulness, we will use it to characterize the concrete versus abstract concept difference and also the differences between congenitally blind and sighted people when conceptualizing the same set of concepts. Both issues have been highly researched topics.

## Differences between concrete and abstract concepts

There is a large literature on the differences between abstract and concrete concepts. Our reading of the literature leads us to conclude that an essential difference between these types of concepts is that, while concrete concepts depend more on perceptual information than abstract concepts, abstract concepts rely to a large extent on social and linguistic input (for a good review of the evidence, see Borghi et al., 2019; for a critical view, see Willems & Casasanto, 2011). Importantly, in our analysis, perceptual information is predicted to introduce greater homogeneity in the semantic properties produced, leading to concrete concepts being more homogeneous than abstract concepts.

Compared to abstract concepts, concrete concepts are easier to learn and process (e.g., Jones, 1985; Walker & Hulme, 1999), are characterized by a larger number of conceptual features (Plaut & Shallice, 1991, 1993), and are more closely related to specific contexts (Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983). A summary of all this research might be that semantic memory (SM) is more densely structured for concrete than for abstract concepts (Jones, 1985; Plaut & Shallice, 1993; Recchia & Jones, 2012; Yap & Pexman, 2016). A richer semantic structure would make concrete concepts easier to access. In contrast, having a less densely structured representation in memory is coherent with abstract concepts having more different senses (Hoffman et al., 2013).

In addition to these differences in semantic richness and context dependence, and as foreshadowed at the beginning of this section, several authors have proposed that the difference between concrete and abstract concepts hinges on the type of features that corresponds to each type of concept. In this view, concrete concepts depend on perceptual content, while abstract concepts depend on linguistic information (e.g., while *dog* may be described by “barks,” “has four legs,” and “is hairy,” *justice* may be described by “fairness” and “law;” Barsalou et al., 2008; Breedin et al., 1994; Paivio, 1986; Wiemer-Hastings & Xu, 2005).

This view is in line with the proposal that conceptual processing involves reactivating perceptual representations (Barsalou, 1999; Feldman, 2010; Gallese & Lakoff, 2005; Prinz, 2002; Pulvermüller, 2005).

Previous studies provide evidence consistent with the idea that people reactivate perceptual features during language comprehension (Lupyan & Ward, 2013; Ostarek & Huettig, 2017), during property verification (i.e., Is *y* a property of concept *x*? Kan et al., 2003; Solomon & Barsalou, 2004), and during semantic property listing (Santos et al., 2011). Consequently, in the current work we hypothesize that concrete concepts are characterized more by perceptual information than abstract concepts, and that this perceptual information introduces a greater homogeneity in conceptualization for concrete versus abstract concepts. For expository purposes, we will call these our *characterizing concreteness hypotheses*.

## Differences in semantic representations between congenitally blind and sighted individuals

As previously discussed, it is possible that concrete concepts are characterized by having more perceptual content than abstract concepts, and that this perceptual content may introduce a greater homogeneity in conceptual representations and processing. If these hypotheses are correct, then they suggest that we should find that congenitally blind individuals, because they lack visual perceptual information, should show differences when processing concrete concepts, but not when processing abstract concepts, which seem to depend more on linguistic and social input (Borghi et al., 2017; Borghi & Cimatti, 2009). For expository purposes, we will call this our *role of vision hypothesis*.

There is in fact previous evidence consistent with our hypothesis that congenitally blind subjects should process concrete concepts differently. Blind individuals show differences in performance relative to sighted individuals when visual information (e.g., color) is critical for judgments (Connolly et al., 2007). Similarly, Kim et al. (2019) found that though blind subjects used general-purpose inferential mechanisms to acquire knowledge about appearances (e.g., that all birds have feathers), they showed systematic differences relative to sighted people when judging similarity based on shape, knowledge that is highly dependent on vision (i.e., choosing the dissimilar item in a triad odd-one-out paradigm, for example, choosing the different animal in the wolf, gorilla, and bear triad).

However, our hypothesis might not be correct. There is a fair amount of evidence, suggesting that conceptual representations are strikingly similar in sighted and blind subjects (Landau & Gleitman, 1985; Marmor, 1978; Zimler & Keenan, 1983) and that though there are some detectable

differences in early development, language acquisition and use is remarkably resilient to the lack of visual input (Pérez-Pereira, 2006). It is likely that congenitally blind subjects can use statistical regularities in language experienced in their communities (Erickson & Thiessen, 2015; Steyvers & Tenenbaum, 2005) to acquire knowledge of semantic relations, even when they do not have direct access to the perceptual information that underlies those statistical regularities (e.g., they know that *zebra* and *penguin* are similar in that they are “black” and “white,” even if they have never had the corresponding visual experiences). Thus, it is an open question whether our hypothesis about processing differences between blind and sighted subjects should hold or not.

### Agreement probability as a measure of homogeneity

The semantic PLT is a procedure widely used in psychology to obtain property-based descriptions of concepts coded in language (Cree & McRae, 2003; Hampton, 1979; McRae et al., 2005; Rosch et al., 1976). Though there are slight differences in the way the task is implemented by different researchers, the general procedure is to ask subjects to produce properties that are typically true of a given concept. Once lists are obtained, they are generally coded into property types (i.e., responses with only superficial differences across subjects are coded as a single property) and accumulated across participants to obtain property frequency distributions. When the PLT is used to collect properties across whole semantic fields, the resulting data can be organized in Conceptual Property Norms (CPNs, e.g., Devereux et al., 2014; Kremer & Baroni, 2011; Lenci et al., 2013; McRae et al., 2005; Montefinese et al., 2013; Vivas et al., 2017).

As a way of measuring homogeneity in the PLT, here we compute *agreement probability* ( $p(a)$ ; Chaigneau et al., 2012), which will be explained in detail in the next section. Conceptually, agreement probability ( $p(a)$ ) is defined as the probability that one property taken randomly from one list produced by an average subject in a PLT is also found in another list produced by a different average subject for the same PLT. By average subject, here we mean a hypothetical participant who on average represents the lists generated across participants, and thus, not a specific individual who produced a particular list. Lists may come from the same concept (the two lists were produced for the same concept  $C1$ ) or from different concepts (the two lists were produced for two different concepts  $C1$  and  $C2$ ). It is called agreement probability because it is a measure of the agreement in the properties being listed. The maximum agreement will be produced when all subjects produce the same list (same properties, same length). In that case,  $p(a) = 1$ . The minimum agreement will be produced when all

subjects produce different lists (different properties, not necessarily different lengths).

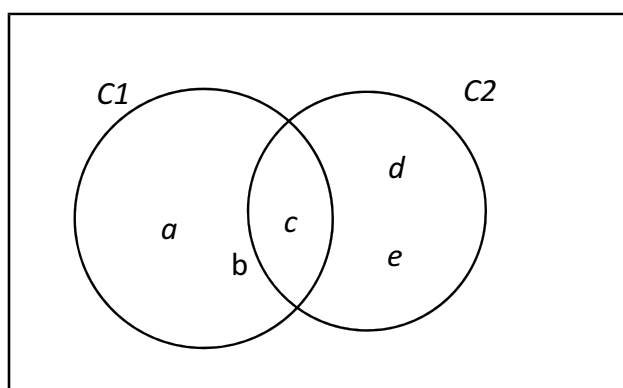
Quantifying homogeneity by using  $p(a)$  is important given that the instantiation of a concept, and thus the properties with which people describe it, depends on multiple factors. Hence,  $p(a)$  may be used as a measure of how sensitive a concept is to those multiple factors, whichever they are. On different occasions, concepts can be instantiated differently (e.g., the concept *to jump* may be instantiated differently in the context of “extreme sports” from the context of “children”). It is likely that concepts are sensitive to the contexts in which they occur in terms of how frequently a given context is associated with a given concept (e.g., *bill* occurs more frequently at a restaurant and less frequently at a beach). It is also likely that concepts are sensitive to context in terms of different senses being associated with different contexts (e.g., *bill* adopts a different sense in the context of restaurant than in the context of government). These factors are likely to introduce non-homogeneity in concepts (i.e., lack of agreement in lists being produced) because people may adopt different points of view when producing property lists after having been cued with a given concept. Other individual factors may also introduce lack of agreement (e.g., subjects being influenced by recent events in memory, or by idiosyncrasies in how a concept was learned or is processed). Therefore, agreement can be interpreted as the degree to which a concept is independent from all those factors, where the higher the  $p(a)$  for a concept, the more independent the concept is from all those factors.

Note that, because homogeneity in conceptualization might be influenced by multiple factors (as discussed immediately above), if measured in a different task, different homogeneity estimates could be obtained. For example, in conversation, it is possible that people will progress to higher agreements due to their history of interactions (Fay et al., 2018). However, and to the best of our knowledge, there is no similar measure in the literature, and we hypothesize that, though other measures could produce different estimates, the results we report here should hold, at least in relative terms. Additionally,  $p(a)$  has the advantage that it makes use of and summarizes/aggregates in a unique indicator, information that is routinely obtained in PLTs, such as the average length of property lists being produced, the total number of unique properties produced for a given concept by a group of subjects, and the property frequency distributions found in a CPN. These advantages will be better appreciated when we present the mathematical properties of  $p(a)$ .

As will be discussed below, computing  $p(a)$  from frequency distributions of conceptual properties involves a combinatorial problem, which makes it impractical to use combinatorial formulae to compute it. Instead, we resort to computational simulations that deliver a close estimate for  $p(a)$ .

### Computing and interpreting the meaning of agreement probability

To understand agreement probability, consider the following simple example. By asking people to produce conceptual properties for two related concepts or two versions of the same concept ( $C1, C2$ ), two property frequency distributions are obtained. For concept  $C1$ , subjects produced properties  $a, b, c$ . For concept  $C2$ , subjects produced properties  $c, d, e$ . This situation is shown in Fig. 1, where for example  $C1 = dog$  and  $C2 = cat$ . Note that to simplify the example, these are equiprobable distributions (i.e., properties in each distribution occur the same number of times). Assume now that subjects produced samples of average size = 2 for  $C1$



**Fig. 1** Two concepts  $C1$  and  $C2$  with their corresponding set of properties obtained in a PLT (for  $C1 = \{a, b, c\}$  and  $C2 = \{c, d, e\}$ ) and intersection ( $C1 \cap C2 = \{c\}$ ), with  $k_1$  (number of  $C1$ 's properties) = 3,  $k_2$  (number of  $C2$ 's properties) = 3,  $u$  (number of properties in the intersection) = 1

(the average number of properties mentioned by people for  $C1$  is 2,  $s_1 = 2$ ) and also 2 for  $C2$  ( $s_2 = 2$ ). Imagine that for concept  $C1$ , subjects produced the following properties:  $a =$  it barks,  $b =$  wags its tail,  $c =$  has four legs and for  $C2$ :  $c =$  has four legs,  $d =$  it meows, and  $e =$  catches mice.

According to conceptual agreement theory (CAT, Chaigneau et al., 2012) agreement probability  $p(a)$  is the probability that one property randomly chosen from a sample of size  $s_2$  of properties extracted from the set of all  $k_2$  properties that are listed in a PLT for a concept  $C2$ , is contained in a sample of size  $s_1$  randomly obtained from the set of all  $k_1$  properties that are listed in a PLT for a concept  $C1$ . CAT's mathematical formulation allows calculating  $p(a)$  using expression (1), where Table 1 defines each of the variables:

$$p(a) = \frac{1}{s_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \#(S_i^1 \cap S_j^2) p_i q_j \tag{1}$$

Equation (1) is the summation of the expected value of the number of common elements between samples  $S_i^1$  of properties listed for  $C1$  and independent samples  $S_j^2$  of properties listed for  $C2$  (i.e., the  $\#(S_i^1 \cap S_j^2)$  term), taking into account the probabilities of each sample (i.e., the  $p_i$  and  $q_j$ ). Assuming that all properties in the  $S_i^1$  and  $S_j^2$  samples have the same probability of being obtained,  $p(a)$  is calculated as the summation in Eq. (1) divided by  $s_2$ . The interested reader may find the complete mathematical and theoretical development of  $p(a)$  in Chaigneau et al. (2012). Here we just present the most important details necessary to understand the present work. Instead, to aid the reader in comprehending Eq. (1) and the definitions in Table 1, we present a simple example, which illustrates the application of such expressions.

Following the previous example, we have that:

**Table 1** Definition of variables used in expressions for calculating  $p(a)$

Variable's symbol	Definition
$s_1$	The average size of a sample of properties obtained from concept $C1$ ( $s_1 \leq k_1$ ), where $k_1$ is the cardinality of the set of unique properties listed for concept $C1$
$n_1$	The number of possible samples of size $s_1$ obtained from the $k_1$ properties that belong to the set of unique properties listed for concept $C1$ . Thus, $n_1 = \frac{k_1!}{(k_1-s_1)!s_1!}$
$p_i$	Probability of obtaining sample $S_i^1$ in Eq. (1)
$s_2$	The average size of a sample of properties obtained from concept $C2$ ( $s_2 \leq k_2$ ), where $k_2$ is the cardinality of the set of unique properties listed for $C2$
$n_2$	The number of possible samples of size $s_2$ obtained from the $k_2$ properties that belong to the set of unique properties listed for concept $C2$ . Thus, $n_2 = \frac{k_2!}{(k_2-s_2)!s_2!}$
$\#(S_i^1 \cap S_j^2)$	The cardinality of the intersection between a sample “ $i$ ” drawn from the set of unique properties listed for concept $C1$ ( $S_i^1$ ) and another independent sample “ $j$ ” drawn from the set of unique properties listed for concept $C2$ ( $S_j^2$ ), i.e., the number of properties that belong to both samples
$q_j$	Probability of obtaining sample $S_j^2$ in Eq. (1)
$u$	Number of common properties between the set of unique properties listed for concept $C1$ and $C2$ . This variable is used in Eq. (3)

Properties listed for  $C1 = \{a, b, c\}$

$k_1 = 3$  and assuming  $s_1 = 2$

$u = 1$  (one common element, i.e.,  $c$ ) and thus,

$n_1 = n_2 = 3!/(3 - 2)!/2! = 3$

Properties listed for  $C2 = \{c, d, e\}$

$k_2 = 3$  and assuming  $s_2 = 2$

the  $n_1$  samples from the properties listed for  $C1$  are  $\{ab, ac, bc\}$  ( $S_i^1 \in \{ab, ac, bc\}$ ) and the  $n_2$  samples from the properties listed for  $C2$  are  $\{cd, ce, de\}$  ( $S_j^2 \in \{cd, ce, de\}$ )

For simplicity, assume that each sample in  $C1$  and in  $C2$  has an equal probability of being selected and thus,  $p_i = 1/n_1 = 1/3$  and  $q_j = 1/n_2 = 1/3$ . Then, using Eq. (1):

$$p(a) = \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \#(S_i^1 \cap S_j^2) \frac{1}{33} = \frac{1}{18} \sum_{i=1}^3 \sum_{j=1}^3 \#(S_i^1 \cap S_j^2) \tag{2}$$

In Eq. (2), the double summation corresponds to the sum of counts of coincidences between each sample  $S_i^1$  and  $S_j^2$ , for example:

$$\begin{aligned} \#(S_1^1 \cap S_1^2) &= \#(ab \cap cd) = 0 \\ \#(S_1^1 \cap S_2^2) &= \#(ab \cap ce) = 0 \text{ and so on until,} \\ \#(S_3^1 \cap S_3^2) &= \#(bc \cap de) = 0 \end{aligned}$$

For this example, each term of the double summation is  $\#(S_i^1 \cap S_j^2) = \{0,0,0,1,1,0,1,1,0\}$  and hence  $p(a) = 4/18 = 2/9$

Probability  $p(a)$  tells us that for individuals who have listed properties for the concept  $C1$  (e.g., *dog*) and  $C2$  (e.g., *cat*), there exists a  $2/9$  probability that if one average participant listed a given property for the concept  $C2$  (e.g., *cat*), that same property will be in the list of properties listed by another different average participant for  $C1$  (e.g., *dog*). Several things are noteworthy. Note that first,  $p(a)$  is a measure of homogeneity because maximal homogeneity will be achieved when all participants in a PLT produce the same list, and minimal homogeneity will be obtained when all participants produce different lists.

Second, the reader may have noted that we assumed that the frequency distribution of the properties is uniform (i.e., properties in the distribution occur the same number of times, and thus  $p_i, q_j$  are the same for all  $i$  and  $j$ , see Eq. (1) and definitions in Table 1). This is an idealized case, and it is highly unlikely that real data would ever produce it. However, idealized models may have the virtue of reducing a problem to its essential characteristics. For such a case, we can demonstrate that (see Appendix A):

$$p(a) = \frac{s_1 u}{k_1 k_2} \tag{3}$$

where  $s_j$  is the average number of properties in a group member’s sample of conceptual content for concept  $C1$  and  $k_j$  is the total number of properties listed at least once for  $C1$ ,

$u$  is the number of common properties between the properties’ distributions of two concepts ( $C1$  and  $C2$ ), and  $k_2$  is the total number of properties listed at least once for concept  $C2$ . Thus,  $p(a)$  is a measure of how well separated two distributions are. This probability depends on the number of common properties between two distributions ( $u$ ). Note that for our simple example above, Eq. (3) necessarily gives the same result as Eq. (2) ( $p(a) = s_1/k_1 \times u/k_2 = 2/3 \times 1/3 = 2/9$ ).

Third, note that  $p(a)$  is not symmetric with respect to concepts  $C1$  and  $C2$ , i.e., calculating  $p(a)$  for concept  $C1$  relative to  $C2$  does not necessarily give the same  $p(a)$  as if we were computing it for concept  $C2$  relative to  $C1$ . Equations (1) and (3) give  $p(a)$  when concept  $C1$  is the reference concept and  $C2$  the comparison concept (i.e.,  $p(a)$  for  $C1$  relative to  $C2$ ). On the other hand, when concept  $C2$  is the reference and  $C1$  is the comparison concept, the expressions are similar, but using  $s_2$  and  $k_2$  and  $k_1$  in (3); i.e.,  $p(a) = s_2/k_2 \times u/k_1$ . This asymmetry tells us that the probability that a property contained in a sample for concept  $C2$  is also obtained in another sample of properties for concept  $C1$  is not necessarily the same as the probability that a property contained in a sample for concept  $C1$  is also obtained in another sample of properties for concept  $C2$ . This asymmetry is an important fact to remember when analyzing  $p(a)$  for concrete and abstract concepts, as well as for blind versus sighted subjects, which we will further use and explain in the corresponding analyses.

Finally, and as already stated,  $p(a)$  may be also computed for the same concept. In this case, there is only one concept  $C1$  and  $p(a)$  is the probability that one property randomly chosen from a sample of size  $s_1$  of properties extracted from the set of all  $k_1$  properties that are listed for a concept  $C1$ , is contained in a different sample of size  $s_1$  randomly obtained from the set of all  $k_1$  properties that are listed for the same concept  $C1$ .

Thus, the same expression (1) and definitions in Table 1 apply, but  $s_1 = s_2, k_1 = k_2, n_1 = n_2, p_i = q_j$ , and samples  $S_i^1$  and  $S_j^2$  are drawn from the same distribution of properties of concept  $C1$ . Hence, based on Eq. (1) and taking into account that now we are computing  $p(a)$  for the same concept  $C1$ , we can write:

$$p(a) = \frac{1}{s_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \#(S_i^1 \cap S_j^1) p_i p_j \tag{4}$$

Note that in Eq. (4) the samples  $S_i^1$  and  $S_j^1$  both have superscript 1, which indicates that they are independently



drawn from the same distribution of properties of concept *CI*. Additionally, note that we replaced  $q_j$  by  $p_j$  so that it is clearer that those probabilities correspond to samples drawn from the same distribution of properties.

With regard to computing  $p(a)$  for the same concept and for uniform property frequency distributions, Eq. (4) becomes:

$$p(a) = \frac{s_1}{k_1} \tag{5}$$

because in Eq. (3) and for the same list of properties for concept *CI*, it will always happen that  $u = k_2$ , i.e., for the same list of properties obtained for a concept, the number of common elements will be the same as the number of properties obtained for the concept (see Appendix A for a more formal demonstration). That fact also tells us that for concepts with uniform property frequency distributions,  $p(a)$  for two different concepts calculated using Eq. (3) will always be lower than  $p(a)$  computed for one of those concepts using Eq. (4), i.e., agreement probability for two different concepts will always be lower than  $p(a)$  for one of the concepts with itself (see Appendix A for a demonstration).

To help understand the computation of  $p(a)$  for the same concept, let's use the same example shown in Fig. 1 and calculate  $p(a)$  for *CI*. Then we have that:

Properties listed for *CI* = {*a, b, c*}  
 $k_1 = 3$  and assuming  $s_1 = 2$   
 $n_1 = 3! / (3 - 2)! / 2! = 3$

the  $n_1$  samples from the properties listed for *CI* are {*ab, ac, bc*} ( $S_i^1 \in \{ab, ac, bc\}$  and  $S_j^1 \in \{ab, ac, bc\}$ )

For simplicity, assume that each sample in the properties listed for *CI* has an equal probability of being selected and thus,  $p_i = p_j = 1/n_1 = 1/3$ .

And thus applying those values to Eq. (4):

$$p(a) = \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \#(S_i^1 \cap S_j^1) \frac{11}{33} = \frac{1}{18} \sum_{i=1}^3 \sum_{j=1}^3 \#(S_i^1 \cap S_j^1) \tag{6}$$

In Eq. (6), the double summation corresponds to the sum of counts of coincidences between each sample  $S_i^1$  and  $S_j^1$ , for example:

$\#(S_1^1 \cap S_1^1) = \#(ab \cap ab) = 2$   
 $\#(S_1^1 \cap S_2^1) = \#(ab \cap ac) = 1$  and so on until,  
 $\#(S_3^1 \cap S_3^1) = \#(bc \cap bc) = 2$

For this example, each term of the double summation is  $\#(S_i^1 \cap S_j^1) = \{2, 1, 1, 1, 2, 1, 1, 1, 2\}$  and hence  $p(a) = 12/18 = 2/3$

Note that  $p(a) = 2/3$  is the same as the one computed by using Eq. (5), i.e.,  $p(a)$  for the same concept =  $s_1/k_1 = 2/3$ . As easily seen from Eq. (5),  $p(a)$  is a measure of conceptual

homogeneity in a group that uses the concept that produced it. Its maximal value is reached only when all group members produce the same set of properties for the concept in question (i.e., when  $s_j = k_j$ ). In contrast, its minimal value is approached when each group member produces unique properties. Also see that  $p(a)$  for concept *CI* relative to *C2* (2/9) is lower than  $p(a)$  for the same concept *CI* (2/3), or for the same concept *C2* =  $s_2/k_2 = 2/3$ .

One last feature that is interesting to note is that  $p(a)$  for the same concept computed with Eq. (5) provides a lower bound for this probability regardless of a distribution's statistical structure (see Appendix A). In other words,  $p(a)$  for the same concept cannot reach a value lower than  $s_j/k_j$ . A direct consequence of this is that statistical structure in property frequency distributions (i.e., nonuniformity) will in general increase homogeneity, which is intuitively correct.

As discussed in Canessa and Chaigneau (2016), computing agreement probability from frequency distributions of conceptual properties involves a combinatorial problem. As shown in Eq. (1), it requires counting coincidences among pairs of samples weighted by their respective probabilities, where a sample means the conceptual content (i.e., properties) provided by an average individual contributing data to the distribution. That equation has the advantage of being formulated for the general case of nonuniform property frequency distributions, but the number of samples (combinations) that need to be taken into account rapidly grows as  $s_1, s_2$  and/or  $k_1, k_2$  increase. For example, for a realistic PLT, a concept may have  $s_j = 7$  and  $k_1 = 30$ , and thus  $n_j = 2,035,800$  (see Table 1 for the expression that calculates  $n_j$ ). That makes expression (1) impossible to use in real PLTs. Therefore, in Canessa and Chaigneau (2016) we present a simulator that emulates the property comparison process underlying  $p(a)$  (i.e., counting the number of times in which a property found in a randomly selected sample is also found in a second randomly selected sample, over the total number of selected samples) and that allows calculating that probability with non-statistically significant differences relative to the exact values that might be computed using Eq. (1). For the detailed simulator's algorithm, the interested reader may consult Canessa and Chaigneau (2016). Here we briefly describe it, so that the parameters that must be inputted to the simulator and will be used in this paper are understood.

The simulator receives the property frequency distribution of properties listed for a concept *CI* and *C2* and their corresponding  $s_1$  and  $s_2$  values. First, the simulator probabilistically gets one independent sample of size  $s_1$  properties without replacement from the properties listed for *CI* and another sample of size  $s_2$  properties without replacement from the properties listed for *C2*. We label the first sample as reference sample and the second one as comparison sample. Note that the sampling probability of each property corresponds to the frequency of the given property relative to the

frequencies of the other properties. In our simple example, if e.g., the frequencies of property  $a = 15$ ,  $b = 20$ , and  $c = 10$  in concept  $CI$ , then, the probability of sampling  $a$  is  $15 / (15 + 20 + 10) = 1/3$  and similarly for  $b = 20/45 = 4/9$  and for  $c = 10/45 = 2/9$ . Then, it randomly selects one property from the comparison sample and if that property is contained in the reference sample, it increments a `pa_counter`. This is done `max_iterations` times. Then,  $p(a)$  is simply approximated by `pa_counter / max_iterations`. Additionally, the simulator has two more inputs. Given that as the simulator iterates, the approximation gets closer to the real value of  $p(a)$  (in fact as `max_iterations` tends to infinity, the approximation reaches the true  $p(a)$  value), we can calculate a moving average of  $p(a)$  as the simulator iterates by using the last `nr_points_moving_avg` iterations. Finally, we can repeat the simulation for `nr_repetitions` times and calculate a mean and standard deviation of  $p(a)$  using each of the values computed in the individual repetitions. This simulator was implemented in NetLogo v. 6.2.1 (Wilensky, 1999) and is available at [https://osf.io/xhfmz/?view\\_only=31c08caa642f42c694425a4f2b46a8b4](https://osf.io/xhfmz/?view_only=31c08caa642f42c694425a4f2b46a8b4) along with data files and instructions on how to use the simulator. For this work, the simulator's parameters were set as follows: `max_iterations = 5000`, `nr_points_moving_avg = 1000` and `nr_repetitions = 50`. The property frequency distributions for each concept may be found in the abovementioned URL and were obtained from Lenci et al. (2013) norms for concrete and abstract concepts, and for sighted and blind individuals, as described in the next section.

## Difference in agreement probability between concrete and abstract concepts, and between sighted and blind individuals

### Participants and data collection procedures

To show one example of the application of agreement probability as a measure of homogeneity in lists produced by subjects in a PLT, we resorted to data collected in Lenci et al. (2013) norms, which report properties for 70 concepts (50 abstract and concrete nouns, and 20 verbs). The Lenci et al. data are freely available on the web. In this work, we use the concrete ( $N_C = 40$ ) and abstract ( $N_A = 10$ ) nouns, which were classified as such in Lenci et al. (2013). Here we provide just the most important details of Lenci et al. (2013) norms; for more particulars see the corresponding paper. Appendix B shows the 40 concrete and 10 abstract nouns. Concrete nouns cover living and nonliving things, most of which were already used in previous norms (Kremer & Baroni, 2011; McRae et al., 2005) or by experiments with blind subjects (Connolly et al., 2007). These concrete concepts included things with salient visual features (e.g., “stripes”

for *zebra*; “yellow” for *banana*). Abstract concepts included emotions (e.g., *jealousy*) and ideals (e.g., *freedom*). Forty-eight Italian subjects ( $N = 48$ ), 22 congenitally blind ( $N_B = 22$ ) and 26 sighted ( $N_S = 26$ ), were included in the study, all of them native Italian speakers. The blind participants were 10 females and 12 males with an average age of 47.2 years (s.d. = 16.5) and with education ranging from junior high school to a master's degree. The 26 sighted participants were selected to match blind subjects as close as possible regarding age, gender, residence, education, and profession. Sighted subjects' average age was 45.1 years (s.d. = 16.8). Subjects were instructed to orally describe the concepts with short phrases and listened to the concepts in random order. To avoid too much fatigue, the 70 concepts were split into two separate sessions, and each session contained a 5-minute break at the middle of it. The entire procedure was done on a laptop, and the oral responses were recorded in digital audio. The oral responses were translated to text using an automated software program. The text was then coded by a trained coder using standard coding procedures (Kremer & Baroni, 2011; McRae et al., 2005).

### Relating visual perceptual strength to agreement probability

According to our *characterizing concreteness hypotheses*, concrete concepts should be characterized by more perceptual information than abstract concepts, and this perceptual information should introduce a greater homogeneity in conceptualization for concrete versus abstract concepts. To test these hypotheses, we resorted to the perceptual modality norms in Vergallito et al. (2020). In those norms, 57 sighted Italian participants rated concepts for their perceptual strength in each of five sensory modalities (i.e., vision, touch, smell, hearing, taste). Subjects were asked to rate, on a scale of 1 to 5, to which extent a given concept was experienced through each of these senses (e.g., the concept *sweet* may receive a high rating for taste and lower for the other modalities). A total of 20 concepts (15 concrete and 5 abstract) in the Vergallito et al. (2020) norms were also present in the Lenci et al. (2013) norms, so we used them in this analysis (see those concepts in Appendix B). Due to our emphasis in the current work on the visual modality, we only used those ratings. As predicted, those 15 concrete concepts showed significantly higher visual strength ratings ( $M = 4.8$ , s.d. = .08) than the five abstract concepts ( $M = 3.3$ , s.d. = .23) ( $t(4.32) = 14.249$ ,  $p < .001$ ; adjusted for unequal variances,  $F = 8.55$ ,  $p = .009$ ). The observed statistical power for this test for an  $\alpha = 0.05$  is above 0.99, and thus, despite the rather small sample size used, this result is reliable.

Regarding our use of only visual perceptual strength, we must note that it is also possible that other perceptual

information (e.g., olfactive, haptic, etc.) would also introduce homogeneity in property listing. Though we believe this is an interesting problem that could be tackled by our measure, it is well beyond the scope of the current work and we defer it for future work.

Regarding our second *characterizing concreteness hypothesis*, data also supported it. Using the  $p(a)$  simulator described in [Computing and interpreting the meaning of agreement probability](#) and the concepts' property frequency distributions obtained from the Lenci et al. (2013) norms, we computed agreement probability for the 15 concrete and 5 abstract concepts for which the Vergallito et al. (2020) norms provided perceptual strength ratings. As predicted, our  $p(a)$  measure positively correlated with visual strength ratings. For sighted subjects the correlation is 0.59 ( $r(20) = .59$ ,  $t(18) = 3.100$ ,  $p = .006$ , observed statistical power for an  $\alpha = 0.05$  equal to 0.87) and for blind subjects it is 0.62 ( $r(20) = .62$ ,  $t(18) = 3.353$ ,  $p = .004$ , observed statistical power for an  $\alpha = 0.05$  equal to 0.92). These positive and statistically significant correlations show that the higher/lower  $p(a)$ s exhibited by concrete/abstract concepts are associated with higher/lower visual strength ratings, which is consistent with the hypothesized homogenizing effect of visual perceptual information on property listing for concrete concepts relative to abstract ones. Also see that the high statistical power attained by those tests suggest that those results are reliable, and are not spurious findings due to underpowered comparisons. Finally, somewhat surprisingly, the correlation between visual perceptual strength and  $p(a)$  is statistically significant for blind subjects, which suggests that visual properties have a homogenizing effect on lists produced by those participants, even though they cannot directly perceive them. We will further elaborate on this issue in the Discussion section.

Additionally, for the 15 concrete and 5 abstract concepts used here, for sighted subjects,  $p(a)$  for concrete concepts ( $M = 0.17$ ,  $s.d. = .03$ ) is higher than for abstract ones ( $M = 0.11$ ,  $s.d. = .03$ ) ( $t(18) = 3.352$ ,  $p = .004$ , observed statistical power for an  $\alpha = 0.05$  equal to 0.91). The same happens for blind, where  $p(a)$  is higher for concrete concepts ( $M = 0.15$ ,  $s.d. = .03$ ) than for abstract concepts ( $M = 0.11$ ,  $s.d. = .02$ ) ( $t(18) = 3.565$ ,  $p = .002$ , observed statistical power for an  $\alpha = 0.05$  equal to 0.85). As we will show in the next subsection, this result agrees with the more general conclusion for the 50 concrete and abstract concepts in the Lenci et al. (2013) norms.

We acknowledge that, because our results are based on subjective ratings of perceptual strength, other explanations are possible. However, we believe that the results we report next provide converging evidence in support of our explanation, so we defer discussing alternative accounts for our [Discussion and conclusions](#). Thus, we proceed now to test our *role of vision hypothesis*. Recall that this hypothesis predicts that lacking visual perceptual information would make concrete concepts less homogeneous for blind subjects than for sighted

participants, reducing the difference in homogeneity between concrete and abstract concepts for a blind population.

### Comparing agreement probability between concrete and abstract concepts for sighted and blind subjects

Using the  $p(a)$  simulator and the concepts' property frequency distributions in Lenci et al.'s (2013) norms, we computed agreement probability for concrete and abstract concepts, within sighted and blind participants. Additionally, recall from our discussion in [Computing and interpreting the meaning of agreement probability](#), that agreement probability can be calculated for the same concept (i.e., a single property frequency distribution) or for two different concepts or versions of the same concept (i.e., two different distributions obtained from different concepts or from two different samples or populations). Given that we have sighted and blind individuals who separately listed properties for the same set of concepts, we have two different property frequency distributions: one for sighted (S) and another for blind (B). Hence,  $p(a)$  may be separately computed using the S distribution and the B distribution, i.e., separately inputting to the simulator S and then B. Those  $p(a)$ s will quantify the agreement probabilities within the sighted group (here we label it:  $S \rightarrow S$ ) and within the blind group ( $B \rightarrow B$ ). We can also compute other two  $p(a)$ s: an intergroup (between-groups) agreement probability from sighted to blind ( $S \rightarrow B$ ) and from blind to sighted ( $B \rightarrow S$ ). Thus, Table 2 presents the results of a two-way analysis of variance (ANOVA) (Type of concept  $\times$  Condition:  $S \rightarrow S$ ,  $B \rightarrow B$ ,  $S \rightarrow B$ ,  $B \rightarrow S$ ).

From Table 2 we can see that the model as a whole is statistically significant, and  $p(a)$  may differ for some comparisons between concrete and abstract concepts and between sighted and blind subjects. Also, there is a significant interaction between those two factors. Note also that the observed statistical power for an  $\alpha = 0.05$  for all the ANOVA results are high and hence, the corresponding results are reliable. Thus, we may now compare and analyze the mean of the eight treatments or cells of the ANOVA. To help visually assess those comparisons, Fig. 2 shows the mean  $p(a)$  and a 95% CI for the eight treatments.

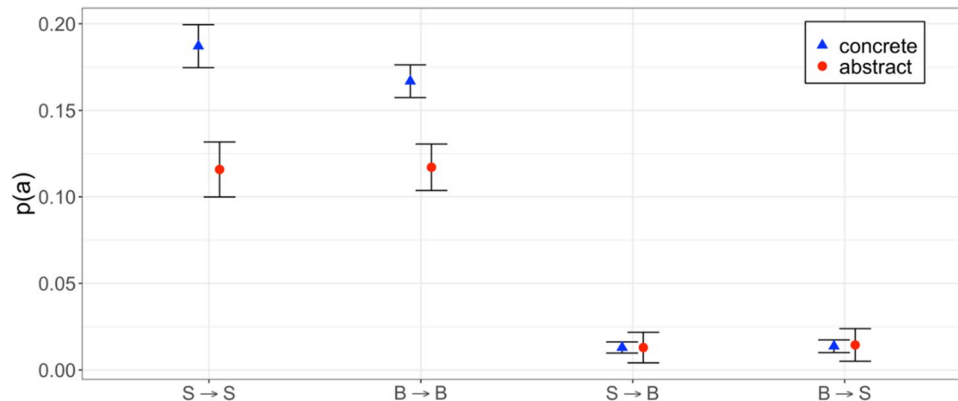
From Fig. 2 we can see that  $p(a)$  is higher for concrete than for abstract concepts for the  $S \rightarrow S$  ( $t(48) = 5.612$ ,  $p < .001$ ) and  $B \rightarrow B$  conditions ( $t(48) = 5.114$ ,  $p = .001$ ) (i.e., within groups). Our results for visual perceptual strength lead us to interpret this as showing that concrete concepts show more homogeneity due to the influence of visual/perceptual information, while abstract concepts are in general less homogeneous due to the influence of social and linguistic information.

Also, from Fig. 2 we can see that  $p(a)$  for concrete concepts and for condition  $S \rightarrow S$  is higher than for conditions  $B \rightarrow B$  ( $t(78) = 2.655$ ,  $p = .01$ ),  $S \rightarrow B$  ( $t(78) = 27.790$ ,  $p < .001$ ),



**Table 2** ANOVA for agreement probabilities for concrete (C) and abstract (A) concepts, and for conditions  $S \rightarrow S$ ,  $B \rightarrow B$ ,  $S \rightarrow B$ ,  $B \rightarrow S$ 

Source	F-test	p-value	Eta-squared	Observed power for $\alpha = 0.05$
Model	$F(7, 192) = 302.735$	<.001	0.917	> .99
Intercept	$F(1, 192) = 1433.623$	<.001	0.882	> .99
Type-concept (C, A)	$F(1, 192) = 50.510$	<.001	0.208	> .99
Condition (SS, BB, SB, BS)	$F(3, 192) = 331.287$	<.001	0.838	> .99
Type-concept $\times$ Condition	$F(3, 192) = 18.317$	<.001	0.223	> .99

**Fig. 2** Agreement probability,  $p(a)$ , for concrete and abstract concepts and for conditions  $S \rightarrow S$ ,  $B \rightarrow B$ ,  $S \rightarrow B$ ,  $B \rightarrow S$ . Bars are 95% CIs. Note that we introduced jitter so that overlapping CIs are better visualized.

and  $B \rightarrow S$  ( $t(78) = 27.403$ ,  $p < .001$ ). However, the difference in  $p(a)$  between conditions  $S \rightarrow B$  and  $B \rightarrow S$  is not statistically significant ( $t(78) = 0.321$ ,  $p = .749$ ). Similarly,  $p(a)$  for concrete concepts and for condition  $B \rightarrow B$  is higher than for conditions  $S \rightarrow B$  ( $t(78) = 31.607$ ,  $p < .001$ ) and  $B \rightarrow S$  ( $t(78) = 30.956$ ,  $p < .001$ ). These results are again consistent with our hypotheses in [Differences between concrete and abstract concepts](#) and [Differences in semantic representations between congenitally blind and sighted individuals](#). Perceptual information is probably dominant and imposes homogeneity on the sighted subjects' sample. Lacking this information in the blind subjects' sample presumably introduces differences in the lists of properties being produced, which in turn is reflected in the comparisons reported above.

An interesting result that Fig. 2 illustrates is that, for abstract concepts, the difference in  $p(a)$  between the  $S \rightarrow S$  and  $B \rightarrow B$  conditions ( $t(18) = 0.138$ ,  $p = .892$ ), as well as between  $S \rightarrow B$  and  $B \rightarrow S$  ( $t(18) = 0.261$ ,  $p = .797$ ), is not statistically significant. Though this is a null result, and should be considered with care, it is expected by our theoretical analysis. Because abstract concepts should be learned by paying attention to the same social and linguistic input in both blind and sighted populations, there is no reason to expect that the respective list of properties should differ in these comparisons.

A final noteworthy results is that, as shown in Fig. 2,  $p(a)$  for abstract concepts is higher for the  $S \rightarrow S$  than for  $S \rightarrow B$  ( $t(18) = 12.791$ ,  $p < .001$ ) and  $B \rightarrow S$  ( $t(18) = 12.408$ ,  $p < .001$ ) conditions. The same happens for  $B \rightarrow B$  with respect to conditions  $S \rightarrow B$  ( $t(18) = 14.666$ ,  $p < .001$ ) and  $B \rightarrow S$  ( $t(18) = 14.167$ ,  $p < .001$ ). Interestingly, lists within groups are more homogeneous than between groups, suggesting that there are factors that operate differently in each group to produce these results (e.g., different learning experiences). This is a surprising result, because it suggests that differences in the property lists that characterize the two groups extend beyond concrete concepts. This was expected for concrete concepts, but we currently have no explanation for why it would happen for abstract concepts. This result awaits replication for further discussion.

### Classification of concrete versus abstract concepts using several machine learning tools and inputs

Given that we found evidence that agreement probability values differ between abstract and concrete concepts, both for sighted and blind subjects, we used several machine learning (ML) techniques to assess whether agreement

probability is not only able to discriminate abstract from concrete at the aggregated average level of analysis, but also at the level of individual concepts. To foreshadow, our results show that  $p(a)$  can be used to classify abstract versus concrete with a good level of certainty. To better generalize and understand our findings, we used the ML tools k-nearest neighbors (KNN), Gaussian naïve Bayes (NB), decision trees (DT), and support vector machines (SVM). Additionally, and as a baseline, we also employed logistic regression (LR), which is a simpler regression tool. The inputs to all those tools were:  $s&k$ :  $s_j$  (mean list length) and  $k_j$  (number of unique properties listed for each concept), equiprobable  $p(a)^{eq}$  for each concept (i.e.,  $p(a)^{eq} = s_j/k_j$ : agreement probability without taking into account the property frequency distribution, see Eq. (5)), and non-equiprobable  $p(a)$  (i.e.,  $p(a)$  computed using the simulator, which takes into consideration the property frequency distribution, see Eq. (4) and description of the simulator). The idea behind using the three aforementioned variables was to assess whether more parsimonious variables achieve a better classification than more elaborated ones (i.e.,  $s&k$  is the most parsimonious variable and  $p(a)$  the least parsimonious one). The classification performance measure used was the  $F_1$  score, which is given by Eq. (7):

$$F_1 score = \frac{2 TP}{2 TP + FP + FN} \quad (7)$$

where we use the values of the confusion matrix (TP: true positives; FP: false positives; FN: false negatives, and TN: true negatives). The  $F_1$  score is the harmonic average of precision ( $TP / (TP+FP)$ ) and recall ( $TP / (TP+FN)$ ) and thus, it balances two objectives: that most of the points which belong to the positive class are correctly classified (i.e., recall), and that most of the points classified as positive class are correctly classified (i.e., precision). The  $F_1$  score varies between 0 and 1, and a high value implies that the model can appropriately classify the positive class and generates a low number of false negatives and false positives, where true positives are associated with the class with fewer labels. We must note that we could have also used accuracy, which is one of the most typical classification performance measures employed in machine learning and that indicates the percentage of correctly classified points over the total number of data points. However, this measure behaves improperly when a class is biased (i.e., when a class has substantially more data points than the rest) because high accuracy is achievable labeling all data points as members of the majority class, which is exactly the situation we face here, i.e., there are 40 concrete and 10 abstract concepts.

Even though there are several classification models in the literature, some of them need a large number of data points, e.g., neural networks, to learn the parameter's model.

For this reason, in this paper, we use the following classic models:

- k-nearest neighbors (KNN): The KNN model is one of the most simple and basic classification models, called a lazy learner (Cover & Hart, 1967). There is no training process and the classification process is based on the distance and class of the k-closest neighbors of a test point. Specifically, when a new data point is presented, the distance over all the training data points is calculated, and the closest k points, with their respective labels, are selected. Based on these k points the probability of belonging to a class is estimated as the total number of points belonging to a specific class over k. For this work, we chose  $k = 3$  to avoid overfitting (i.e., avoid memorization of the training data, obtaining a high test error).
- Gaussian naïve Bayes (NB): The NB model is based on Bayes' theorem and probability conditional independence. For a given series of known inputs or variables, the model quantifies the conditional probability that the analyzed record belongs to a specific category of the class label (Langley et al., 1992). However, given the difficulty of finding the conditional probability of the data for a specific class label, the model assumes independence between variables given the class. Once the parameters are learned and given a new data point, the model calculates the probability of belonging to each class (standardizing the proportional probabilities generates corresponding probabilities).
- Decision trees (DT): A DT is a structure composed by nodes, leaves, and branches, where each node corresponds to a decision (or a test applied on some attribute), and each branch represents a possible path of this decision or test. When a data point is inserted in the model, the tree is traveled until a leaf is reached. Each "leaf" determines the probability that the data point belongs to one of the two possible classes (Quinlan, 1986). For this work, we restricted the depth of the tree to two levels to avoid overfitting.
- Support vector machines (SVM): SVM uses a hyperplane to separate between classes. The training algorithm searches for the hyperplane with the highest "margin," i.e., the hyperplane such as the distance to the support vector points (closest points of each class to the hyperplane) is maximized (Boser et al., 1992). For complex problems, the dimension of the data points can be increased artificially by a kernel function, and the hyperplane in this new dimension can be found.
- Logistic regression (LR): LR is a nonlinear regression that allows predicting binary variables. The model calculates the probability that a data point belongs to one of the two possible classes, using a logistic function (Fang, 2013).

The  $F_1$  score results for abstract versus concrete concept classification can be seen in Table 3, where we used abstract concepts as the positive class. We evaluated three different datasets: sighted (26 participants), blind (22 participants), and both (sighted and blind combined, with a total of 48 participants). The classification results, for each dataset, correspond to the average of the test folds using a five-fold stratified cross-validation approach. This approach separates the selected dataset into five folds, using four folds for training and another fold for testing (the stratification forces each fold to have two abstract concepts). The process is repeated five times, using each fold as a test set. All models were also checked for overfitting, obtaining a test error similar to the training error.

† shows cases where the average value of  $F_1$  score achieved by using  $p(a)$  or  $p(a)^{eq}$  compared to  $s\&k$  are higher and statistically significant at least at the 0.05 level. An \* indicates that the  $F_1$  score achieved by using  $p(a)$  is higher than the one of  $p(a)^{eq}$  and of  $s\&k$ , and is statistically significant at least at the 0.05 level.

As can be seen from Table 3, most of the models using  $s\&k$  are unable to obtain good performance, the lowest score being 0.07. In contrast, by comparing  $s\&k$  with  $p(a)^{eq}$  and  $p(a)$ , the  $F_1$  scores for the agreement probability values are higher on 12 and 13 occasions, respectively. From those comparisons, the  $F_1$  scores achieved by  $p(a)^{eq}$  and  $p(a)$  are higher and statistically significant at least at the 0.05 level in 7 and 12 cases, respectively. This shows that  $p(a)^{eq}$  and  $p(a)$  better differentiate concrete from abstract concepts in contrast to using the more parsimonious  $s\&k$  combination. Additionally, comparing the  $F_1$  scores achieved with  $p(a)$  and  $p(a)^{eq}$ , we can see that they are equal to or higher for  $p(a)$  in all cases, differences that reach statistical significance at the 0.05 level in three comparisons. Thus, all in all, we may say that the classification performance attained by  $p(a)$  is the best, followed by  $p(a)^{eq}$ , and those two, trailed by  $s\&k$ .

Finally, note from Table 3 that the  $F_1$  scores for  $p(a)$  suggest that the discrimination between concrete and abstract concepts in the sighted population is better than in the blind population. All  $F_1$  scores for the five classification tools

used are statistically significantly higher for sighted than for blind, except for KNN ( $t(8)$  and  $p$  value in parenthesis; respectively, 0.746 (0.477), 3.751 (.006), 4.399 (0.002), 3.651 (0.006), and 2.800 (0.023), and note that the values for making these comparisons come from executing the test fold for each tool five times). This is consistent with our hypothesis that the difference between concrete and abstract concepts would be more conspicuous in sighted than in blind, because the blind population tends to learn abstract concepts in much the same way it learns concrete concepts due to a lack of visual perceptual properties. Hence, this similarity in learning concrete and abstract concepts blurs their distinction in the blind population.

### Discussion and conclusions

In the current work, we have discussed agreement probability, a measure of homogeneity of concept instantiations in the Property Listing Task. Being a probability, the measure has the positive characteristic of being naturally bounded in the 0 to 1 range. Relatedly, the 0 and 1 values are interpreted in a clear and straightforward fashion (i.e., respectively, total heterogeneity and total homogeneity). Additionally, agreement probability naturally integrates information produced when property listing data is collected into a single value that depends on the average list length produced by subjects ( $s$ ), the total number of unique properties produced by the subject sample ( $k$ ), and the frequency distribution of those properties. Finally, as shown in Appendix A, agreement probability also has the nice feature of directly implying that nonuniform property probability distributions reflect greater homogeneity in property lists (see the lower bound demonstration in Appendix A) and so they should be considered in a homogeneity index.

We assume that heterogeneity is an inherent characteristic of naturally occurring concepts coded in language. Many factors could influence this heterogeneity in the real world. Consequently,  $p(a)$  could be used to gauge these factors' relative influence when comparing types of concepts or

**Table 3**  $F_1$  score results for abstract versus concrete concept classification using fivefold stratified cross-validation (mean, std. dev. in parentheses)

F1	Sighted			Blind			Sighted and blind		
	s&k	$p(a)^{eq}$	$p(a)$	s&k	$p(a)^{eq}$	$p(a)$	s&k	$p(a)^{eq}$	$p(a)$
<b>KNN</b>	0.89(0.07)	0.69(0.23)	0.84(0.06)	0.39(0.22)	0.63(0.06)	0.82(0.05) †*	0.70(0.12)	0.68(0.04)	0.87(0.07) †*
<b>NB</b>	0.34(0.18)	0.69(0.23) †	0.94(0.08) †	0.07(0.09)	0.59(0.08) †	0.74(0.09) †*	0.44(0.15)	0.67(0.06) †	0.67(0.05) †
<b>DT</b>	0.42(0.13)	0.74(0.25) †	0.84(0.06) †	0.42(0.21)	0.61(0.06)	0.65(0.08) †	0.54(0.04)	0.70(0.08) †	0.77(0.06) †
<b>SVM</b>	0.63(0.06)	0.69(0.23)	0.83(0.10) †	0.25(0.13)	0.61(0.06) †	0.64(0.06) †	0.63(0.06)	0.67(0.06)	0.67(0.05)
<b>LR</b>	0.43(0.25)	0.69(0.23)	0.81(0.11) †	0.22(0.13)	0.61(0.06) †	0.67(0.02) †	0.69(0.07)	0.65(0.02)	0.67(0.05)

types of conceptualizers. To show that this is the case, we compared conceptual agreement values between two types of concepts and two types of conceptualizers.

### What have we learned from the concrete/abstract and blind/sighted comparisons

A large literature strongly suggests that concrete concepts are different from abstract concepts. In that literature, evidence is discussed that when conceptualizing, people routinely reenact perceptual content associated with the corresponding concepts (Kan et al., 2003; Lupyan & Ward, 2013; Ostarek & Huettig, 2017; Santos et al., 2011; Solomon & Barsalou, 2004), which is characteristic of concrete concepts. In contrast, abstract concepts appear not to be characterized as much by perceptual content, but rather by social and linguistic associations (Barsalou et al., 2008; Borghi et al., 2017; Borghi & Cimatti, 2009; Breedin et al., 1994; Paivio, 1986; Wiemer-Hastings & Xu, 2005).

From this literature, we posited our *characterizing concreteness hypotheses*, which holds that concrete concepts are characterized more by perceptual information than abstract concepts. Importantly, we also hypothesized that this perceptual information introduces a greater homogeneity in conceptualization for concrete than for abstract concepts. Consistently with this hypothesis, we found visual strength subjective ratings obtained from Vergallito et al. (2020) to be higher for concrete than for abstract concepts, and we also found that visual strength ratings positively correlated with our  $p(a)$  measure, confirming that visual information is associated with increasing homogeneity across participants.

A somewhat surprising result is that the positive correlation between visual strength ratings and  $p(a)$  also occurs when blind subjects' data are analyzed. This suggests that blind participants not only have information about visual perceptual information (e.g., that “black” and “white” can be used to describe *zebras*), which is likely to be obtained from blind subjects' interactions with the sighted community (cf., Louwerse, 2018), but also that this linguistic source introduces homogeneity in their lists, similarly to what occurs with sighted subjects. In fact, there is recent evidence which is consistent with this. The concreteness advantage effect consists in people showing faster processing for concrete than for abstract words, presumably due to the effect of perceptual information. However, Bottini et al. (2022) report that early blind subjects show this effect even when a word's concreteness depends mostly on its reliance on visual information (e.g., “blue”).

Because discrimination tasks that rely on visual information can detect differences between sighted and blind subjects (Connolly et al., 2007; Kim et al., 2019), in our *role of vision hypothesis*, we posited that visual reenactments should introduce greater homogeneity for concrete concepts

in sighted compared to blind participants. This seems at odds with our finding discussed in the immediately preceding paragraph, which suggests that blind participants do have information about visual perceptual information, presumably acquired through regularities experienced in language, and that this information does introduce relative homogeneity in the lists they produce. However, as discussed next, we did find evidence consistent with our *role of vision hypothesis*.

As shown in Fig. 2, concrete concepts are less homogeneous for blind than for sighted participants. To explain this apparent contradiction, here we further hypothesize that when visual reenactments occur, they capture attention and guide property listing. Thus, even if blind subjects have the linguistically represented perceptual information, their lists rely on linguistic associations and not on the highly salient visual reenactments. In contrast, for sighted subjects, visual reenactments capture attention and guide listing, thus introducing homogeneity to a larger extent than would be expected only from linguistic regularities.

An additional and interesting finding is that, as shown in Fig. 2,  $p(a)$  computations indicate that property lists differ across groups of conceptualizers, suggesting that perhaps different learning experiences lead to different category memory representations. When comparisons were made across groups (i.e., between blind and sighted participants),  $p(a)$  values were consistently lower than when those comparisons were made within the same groups. Evidently, property frequency distributions were not the same across our groups.

Though being able to use  $p(a)$  to make group level comparisons (i.e., groups of concepts and groups of conceptualizers) is already interesting, we also showed that  $p(a)$  can be used to discriminate between individual concepts. If abstract concepts are characterized by producing more variable instantiations in the PLT than concrete concepts, then,  $p(a)$  might also allow discriminating between concepts at the individual concept level (i.e., showing that a particular concept can be classified as concrete or abstract based on its agreement probability value). To this effect, we introduced a simple measure consisting of  $s$  (the average number of properties produced by subjects) and  $k$  (the total number of unique properties produced by the whole subject sample) and contrasted it with  $p(a)^{eq}$  and  $p(a)$  in their capacity to discriminate concrete and abstract concepts. These three variables were submitted to machine learning algorithms and their classification performances contrasted. Overall, our data showed that the best classification performance was achieved by  $p(a)$ . Three consequences ensue: agreement probability carries more useful information about concepts than its  $s$  and  $k$  constituents considered in isolation; information about property frequency distribution needs to be considered in the computation of agreement probability; abstract concepts effectively are more heterogeneous than concrete



concepts, not only as a group, but also at the level of individual concepts. Additionally, classification results for  $p(a)$  between sighted and blind indicated that a better discrimination between concrete and abstract concepts is achieved among sighted individuals. This lends further support to the theory that blind people learn abstract concepts much in the same way as concrete concepts due to lack of visual perceptual information. Hence, this similarity in learning concrete and abstract concepts blurs their distinction in the blind population, which makes distinguishing between concrete and abstract concepts more difficult for blind people.

### What more might $p(a)$ enable

As discussed above, the work we report here shows that variability in the PLT is not necessarily noise. Rather, variability in the PLT contains information that can be meaningfully related to the literature on the abstract concept versus concrete concept distinction, and to the literature on the effect that lack of sight has on conceptual representations. That a simple task like the PLT contains such a wealth of information is surprising. In what follows, we want to suggest other issues to which  $p(a)$  could be applied to gain theoretical insights.

**Studying conceptualizations in social groups** The PLT and ensuing CPNs have been used to characterize shared semantic memory in social groups, either using them in isolation or in combination with other techniques (e.g., Hood, 2020; Mazzuca et al., 2020; Sunohara et al., 2022; Weiler & Jacobsen, 2021). The aim of these researchers has been to characterize shared semantic concepts in a particular social group (e.g., to characterize knowledge of foods in children; to characterize the meaning of tattoos in older adults). However, it is not trivial to claim that a certain semantic structure (1) is shared across members of a social group, and (2) is also specific to that social group, in contrast to being relatively invariant across different social groups. Following our comparison between blind and sighted subjects, we envision that, by using  $p(a)$ , it should be possible to compare linguistically coded concepts in different social groups. A group would have a shared and distinctive conceptualization if the within-group  $p(a)$  is greater than the between-group  $p(a)$ , just as our analyses illustrate.

**Analyzing the effect of coding on CPN results** Because the PLT task is highly productive and properties can be expressed in numerous ways (e.g., people who are cued with the concept *democracy* may use “a president is elected” and “there are presidential elections” to essentially refer to the same property), PLT data needs to be coded. This coding process typically involves several coders, and inter-coder reliability is always a concern. Only recently have there been

attempts to develop methods oriented to promoting highly reliable codings (Buchanan et al., 2020; Reid & Katz, 2022). Note that low or even moderate reliabilities make it difficult to aim for replicable studies.

The problems introduced by coding are partly responsible for why one can hardly find in the literature studies that make use of coding procedures that were developed by other researchers and why CPN studies are seldom replicated. A closely related problem is the following. Coders in different CPN studies could code highly related sets of raw properties with slightly different labels, and their codes could produce somewhat different partitions of the raw properties, such that inter-studies comparisons are made difficult to carry out (i.e., How do we know if the coded properties yield similar data structures to the extent that both studies should be considered replications?). Note that these problems only increase when concepts of interest are abstract, because people tend to produce more unique properties. We hope that computing  $p(a)$  could help to solve these issues, given that different coding systems over the essentially same raw property data should produce comparable agreement values.

**Testing the effect of context on the instantiation of a concept** It has been long argued that contextual knowledge plays a central role in categorization and cognition (Chaigneau et al., 2009; Kiefer & Pulvermüller, 2012; Lin & Murphy, 2001; Roth & Shoben, 1983; Wenchi & Barsalou, 2006). Furthermore, evidence supports the idea that conceptual properties can be meaningfully divided into those that are context dependent (i.e., those that become active depending on specific contexts, e.g., that a *basketball* “can float”) and those that are relatively independent from context (i.e., those that become active across different contexts, e.g., that *dogs* “bark”) (Barsalou, 1982). If, as we hypothesize (see our [Agreement probability as a measure of homogeneity](#)), concepts with lower  $p(a)$  are those for which people may adopt different points of view when conceptualizing them, then, manipulating contexts should change a concept’s  $p(a)$ . To test this hypothesis, we envision experiments where property lists are obtained after subjects have been primed with specific contexts, and we would predict that  $p(a)$  should increase when specific relevant contexts are introduced, and that perhaps abstract concepts should be relatively more influenced by this manipulation. However, these experiments are beyond the scope of the current work, and we defer them for future work.

On closing, we want to highlight that our  $p(a)$  measure is consistent with views that see an intimate link between cognition and culture (Atran, 2003; Berntsen & Rubin, 2004; DiMaggio, 1997; Lehman et al., 2004; McCauley et al., 2022; ojaletto & Medin, 2015; Patterson, 2014; Roberson et al., 2000; Talmy, 2000; Waxman et al., 2007), where cognition is thought to reflect objective cultural practices in the subjective domain

(Kashima, 2016; Nisbett et al., 2001; Nisbett & Masuda, 2003; Nisbett & Miyamoto, 2005; Romney & Moore, 1998). Thus, we believe that  $p(a)$  has a wide range of application and will be pleased if it does indeed live up to this standard.

## Appendix A

### Mathematical demonstrations

#### 1. Initial definitions

Let's consider two sets of coded properties listed in a PLT for two concepts  $C1$  and  $C2$ :

$$C1 = \{a_1, a_2, \dots, a_{k_1}\} \quad \text{and} \quad C2 = \{c_1, c_2, \dots, c_{k_2}\} \quad (A1)$$

In this case, the cardinality of  $C1$  is  $k_1$  and the one of  $C2$  is  $k_2$ . Let's also denote the size of a sample extracted from  $C1$  as  $s_1$  and the size of a sample taken from  $C2$  as  $s_2$ . Then, the number of such possible independent samples will be:

$n_1$ : total number of possible samples of size  $s_1$  obtained from  $C1$

$n_2$ : total number of possible samples of size  $s_2$  obtained from  $C2$

Thus,

$$n_1 = \binom{k_1}{s_1} \quad \text{and} \quad n_2 = \binom{k_2}{s_2}. \quad (A2)$$

If each independent sample is an event of a random variable, we can define the set of all possible events as:

$$M_1 = \{S \subseteq C_1 : \#S = s_1, S_i \neq S_j \quad \forall i, j\} = \{S_1^1, S_2^1, \dots, S_{n_1}^1\} \quad (A3)$$

$$M_2 = \{S \subseteq C_2 : \#S = s_2, S_i \neq S_j \quad \forall i, j\} = \{S_1^2, S_2^2, \dots, S_{n_2}^2\} \quad (A4)$$

So, we have the following probability:

$$p(a) = \frac{1}{s_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \#(S_i^1 \cap S_j^2) p_i q_j, \quad \sum_{i=1}^{n_1} p_i = 1, \quad p_i \geq 0 \quad \sum_{j=1}^{n_2} q_j = 1, \quad q_j \geq 0. \quad (A5)$$

where  $p_i$  is the probability of obtaining sample  $S_i^1$  and  $q_j$  is the probability of obtaining sample  $S_j^2$ .

Note that throughout this appendix we will use  $C1$  and  $C2$  to designate two concepts, as well as the set of unique coded properties listed for such concepts, where the corresponding meaning is given by the context in which those labels are used.

The probability in Eq. (A5) may be seen as a quadratic form by defining the following matrices:

$$B = (b_{ij}) \quad \text{where} \quad b_{ij} = \#(S_i^1 \cap S_j^2). \quad (A6)$$

$$p = (p_1, p_2, \dots, p_{n_1}) \quad (A7)$$

$$q = (q_1, q_2, \dots, q_{n_2}) \quad (A8)$$

$$p(a) = \frac{1}{s_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} p_i b_{ij} q_j = \frac{1}{s_2} p' B q \quad (A9)$$

#### 2. Demonstration $p(a)^{eq}$ for two different concepts $C1$ and $C2$ is equal to $\frac{s_1}{k_1} \frac{u}{k_2}$

Beginning with Eq. (A9), for equiprobable elements of  $C1$  and  $C2$ , we can see that the samples obtained from  $C1$  or  $C2$  are also equiprobable (they have the same probability of being obtained), and thus  $p_j = \frac{1}{n_1}$  and  $q_j = \frac{1}{n_2}$ . In that case,

$$p(a)^{eq} = \frac{1}{s_2 n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} b_{ij} = \frac{1}{s_2 n_1 n_2} \mathbf{1}' B \mathbf{1} \quad (A10)$$

where  $p(a)^{eq}$  denotes that we are calculating such probability for equiprobable elements of  $C1$  and  $C2$ .

Using the notation  $B = (Z_1, Z_2, \dots, Z_{n_1})$ , where  $Z_i$  is the  $i$ -th column of matrix  $B$ , we have that  $\mathbf{1}' Z_i$  is the sum of each column. Also, if we denote  $G(Z_i)$  as the sum of the  $i$ -th column, we can rewrite Eq. (A10) as:

$$p(a)^{eq} = \frac{1}{s_2 n_1 n_2} \mathbf{1}' B \mathbf{1} = \frac{1}{s_2 n_1^2} (G(Z_1), \dots, G(Z_{n_1}))' \mathbf{1} = \frac{1}{s_2 n_1 n_2} \sum_{i=1}^{n_1} G(Z_i). \quad (A11)$$

Let  $l_i$  be the maximum number of common elements between the  $i$ -th element of  $M_1$  (for example  $S_i^1$ ) and the set  $M_2$ . Then, the sum of each column  $G(Z_i)$  in matrix  $B$  depends on  $l_i$ .

Considering that  $r$  is the number of common elements between two independent samples of size  $s_1$  and  $s_2$  obtained from the  $k_1$  and  $k_2$  elements of  $C1$  and  $C2$ , respectively, then  $G(Z_i)$  will be  $r$  multiplied by the number of samples that contain those  $r$  common elements, and summed over all possible  $r$  values. The number  $r$  can be modelled as a hypergeometric random variable as follows:

$R \sim \text{HyperGeo}(s_2, k_2, l_i)$ , whose probability mass function is given by:

$$f(r) = \frac{\binom{l_i}{r} \binom{k_2 - l_i}{s_2 - r}}{\binom{k_2}{s_2}}, \quad r \in \text{Range}(R) = \{\max\{0, s_2 + l_i - k_2\}, \dots, \min\{s_2, l_i\}\}. \quad (A12)$$

We know that the expected value of a hypergeometric random variable as already defined is:  $E[R] = \frac{s_2 l_i}{k_2}$  and by definition of  $E[R]$ :

$$\begin{aligned}
 E[R] &= \sum_{r \in \text{Rge}(R)} r f(r) = \frac{s_2 l_i}{k_2} \Rightarrow G(Z_i) \\
 &= \sum_{v \in \text{Rge}(R)} r \binom{l_i}{r} \binom{k_2 - l_i}{s_2 - r} = \frac{s_2 l_i}{k_2} \binom{k_2}{s_2} = \frac{n_2 s_2 l_i}{k_2}. \tag{A13}
 \end{aligned}$$

Then,

$$p(a)^{eq} = \frac{1}{s_2 n_1 n_2} \sum_{i=1}^{n_1} G(Z_i) = \frac{1}{s_2 n_1 n_2} \sum_{i=1}^{n_1} \frac{n_2 s_2 l_i}{k_2} = \frac{1}{n_1 k_2} \sum_{i=1}^{n_1} l_i. \tag{A14}$$

Recall that  $l_i$  is the maximum number of common elements between the  $i$ -th element of  $M_1$  and the set  $M_2$ . Let  $u$  be the maximum number of common elements among all possible combinations of the sets  $M_1$  and  $M_2$  (note that in Table 1  $u$  is differently defined as “number of common properties between the set of unique properties listed for concept  $C1$  and  $C2$ ,” but that definition is totally consistent with the one we use here), then  $l_i$  varies between  $\max\{0, s_1 + u - k_1\}$  and  $\min\{s_1, u\}$ .

Now, we can rewrite  $\sum_{i=1}^{n_1} l_i$  as the possible values of  $l_i$  multiplied by the number of times that each  $l_i$  value appears in its summation, i.e.,

$$\sum_{i=1}^{n_1} l_i = \sum_{v=\max\{0, s_1+u-k_1\}}^{\min\{s_1, u\}} (v \cdot (\text{nr. of times } v \text{ appears})).$$

As can be observed,  $v$  is the number of common elements between two independent samples of size  $s_1$  obtained from the  $k_1$  elements of  $C1$ . Thus, the number  $v$  can be modelled as a hypergeometric random variable as follows:

$V \sim \text{HyperGeo}(s_1, k_1, u)$ , whose probability mass function is given by:

$$\begin{aligned}
 f(v) &= \frac{\binom{u}{v} \binom{k_1 - u}{s_1 - v}}{\binom{k_1}{s_1}}, \quad v \in \text{Range}(V) \\
 &= \{\max\{0, s_1 + u - k_1\}, \dots, \min\{s_1, u\}\}. \tag{A15}
 \end{aligned}$$

We know that the expected value of a hypergeometric random variable as already defined is:  $E[V] = \frac{s_1 u}{k_1}$  and by definition of  $E[V]$ :

$$\begin{aligned}
 E[V] &= \sum_{v \in \text{Rge}(V)} v f(v) = \frac{s_1 u}{k_1} \Rightarrow \sum_{i=1}^{n_1} l_i \\
 &= \sum_{v \in \text{Rge}(V)} v \binom{u}{v} \binom{k_1 - u}{s_1 - v} \\
 &= \frac{s_1 u}{k_1} \underbrace{\binom{k_1}{s_1}}_{n_1} = \frac{n_1 s_1 u}{k_1}. \tag{A16}
 \end{aligned}$$

Replacing Eq. (A16) in Eq. (A14), we obtain:

$$p(a)^{eq} \text{ for two concepts } C1 \text{ and } C2 = \frac{1}{n_1 k_2} \sum_{i=1}^{n_1} l_i = \frac{1}{n_1 k_2} \frac{n_1 s_1 u}{k_1} = \frac{s_1 u}{k_1 k_2}. \tag{A17}$$

**3. Corollary  $p(a)^{eq}$  for the same concept  $C1$  is equal to  $\frac{s_1}{k_1}$**

Beginning with Eq. (A17), we can see that if  $C1 = C2$ , then  $u = k_2$ , i.e., for the same concept  $C1$ , the number of common properties between the same two concepts must be equal to the number of properties for  $C1$ . Thus, in Eq. (A17)  $u/k_2 = 1$ , and we can write:

$$p(a)^{eq} \text{ for the same concept } C1 = \frac{s_1}{k_1} \tag{A18}$$

Additionally, replacing  $s_1/k_1$  in Eq. (A17) by Eq. (A18), we can also write:

$$p(a)^{eq} \text{ for two concepts } C1 \text{ and } C2 = (p(a)^{eq} \text{ for concept } C1) \left(\frac{u}{k_2}\right) \tag{A19}$$

**4. Corollary  $p(a)^{eq}$  for two different concepts  $C1$  and  $C2 < p(a)^{eq}$  for the same concept  $C1$**

Note that for equiprobable elements of  $C1$  and  $C2$  Eq. (A19) implies that  $p(a)^{eq}$  for the same concept  $C1$  will condition the value of the corresponding  $p(a)^{eq}$  for two different concepts  $C1$  and  $C2$ , depending on the amount of overlap between the  $C1$  and  $C2$  concepts. The only way in which  $p(a)^{eq}$  for two different concepts  $C1$  and  $C2$  could be equal to  $p(a)^{eq}$  for the same concept  $C1$  is if  $u = k_2$ , i.e., if the cardinality of the intersection between  $C1$  and  $C2$  (i.e., the number of properties that belong to both  $C1$  and  $C2$ ) is equal to the cardinality of  $C2$ , which means that  $C2 \subseteq C1$ . Thus, for two different concepts (i.e.,  $C2 \not\subseteq C1$ ), it must hold that  $p(a)^{eq}$  for two different concepts  $C1$  and  $C2 < p(a)^{eq}$  for the same concept  $C1$ .

**5. Lower bound of  $p(a)$  for the same concept  $C1$  is equal to  $p(a)^{eq}$  for the same concept  $C1$**

If we calculate  $p(a)$  for the same concept  $C1$ , then we can rewrite Eq. (A5) as:

$$\begin{aligned}
 p(a) &= \frac{1}{s_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \#(S_i^1 \cap S_j^1) p_i p_j, \\
 \sum_{i=1}^{n_1} p_i &= 1, \quad p_i \geq 0 \quad \sum_{j=1}^{n_1} p_j = 1, \quad p_j \geq 0. \tag{A20}
 \end{aligned}$$

Where  $S_i^1$  and  $S_j^1$  are samples drawn from the same concept  $C1$  and because of that we use the same superscript 1. Note also that  $p_i$  and  $p_j$  are the corresponding probabilities of obtaining such samples.

The probability in Eq. (A20) may be seen as a quadratic form by defining the following matrices:

$$A = (a_{ij}) \text{ where } a_{ij} = \#(S_i^1 \cap S_j^1) \tag{A21}$$

$$p = (p_1, p_2, \dots, p_{n_1}) \tag{A22}$$

$$p(a) = \frac{1}{s_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} p_i a_{ij} p_j = \frac{1}{s_1} p^t A p \tag{A23}$$

To calculate a lower bound on  $p(a)$  for the same concept  $CI$ , we must minimize expression (A23). This minimization problem expressed in matrix notation is:

$$\begin{aligned} & \underset{s_1}{\min} \frac{1}{2} p^t A p, \\ & \text{s. t. } p^t \mathbf{1} = 1 \\ & p \geq 0. \end{aligned} \tag{A24}$$

The properties of  $A$  are:

1. Symmetric, i.e.,  $A^t = A$ .
2.  $a_{ii} = s_1 \forall i$ .
3.  $s_j = a_{ii} > a_{ij} \geq 0 \forall i, j$
4.  $A\mathbf{1} = \alpha\mathbf{1}$  or  $\mathbf{1}^t A = \alpha\mathbf{1}^t$ , where  $\alpha$  is the sum of a row  $G(\mathbf{Z}_i)$  calculated in (A13), but now  $s_2 = s_1$  and  $k_2 = k_1$ . Thus:

$$\begin{aligned} \alpha = G(\mathbf{Z}_i) &= \sum_{v \in \text{Rge}(R)} r \binom{l_i}{r} \binom{k_1 - l_i}{s_1 - r} \\ &= \frac{s_1 l_i}{k_1} \underbrace{\binom{k_1}{s_1}}_{n_1} = \frac{n_1 s_1 l_i}{k_1} = \frac{n_1 s_1^2}{k_1}. \end{aligned} \tag{A25}$$

where you should also remember that  $l_i$  is the maximum number of common elements but now for the same concept  $CI$ , and hence  $l_i = s_j$ .

The first property indicates that all the eigenvalues of  $A$  are real numbers and the last property shows that all the rows, and by symmetry, all the columns of  $A$  sum the same value. This same property indicates that  $\alpha$  is an eigenvalue of  $A$  and the vector  $\mathbf{1}$  is its respective eigenvector.

Before solving (A24), we must note that  $\frac{2}{s_1}$  is a constant, and thus it will not be considered in solving the problem. Additionally, considering an objective function  $f(p) = \frac{1}{2} p^t A p$  and that the solution set where we are working is  $D = \{p \in \mathfrak{R}^{n_1} : p^t \mathbf{1} = 1, p \geq 0\}$ , it is easy to see that the problem has a solution, given that the objective function is continuous and  $D$  is a closed and bounded set, reaching the minimum value either in the interior of  $D$ , when  $p > 0$ , or at the frontier of  $D$ , when a component of  $p = 0$ . The Weierstrass theorem guaranties the existence of the minimizer in  $D$ .

To solve problem (A24), let's consider the Lagrangian:

$$L(p, \lambda, \mu) = \frac{1}{2} p^t A p + \lambda(p^t \mathbf{1} - 1) - \mu^t p. \tag{A26}$$

Note that the restrictions of the type  $p_i \geq 0$  or  $p_i \leq 0$  are not included in the Lagrangian, but the so-called complementary conditions, which are  $\mu_i p_i = 0$ . Thus, the problem to be solved, KKT (Karush-Kuhn-Tucker) conditions, is:

$$\begin{aligned} \nabla_p L(p^*, \lambda^*, \mu^*) &= 0, \\ (p^t \mathbf{1} - 1) &= 0, \\ -p &\leq 0, \\ \mu &\geq 0, \\ \mu_i p_i &= 0, \quad i = 1, \dots, n_1 \end{aligned} \tag{A27}$$

Remembering matrix derivatives, we have that:

$$\nabla_p L(p, \lambda, \mu) = \begin{pmatrix} \frac{\partial L}{\partial p_1} \\ \vdots \\ \frac{\partial L}{\partial p_{n_1}} \end{pmatrix} = A p + \lambda \mathbf{1} - \mu \tag{A28}$$

To solve the Lagrangian, let's first consider that  $\mu_i = 0$ , that is, we will for now suppose that the restrictions that  $p$  must be positive are inactive. Then,

$$\begin{aligned} A p + \lambda \mathbf{1} &= 0, \\ \mathbf{1} A p + \lambda \mathbf{1}^t \mathbf{1} &= 0, \\ \alpha \mathbf{1}^t p + \lambda n_1 &= 0, \\ \alpha + \lambda n_1 = 0 &\Rightarrow \lambda^* = -\frac{\alpha}{n_1} \end{aligned}$$

Replacing the value of  $\lambda^*$  in  $A p + \lambda \mathbf{1} = 0$  and using the fact that  $A \mathbf{1} = \alpha \mathbf{1}$ , we have that:

$$\begin{aligned} A p - \frac{\alpha}{n_1} \mathbf{1} &= 0, \\ A p - \frac{1}{n_1} A \mathbf{1} &= 0, \\ A \left( p - \frac{1}{n_1} \mathbf{1} \right) &= 0. \end{aligned}$$

If  $\det(A) \neq 0$ , then  $p^*$  is unique and equal to:

$$p^* = \frac{1}{n_1} \mathbf{1} = \begin{pmatrix} 1/n_1 \\ \vdots \\ 1/n_1 \end{pmatrix} \tag{A29}$$

If  $\det(A) = 0$ , there exist infinite solutions and one of them is the one already found in Eq. (A29). The rest of the solutions are of the form:

$$q^* = p^* + b v, \quad \text{where } v \in \text{Ker}(A) = \{v \in \mathfrak{R}^{n_1} : A v = 0\} \tag{A30}$$

The constant  $b \in \mathfrak{R}$ , is any that gives  $q^* \geq 0$ . The vector  $v$ , is a linear combination of the eigenvectors of matrix  $A$  associated with the zero eigenvalues. Given that  $p^*$  and  $q^*$  will allways be positive, we can state that the restriction that  $p$  must be positive is met.

Notably, see that the solution in Eq. (A29) means that a lower bound to  $p(a)$  for the same concept  $CI$  occurs when all samples drawn from  $CI$  are equiprobable, and thus, all



elements in  $CI$  must also be equiprobable, i.e., a lower bound to  $p(a)$  for the same concept  $CI$  is equal to  $p(a)^{eq}$  for the same concept  $CI$ .

## Appendix B

### List of concrete and abstract concepts used in this study

The following concrete (C) ( $N_C = 40$ ) and abstract (A) ( $N_A = 10$ ) concepts from the Lenci et al. (2013) norms were used in this study, where Lenci et al. classified each noun as concrete or non-concrete (i.e., abstract). Note that concepts marked with an asterisk (\*) correspond to those that are also contained in the Vergallito et al. (2020) study (15 concrete and 5 abstract concepts).

Concept (alphabetical order)	Type (C, A)
apartment	C
apple	C
banana	C
bar	C *
beach	C *
canary	C
car	C *
carrot	C
cat	C *
cathedral	C
cheerfulness	A
cherry	C
comb	C
crow	C
democracy	A
dog	C *
eggplant	C
freedom	A *
friendship	A
giraffe	C
hammer	C *
horse	C *
jealousy	A *
justice	A *
kiwi	C
knife	C *
lawn	C *
lettuce	C
motorcycle	C *
mountain	C *
pain	A *
passion	A *
pencil	C *
penguin	C
pineapple	C

potato	C
religion	A
restaurant	C *
screwdriver	C
sea	C
seagull	C
ship	C *
shop	C
swan	C
tomato	C
train	C *
woods	C
worry	A
zebra	C

**Acknowledgements** This research was carried out with funds provided by ANID, Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) of the Chilean government grant 1200139. We gratefully acknowledge Carlos Barra, dear friend and collaborator in various research projects, for his valuable insights and viewpoints regarding agreement probability, which greatly helped us refine the ideas presented in this paper. Carlos recently passed away, and thus we dedicate this paper to his memory.

**Funding** This research was carried out with funds provided by ANID, Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) of the Chilean government grant 1200139.

**Data availability** Data and materials used for the analyses reported here are available at [https://osf.io/xhfmz/?view\\_only=31c08caa642f42c694425a4f2b46a8b4](https://osf.io/xhfmz/?view_only=31c08caa642f42c694425a4f2b46a8b4).

**Code availability** Software and code are available at [https://osf.io/xhfmz/?view\\_only=31c08caa642f42c694425a4f2b46a8b4](https://osf.io/xhfmz/?view_only=31c08caa642f42c694425a4f2b46a8b4).

## Declarations

**Conflicts of interest/competing interests** The authors declare not having any known conflict of interests.

**Ethics approval** Data used in this study was obtained in a previously reported study for which the original authors obtained Ethics approval.

**Consent to participate** Informed consent was obtained from all individual participants included in the original study from which we obtained data.

**Consent for publication** Not applicable.

## References

- Atran, S. (2003). Théorie cognitive de la culture. *L'Homme*, 166, 107–144.
- Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, 10, 82–93.
- Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge University Press.

- Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 29–101). Lawrence Erlbaum Associates.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–660. <https://doi.org/10.1017/S0140525X99002149>
- Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 245–284). Oxford University Press: Oxford. <https://doi.org/10.1093/acprof:oso/9780199217274.003.0013>
- Berntsen, D., & Rubin, D. C. (2004). Cultural life scripts structure recall from autobiographical memory. *Memory and Cognition*, 32(3), 427–442. <https://doi.org/10.3758/BF03195836>
- Borghii, A. M., & Cimatti, F. (2009). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia*, 48(3), 763–773. <https://doi.org/10.1016/j.neuropsychologia.2009.10.029>
- Borghii, 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://doi.org/10.1037/bul0000089>
- Borghii, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2019). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*, 29, 120–153. <https://doi.org/10.1016/j.plrev.2018.12.001>
- Boser, B., Guyon, I., & Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In: *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144–152.
- Bottini, R., Morucci, P., D'Urso, A., Collignon, O., & Crepaldi, D. (2022). The concreteness advantage in lexical decision does not depend on perceptual simulations. *Journal of Experimental Psychology: General*, 151(3), 731–738. <https://doi.org/10.1037/xge0001090>
- Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology*, 11(6), 617–660. <https://doi.org/10.1080/02643299408251987>
- Buchanan, E. M., De Deyne, S., & Montefinese, M. (2020). A practical primer on processing semantic property norm data. *Cognitive Processing*, 21, 587–599. <https://doi.org/10.1007/s10339-019-00939-6>
- Canessa, E. C., & Chaigneau, S. E. (2016). When are concepts comparable across minds? *Quality and Quantity*, 50(3), 1367–1384. <https://doi.org/10.1007/s11135-015-0210-4>
- Chaigneau, S. E., Barsalou, L. W., & Zamani, M. (2009). Situational information contributes to object categorization and inference. *Acta Psychologica*, 130(1), 81–94. <https://doi.org/10.1016/j.actpsy.2008.10.004>
- Chaigneau, S. E., Canessa, E., & Gaete, J. (2012). Conceptual agreement theory. *New Ideas in Psychology*, 30(2), 179–189.
- Connolly, A. C., Gleitman, L. R., & Thompson-Schill, S. L. (2007). Effect of congenital blindness on the semantic representation of some everyday concepts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20), 8241–8246. <https://doi.org/10.1073/pnas.0702812104>
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27.
- Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132(2), 163–201. <https://doi.org/10.1037/0096-3445.132.2.163>
- Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). The Centre for speech, language and the brain (CSLB) concept property norms. *Behavior Research Methods*, 46(4), 1119–1127. <https://doi.org/10.3758/s13428-013-0420-4>
- DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology*, 23, 263–287.
- Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, 37, 66–108. <https://doi.org/10.1016/j.dr.2015.05.002>
- Fang, J. (2013). Why logistic regression analyses are more reliable than multiple regression analyses. *Journal of Business and Economics*, 4, 620–633.
- Fay, N., Walker, B., Swoboda, N., & Garrod, S. (2018). How to create shared symbols. *Cognitive Science*, 42, 241–269. <https://doi.org/10.1111/cogs.12600>
- Feldman, J. (2010). Embodied language, best-fit analysis, and formal compositionality. *Physics of Life Reviews*, 7(4), 385–410. <https://doi.org/10.1016/j.plrev.2010.06.006>
- Frege, G. (1893/1952). On sense and reference. In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). : Blackwell.
- Gallese, V., & Lakoff, G. (2005). The Brain's concepts. *Cognitive Neuropsychology*, 22, 455–479. <https://doi.org/10.1080/02643290442000310>
- Glock, H. J. (2009). Concepts: Where subjectivism goes wrong. *Philosophy*, 84(1), 5–29.
- Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18(4), 441–461. [https://doi.org/10.1016/S0022-5371\(79\)90246-9](https://doi.org/10.1016/S0022-5371(79)90246-9)
- Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45, 718–730. <https://doi.org/10.3758/s13428-012-0278-x>
- Hood, J. H. (2020). Cultural models of democracy among Burmese residents in the United States. *Journal of Cultural Cognitive Science*, 4(1), 107–122. <https://doi.org/10.1007/s41809-019-00033-5>
- Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain & Language*, 24, 1–19.
- Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, 20, 525–540.
- Kashima, Y. (2016). *Cultural Dynamics. Current Opinion in Psychology*, 8, 93–97. <https://doi.org/10.1016/j.copsyc.2015.10.019>
- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48(7), 805–825. <https://doi.org/10.1016/j.cortex.2011.04.006>
- Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences of the United States of America*, 166(23), 11213–11222. <https://doi.org/10.1073/pnas.1900952116>
- Kremer, G., & Baroni, M. (2011). A set of semantic norms for German and Italian. *Behavior Research Methods*, 43(1), 97–109. <https://doi.org/10.3758/s13428-010-0028-x>
- Landau, B., & Gleitman, L. R. (1985). *Cognitive science series, 8. Language and experience: Evidence from the blind child*. Harvard University Press.
- Langley, P., Iba, W., & Thompson, K. (1992). An Analysis of Bayesian Classifiers. *Proceedings of the Tenth National Conference on Artificial Intelligence*, 223–228.
- Lehman, D. R., Chiu, C. Y., & Schaller, M. (2004). Psychology and culture. *Annual Review of Psychology*, 55, 689–714. <https://doi.org/10.1146/annurev.psych.55.090902.141927>
- Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. (2013). BLIND: A set of semantic feature norms from the congenitally blind.

- Behavior Research Methods*, 45(4), 1218–1233. <https://doi.org/10.3758/s13428-013-0323-4>
- Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, 130(1), 3–28. <https://doi.org/10.1037/0096-3445.130.1.3>
- 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.
- Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences of the United States of America*, 110(35), 14196–14201. <https://doi.org/10.1073/pnas.1303312110>
- Marmor, G. S. (1978). Age at onset of blindness and the development of the semantics of color names. *Journal of Experimental Child Psychology*, 25(2), 267–278. [https://doi.org/10.1016/0022-0965\(78\)90082-6](https://doi.org/10.1016/0022-0965(78)90082-6)
- Mazucca, C., Majid, A., Lugli, L., Nicoletti, R., & Borghi, A. M. (2020). Gender is a multifaceted concept: Evidence that specific life experiences differentially shape the concept of gender. *Language and Cognition*, 12(4), 649–678. <https://doi.org/10.1017/langcog.2020.15>
- McCauley, T. G., Billingsley, J., & McCullough, M. E. (2022). An evolutionary psychology view of forgiveness: Individuals, groups, and culture. *Current Opinion in Psychology*, 44, 275–280. <https://doi.org/10.1016/j.copsyc.2021.09.021>
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4), 547–559. <https://doi.org/10.3758/BF03192726>
- Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). Semantic memory: A feature-based analysis and new norms for Italian. *Behavior Research Methods*, 45(2), 440–461. <https://doi.org/10.3758/s13428-012-0263-4>
- Nisbett, R. E., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences of the United States of America*, 100(19), 11163–11170. <https://doi.org/10.1073/pnas.1934527100>
- Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, 9(10), 467–473. <https://doi.org/10.1016/j.tics.2005.08.004>
- Nisbett, R. E., Choi, I., Peng, K., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108(2), 291–310. <https://doi.org/10.1037/0033-295X.108.2.291>
- Ojalehto, B., & Medin, D. L. (2015). Perspectives on culture and concepts. *Annual Review of Psychology*, 66, 249–275.
- Ostarek, M., & Huettig, F. (2017). A task-dependent causal role for low-level visual processes in spoken word comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(8), 1215–1224. <https://doi.org/10.1037/xlm0000375>
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.
- Patterson, O. (2014). Making sense of culture. *Annual Review of Sociology*, 40(1), 1–30.
- Pérez-Pereira, M. (2006). Language development in blind children. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (Vol. 6, Second ed., pp. 357–361). Elsevier.
- Plaut, D. C., & Shallice, T. (1991). Effects of word abstractness in a connectionist model of deep dyslexia. *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 73–78). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377–500.
- Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. MIT Press.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews. Neuroscience*, 6(7), 576–582. <https://doi.org/10.1038/nrn1706>
- Quinlan, J. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
- Recchia, G., & Jones, M. N. (2012). The semantic richness of abstract concepts. *Frontiers in Human Neuroscience*, 6, 315. <https://doi.org/10.3389/fnhum.2012.00315>
- Reid, J. N., & Katz, A. (2022). The RK processor: A program for analysing metaphor and word feature-listing data. *Behavior Research Methods*, 54(1), 174–195. <https://doi.org/10.3758/s13428-021-01564-y>
- Roberson, D., Davidoff, J., & Davies, I. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, 129(3), 369–398. <https://doi.org/10.1037/0096-3445.129.3.369>
- Romney, A. K., & Moore, C. C. (1998). Toward a theory of culture as shared cognitive structures. *Ethos*, 26(3), 314–337. <https://doi.org/10.1525/eth.1998.26.3.314>
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382–439. [https://doi.org/10.1016/0010-0285\(76\)90013-X](https://doi.org/10.1016/0010-0285(76)90013-X)
- Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, 15(3), 346–378. [https://doi.org/10.1016/0010-0285\(83\)90012-9](https://doi.org/10.1016/0010-0285(83)90012-9)
- Russell, B. (1997). *The problems of philosophy*. Oxford University Press.
- Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, 3, 83–119.
- Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1), 82–102. <https://doi.org/10.1037/0278-7393.9.1.82>
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27(5), 499–520. [https://doi.org/10.1016/0749-596X\(88\)90022-8](https://doi.org/10.1016/0749-596X(88)90022-8)
- Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, 32, 244–259.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1), 41–78. [https://doi.org/10.1207/s15516709cog2901\\_3](https://doi.org/10.1207/s15516709cog2901_3)
- Sunohara, M., Sasaki, J., Kogo, S., & Ryder, A. G. (2022). Japanese clinical psychologists' consensus beliefs about mental health: A mixed-methods approach. *Japanese Psychological Research*. <https://doi.org/10.1111/jpr.12410>
- Talmy, L. (2000). *Toward a cognitive semantics*. In *Volume II: Typology and process in concept structuring*. The MIT press.
- Vergallito, A., Petilli, M. A., & Marelli, M. (2020). Perceptual modality norms for 1,121 Italian words: A comparison with concreteness and imageability scores and an analysis of their impact in word processing tasks. *Behavior Research Methods*, 52, 1599–1616. <https://doi.org/10.3758/s13428-019-01337-8>
- Vivas, J., Vivas, L., Comesaña, A., Coni, A. G., & Vorano, A. (2017). Spanish semantic feature production norms for 400 concrete concepts. *Behavior Research Methods*, 49(3), 1095–1106. <https://doi.org/10.3758/s13428-016-0777-2>
- Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1256–1271. <https://doi.org/10.1037/0278-7393.25.5.1256>
- Waxman, S., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early

- essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, 43(2), 294–308. <https://doi.org/10.1037/0012-1649.43.2.294>
- Weiler, S. M., & Jacobsen, T. (2021). “I’m getting too old for this stuff”: The conceptual structure of tattoo aesthetics. *Acta Psychologica*, 219. <https://doi.org/10.1016/j.actpsy.2021.103390>
- Wenchi, Y. E. H., & Barsalou, L. W. (2006). The situated nature of concepts. *American Journal of Psychology*, 119(3), 349–384. <https://doi.org/10.2307/20445349>
- Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, 29(5), 719–736. [https://doi.org/10.1207/s15516709cog0000\\_33](https://doi.org/10.1207/s15516709cog0000_33)
- Wilensky, U. (1999). *NetLogo*. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling. Northwestern University.
- Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, 2(JUN). <https://doi.org/10.3389/fpsyg.2011.00116>
- Yap, M. J., & Pexman, P. M. (2016). Semantic richness effects in syntactic classification: The role of feedback. *Frontiers in Psychology*, 7(SEP). <https://doi.org/10.3389/fpsyg.2016.01394>
- Zimler, J., & Keenan, J. M. (1983). Imagery in the congenitally blind: How visual are visual images? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(2), 269–282. <https://doi.org/10.1037/0278-7393.9.2.269>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.