

The role of variability in the property listing task

Sergio E. Chaigneau^{1,2} · Enrique Canessa^{1,3} · Carlos Barra⁴ · Rodrigo Lagos^{5,6}

Published online: 22 June 2017
© Psychonomic Society, Inc. 2017

Abstract It is generally believed that concepts can be characterized by their properties (or features). When investigating concepts encoded in language, researchers often ask subjects to produce lists of properties that describe them (i.e., the Property Listing Task, PLT). These lists are accumulated to produce Conceptual Property Norms (CPNs). CPNs contain frequency distributions of properties for individual concepts. It is widely believed that these distributions represent the underlying semantic structure of those concepts. Here, instead of focusing on the underlying semantic structure, we aim at characterizing the PLT. An often disregarded aspect of the PLT is that individuals show intersubject variability (i.e., they produce only partially overlapping lists). In our study we use a mathematical analysis of this intersubject variability to guide our inquiry. To this end, we resort to a set of publicly available norms that contain information about the specific properties that were informed at the individual subject level. Our results suggest that when an individual is performing the PLT, he or she generates a list of properties that

is a mixture of general and distinctive properties, such that there is a non-linear tendency to produce more general than distinctive properties. Furthermore, the low generality properties are precisely those that tend not to be repeated across lists, accounting in this manner for part of the intersubject variability. In consequence, any manipulation that may affect the mixture of general and distinctive properties in lists is bound to change intersubject variability. We discuss why these results are important for researchers using the PLT.

Keywords Property listing task · Conceptual property norms · Inter-subject variability · Conceptual agreement theory

Introduction

It is generally considered that concepts (or at least concrete concepts) can be characterized by their properties (also called features) and those properties' distributions. In particular, to study natural concepts encoded in language, researchers often ask subjects to provide a list of properties related to a given concept (i.e., the Property Listing Task, PLT; also known as the Feature Listing Task). Accumulating these lists across multiple subjects, Conceptual Property Norms (CPNs) are built. These norms are widely believed to contain information about the underlying semantic structure of those concepts contained in CPNs.

Despite the wide use of the PLT, little attention has been devoted to understanding it. In particular, an often overlooked fact is that the PLT produces intersubject variability (i.e., different people produce somewhat different lists of properties). As we will discuss briefly, this is not a trivial fact. In the current study, we focus on how to understand this variability, how it should be measured, and we explore two variables that seem good *a priori* candidates to affect it: concept exposure and property generality. To this end, we analyze a large set of normed concepts, with a

✉ Sergio E. Chaigneau
sergio.chaigneau@uai.cl

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

² Center for Social and Cognitive Neuroscience, School of Psychology, Universidad Adolfo Ibáñez, Las Condes, Santiago, Chile

³ Faculty of Engineering and Science, Universidad Adolfo Ibáñez, Viña del Mar, Chile

⁴ Independent Researcher, Viña del Mar, Chile

⁵ Universidad Autónoma de Chile, San Miguel, Chile

⁶ Programa Magister en Bioestadística, Escuela de Salud Pública, Universidad de Chile, Santiago, Chile

focus on the PLT rather than on the underlying semantic or conceptual structure. In what remains of this introduction, we discuss the relation between concepts and properties, how the PLT is used to build CPNs and what these norms are used for, and how variability is currently handled in CPNs and other studies.

Concepts and conceptual properties

Two central tenets of the prevailing view of concepts are that representations are perceptually based and that they are learned (Prinz, 2005). For example, through experience, people learn that a face is characterized by properties such as a given nose length, eye separation, mouth width, etc. (e.g., Tversky, 1977). This has come to be a substantial part of the standard theory of concepts, and it holds that concepts are learned by observing category exemplars, extracting the relevant properties (Schyns, Goldstone, & Thibaut, 1998), and estimating the frequency distribution of those relevant properties (Ashby & Alfonso-Reese, 1995; Griffiths, Sanborn, Canini, Navarro, & Tenenbaum, 2011). This then allows organizing a semantic structure and making category judgments (e.g., How typical is a given exemplar of category X? Does the exemplar belong to category X or category Y? How central is property *j* for category X?).

The nature of the extraction and estimation processes are the concern of many theories, which attempt to explain how properties are combined by the cognitive system in order to achieve correct classification. We pay limited attention to those theories, but the interested reader may consult any of several books devoted to that topic (e.g., Murphy, 2002; Pothos & Wills, 2011). Much of this research has been done in laboratory experiments with artificial categories. However, it is generally assumed that the same processes extrapolate to the real world. Notably, it is assumed that conceptual properties play an important role in categorization both inside and outside the laboratory (e.g., McRae, Cree, Seidenberg, & McNorgan, 2005).

A well-known finding about concepts is that they relate only probabilistically to conceptual properties (Rosch, 1973). This means, among other things, that two concepts that may be applied to a situation or object are generally not able to be discriminated between through a logical rule (i.e., by necessary and sufficient properties), but show a typicality structure instead (Rosch, Simpson, & Miller, 1976). Those exemplars that exhibit frequent properties are more typical than those exhibiting less frequent properties (Rosch & Mervis, 1975) (e.g., an ostrich would be a low typicality exemplar of the BIRD category, whereas a dove would be a typical exemplar). The typicality structure also means that an object can be a member of different categories, though probably to different degrees (e.g., a Chihuahua may be a low typicality exemplar of the DOG category, while simultaneously being a relatively more typical case of the PET category). Importantly for the work we report here, note that all this implies

that concepts will to a large extent share properties (e.g., being friendly is a property of DOG, but also of DOLPHIN; being made of metal may be a property of FLUTE, but also of FORK). Henceforth, we refer to this as the degree of generality of properties. General properties are those that are shared across many concepts (e.g., being made of metal). Low generality properties are sometimes referred to as distinctive properties (Cree & McRae, 2003), i.e., those that relate to only a few concepts (e.g., being used to treat headaches). Note, however, that general and distinctive properties are simply two extremes of the same continuous dimension.

Conceptual property norms

When researchers want to learn about natural concepts outside of the control offered by the laboratory, they might collect data about linguistic productions, and build Conceptual Property Norms (CPNs). Whereas categories used in laboratory experiments are often arbitrary (e.g., patterns of dots on a screen), categories in the real world are typically conventional and encoded in language (e.g., DOG, CAR, ALLIGATOR). An important tool used to construct CPNs is the Property Listing Task (PLT; e.g., Cree & McRae, 2003; Hampton, 1979; McRae, Cree, Seidenberg, & McNorgan, 2005; Rosch & Mervis, 1975; Rosch et al., 1976). In this task, subjects are asked to produce features typically associated with a given concept (e.g., to the question “What properties are typically true of a DOG?”, “has four legs” may be produced). After coding responses into property types (i.e., grouping together responses with only superficial differences across subjects, e.g., “has four legs” and “is a quadruped”), responses are accumulated to build the CPNs. These norms are frequency distributions of property types produced by a sample of subjects for a large number of concepts encoded in language.

Conceptual Property Norms contain a wealth of information about concepts and their properties. They are widely believed to measure underlying conceptual content, which in turn is thought to reflect the statistical structure of the categories under study (e.g., McNorgan, Kotack, Meehan, & McRae, 2007). This information may be used for different purposes. Norming studies can be used to test theories. For example, it is possible to make predictions based on these norms’ distributional properties, e.g., the pattern of inter-property correlations that typically emerge in these norms (e.g., characterizing differences between tools and animals; Cree & McRae, 2003), or the way properties cluster by sensory modality (e.g., “it makes pleasant sounds” and “it can sing,” both correspond to the auditory modality; e.g., Vigliocco, Vinson, Lewis, & Garrett, 2004; Wu & Barsalou, 2009). The norms may also be used as a source of control variables for experiments, e.g., controlling for frequency of properties, cue validity, number of properties produced, and other such variables (e.g., McRae, Cree, Westmacott, & de Sa, 1999).

Conceptual variability

A fact that may be overlooked is that conceptual content varies across individuals (i.e., intersubject conceptual variability) and even within individuals over time (i.e., intrasubject conceptual variability; see Barsalou, 1987, 1993). When asked to list conceptual properties for a concept, different people produce somewhat different properties and the same person may produce somewhat different properties over time. This is often considered problematic. Based on this, Barsalou (1987, 1993) doubted that concepts are entities in the mind. Converse (1964), based on the instability of public opinion, doubted that public opinion exists. More recently, Gabora, Rosch, and Aerts (2008) suggested that concepts should be modeled with quantum mechanics formalisms, because in their view, concepts, like quantum entities, do not have definite properties unless in a particular measurement situation. Notably, this whole debate pre-dates psychological discussions (see Frege, 1893; Glock, 2009; Russell, 1997).

In the context of CPNs, variability has been treated as measurement error; a less radical option than those discussed above. In particular, a common practice in CPN studies is to weed out from the norms those properties produced by less than a given proportion of the sample, on the account that low frequency responses would obscure the general patterns in the data (e.g., in McRae et al., 2005, features produced by less than five out of 30 participants were not included in the analyses).

In the current work, rather than sweeping it under the rug, we deliberately focus on conceptual variability. In our analysis, the issue of variability is intimately linked with the PLT. Several reasons for conceptual variability in the PLT come to mind. People may produce different conceptual content because of contextual effects (i.e., slight differences in the context present when individuals are producing conceptual properties result in different conceptual content being accessed), because of differences in interest or salience (i.e., people may be differentially motivated to produce content for a given concept), because of exposure (i.e., being differentially exposed to a concept may lead to changes in the identity and number of properties learned and reported), or because of differences in the level of generality of properties being produced (i.e., people producing more general properties may achieve less intersubject variability than people producing more distinctive properties). Of these potential factors, in the current work we focus on the latter two.

Theory

As noted by Barsalou and collaborators (Santos, Chaigneau, Simmons, & Barsalou, 2011; Wu & Barsalou, 2009), though the PLT is used by researchers across psychology (e.g., in cognitive psychology, social psychology, cognitive neuroscience, neuropsychology, consumer psychology), little is known about

its characteristics. Hence, our goal in the current work is to analyze the PLT. In particular, we will show how to compute an index of intersubject variability, and then use this formulation to test the effects of exposure and generality on PLT variability. A mathematical tool will aid us in this undertaking. In the current work, we use the probability of *true agreement* (event labeled $a1$, being $p(a1)$ its corresponding probability) as a measure of intersubject conceptual variability (Canessa & Chaigneau, 2016; Chaigneau, Canessa, & Gaete, 2012).

In a CPN study, after raw data has been collected and properties have been coded, all that researchers have are different lists of properties produced by individual subjects. Each of these lists can be thought of as a sample of conceptual content taken from the total population of properties produced by participants in the CPN study. These samples vary in length and in content (respectively, the number of properties and the specific properties they contain). We can quantify these differences via the probability of true agreement ($p(a1)$). This is defined as the probability that a property that is picked randomly from one of these samples is also found in another randomly picked sample. Note that $p(a1)$ is a measure of intersubject variability in conceptual content because it will achieve a value of 1 if and only if all subjects produce the same list of properties in the PLT (i.e., same content and same length), and it will achieve a value of 0 (zero) if and only if all subjects produce completely different lists of properties in the PLT. Thus, $p(a1) = 1$ means that there is no conceptual variability, and $p(a1) = 0$ means that conceptual variability is maximal. A desirable property of $p(a1)$ is that it is strictly comparable across conditions. As will become evident from our discussions, this probability is applicable across concepts, regardless of how their properties are distributed. (Note that our measure could be applied to other data collection procedures in which participants' responses are accumulated to obtain frequency distributions, and where intersubject variability is likely to arise, such as word association and top-of-mind procedures. However, we will not explore this issue further.)

Expression (1) puts these ideas into mathematical form. Due to space restrictions, we do not derive the equation here. The interested reader may find its development and a fuller explanation in Chaigneau, Canessa and Gaete (2012). In lieu of that, after explaining the terms in the equation, we provide a simplified example of its computation illustrating its meaning.

$$p(a1) = \frac{1}{s_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \#(S_i^1 \cap S_j^1) p_i p_j \quad (1)$$

Equation (1) is the summation of the expected value of the number of common properties between all possible pairs of samples (lists of size s_j) of properties randomly obtained from a concept in a PLT (i.e., the S_i^1 and S_j^1 in the equation, and the corresponding cardinality of the intersection, i.e., the $\#(S_i^1 \cap S_j^1)$ term), taking into account the probabilities of each sample

(i.e., the p_i and p_j). Assuming that all properties in the samples have the same probability of being selected from a given sample, $p(aI)$ is calculated as the summation in (1) divided by s_I . The terms in Eq. (1) are defined in Table 1.

The following is a worked out example of how $p(aI)$ is computed. Let us assume a PLT for the concept DOG, and that after coding, three properties are obtained. In this case, $a =$ “is a pet”, $b =$ “has four legs,” $c =$ “is hairy.” Also assume that, on average, each of the lists produced by subjects has two properties, i.e., $s_I = 2$. Figure 1 shows this hypothetical example.

The relevant quantities that are necessary to compute $p(aI)$ are:

$$k_I = 3$$

$$s_I = 2$$

and

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

That is, the $n_I S_i^1$ and S_j^1 possible samples are $\{ab, ac, bc\}$

Though Eq. (1) allows handling any distribution of properties and sample probabilities, assume for simplicity that each sample has an equal probability of being selected and, thus, $p_i = p_j = 1/n_I = 1/3$ (for an example with an unequal frequency distribution of properties, see Appendix A). Then, using Eq. (1),

$$\begin{aligned}
 p(aI) &= \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \#(S_i^1 \cap S_j^1) \frac{1}{3} \frac{1}{3} \\
 &= \frac{1}{18} \sum_{i=1}^3 \sum_{j=1}^3 \#(S_i^1 \cap S_j^1) \tag{2}
 \end{aligned}$$

In Eq. (B4), the double summation corresponds to the sum of counts of coincidences between each sample S_i^1 and S_j^1 , for example:

$$\#(S_1^1 \cap S_1^1) = \#(ab \cap ab) = 2$$

$$\#(S_1^1 \cap S_2^1) = \#(ab \cap ac) = 1 \text{ and so on until,}$$

$$\#(S_3^1 \cap S_3^1) = \#(bc \cap bc) = 2$$

For this example, each term of the double summation is

$$\#(S_i^1 \cap S_j^1) = \{2, 1, 1, 1, 2, 1, 1, 1, 2\}$$

Hence, $p(aI) = 12/18 = 2/3$

From inspecting the example presented in Fig. 1, it is easily seen that a given property contained in a sample of size s_I will be also contained in two out of all possible three randomly selected samples of the same size. Note that this is precisely the value obtained above (i.e., 2/3). Although not readily apparent, we treated the example above as a case of equiprobable properties (all properties have an equal chance of being selected in a sample, and thus all samples have an equal probability of occurring). Importantly, note also that $s_I/k_I = 2/3$. This is

not coincidental. In fact, it is always true that for equiprobable properties, Eq. (1) reduces to:

$$p(aI) = \frac{s_I}{k_I} \tag{3}$$

This is an important property, and it aids in understanding $p(aI)$ (a demonstration may be found in Canessa & Chaigneau, 2016). Examining Eq. (3) readily shows that the probability of true agreement is a measure of the agreement in conceptual content among subjects, so that the larger the total number of properties describing a concept (k_I) compared with the average number of properties that describe the concept in people’s conceptual content samples (s_I), the smaller $p(aI)$, and vice-versa. Thus, it is clear that k_I and s_I are mathematically related to $p(aI)$. All other things being equal, as k_I increases while leaving s_I constant, $p(aI)$ decreases. A smaller s_I relative to k_I implies that it is more difficult to find coincidences between the s_I sized samples. In contrast, as s_I increases while leaving k_I constant, $p(aI)$ increases.

The same discussion above about s_I and k_I , is valid for the general (and more realistic) case of non-equiprobable properties expressed in Eq. (1). To understand why this is the case, first consider that Eq. (3) is also a lower bound on $p(aI)$, i.e., $p(aI)$ cannot take a value smaller than the one in Eq. (3) (see a demonstration in Appendix B). There is an intuitive explanation for the lower bound. If all samples (i.e., combinations of properties) were equiprobable, and we increased the probability of one of those samples (i.e., if we made the samples not equiprobable), then there would be an increased probability of that sample being chosen for comparison. By Eq. (1), this means that $p(aI)$ increases. In contrast, if we were to decrease the probability of one of those samples, other samples would increase in probability (given that the sum of all the probabilities for all samples must be one), meaning that $p(aI)$ could not in any case be smaller than s_I/k_I . Consequently, $p(aI)$ calculated by Eq. (1) will always be larger than s_I/k_I .

We can express the idea above as $p(aI) = \theta s_I/k_I$, with $\theta > 1$. Of course, the value of factor θ will vary depending on the parameters p_i, p_j (i.e., the frequency distribution of properties), but the key issue is that θ will always be bigger than one. Now, given that $p(aI) = \theta s_I/k_I$, and $\theta > 1$, this means that an increase in s_I will also increase $p(aI)$ (apart from the effect of θ) and an increase in k_I will also decrease $p(aI)$ (apart from the effect of θ). In our analyses, we will use this mathematical relation as a validity check. If our formulations are correct and in fact have empirical content, they should pan out in data from any CPN study.

Equation (1) does not exhaust the issues of intersubject variability. What is not clear from Eq. (1) is that k_I and s_I are also related to each other. An analysis of the role of the PLT in CPNs shows that the only way in which k_I can increase is by an increase in s_I (i.e., everybody produces longer lists

Table 1 Definition of variables used in equation (1)

Variable symbol	Definition
s_I	The average size of a sample of properties obtained from a concept (i.e., the average length of lists of properties obtained in a PLT for a concept) ($s_I \leq k_I$), where k_I is the cardinality of the concept (i.e., the total number of properties obtained in a PLT for a concept, after coding)
n_I	The number of possible samples of size s_I obtained from the k_I properties that belong to a concept. Thus, $n_I = \frac{k_I!}{(k_I - s_I)! s_I!}$
$\#(S_i^1 \cap S_j^1)$	The cardinality of the intersection between a sample “ i ” drawn from a concept (S_i^1) and another independent sample “ j ” drawn from the same concept (S_j^1), i.e., the number of properties that belong to both samples or lists obtained in a PLT for a concept
$p_i \cdot p_j$	Probability of obtaining sample S_i^1 or S_j^1 in Eq. (1)

with the same content, and thus k_I will be larger), or by an increase in the diversity from one list to another (i.e., if lists contain low frequency properties that tend not to be repeated in other lists, then k_I will increase because k_I is the combination of all the different lists). Furthermore, it seems that the longer the individual lists are, the higher the probability that those lists will contain low frequency properties that are not repeats in other lists (e.g., idiosyncratic properties). When lists are accumulated across subjects, as is done in a CPN study, it seems intuitive that the longer the lists, the higher the probability that those lists will contain properties that are not frequently repeated in other lists, thus increasing the total number of properties listed across subjects. Note that this implies an exponential relation. This is our Hypothesis 1a, which states that s_I and k_I will be positively related and that their relation will be nonlinear.

Another consequence of this analysis is that we can predict that, though both k_I and s_I should correlate with $p(aI)$ (correspondingly, a negative and a positive correlation), it is likely that k_I will show a higher correlation than s_I , given that the s_I parameter contains information only about the number of properties, whereas the k_I parameter contains additional information about the diversity of properties across subjects, and this latter information is precisely what $p(aI)$ captures. This is our Hypothesis 1b.

From the discussion in the preceding paragraphs, we hope it is apparent that the list length in the PLT (variable s_I) should be the critical variable regarding conceptual variability. As it changes, it drives changes in k_I and in $p(aI)$. Now, recall that in the section on conceptual variability we suggested two

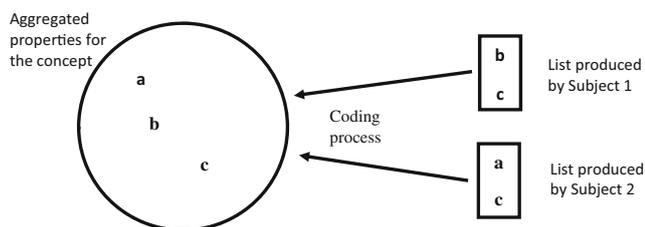


Fig. 1 A concept with its corresponding properties $\{a, b, c\}$, with k_I (number of properties) = 3, and two lists of properties obtained in a PLT, each on average of size $s_I = 2$

different factors that could affect the length and composition of lists in the PLT. We will discuss each in turn, explaining why they should affect s_I (and consequently, k_I and $p(aI)$).

One such factor is exposure. Assuming that increasing exposure leads to learning more about a concept, and therefore to longer lists, we predict that exposure will be positively correlated to s_I and to k_I , but negatively to $p(aI)$. Thus, because as people have more to report due to exposure, there will presumably be a greater chance that their reported list will contain different properties, decreasing $p(aI)$. This is our Hypothesis 2.

Another factor that could affect s_I is the generality of properties. As we have already discussed, the statistical structure of concept to properties relations implies that some properties are shared across concepts. Here, we refrain from speculating about why this is the case. The fact is that properties vary in how general or distinctive they are, depending on how many concepts contain them. Assuming that general properties are fewer than distinctive properties, we predict that generality will be associated with shorter lists, a lower chance of having different properties across lists, and an increased $p(aI)$. This is our Hypothesis 3.

Data

In this work, we used two publicly available norm studies. Our main data comes from the Centre for Speech, Language and the Brain concept property norms (from here and on, the CSLB norms; Devereux, Tyler, Geertzen, & Randall, 2014). To perform other secondary analyses, we used the SUBTLEX-UK British English word frequency database (from here and on, the SUBTLEX-UK norms; van Heuven, Mandera, Keuleers, & Brysbaert, 2014). We will describe each in turn. Table 2 shows a list of all variables used throughout the analyses that follow. For each variable, we indicate its name, in which of the two sets of norms it is found, and how it is computed.

Participants in the CSLB norms ($n = 123$; 18–40 years of age) were presented with a set of 30 concrete concepts, and

Table 2 Variables analyzed in the current work

Variable name	Definition
s_l	Mean list length. The average number of properties produced for a given concept by participants in the CSLB norms study
k_l	The total number of distinct properties produced for a given concept by participants in the CSLB norms study
$p(al)$	The true agreement probability of a given concept in the CSLB norms. To compute this value, it is not practical to use Eq. (1), because it soon leads to a combinatorial explosion. Instead, we used a computational program described in Canessa and Chaigneau (2016), which uses the frequency distribution of properties and the mean list length to closely approximate the exact value obtained by the equation
Generality	This is a property level value. It is the number of concepts in the CSLB norms in which a given property occurs. The greater the number of concepts in which it appears, the greater the generality of the property. When analyses were performed at the property level, we used a base 10 logarithmic transformation (see the text for further detail). (Note that in other studies sometimes the inverse of this variable is used, in which case it is called “distinctiveness”)
Average generality	The average generality of the k_l properties listed for a given concept in the CSLB norms. These average values were not transformed
Frequency	This is a property by concept level value. It is the number of subjects that produced a given property for a given concept in the CSLB norms
Exposure	This is a concept level variable. It is the frequency in which a concept in the CSLB norms is found in the SUBTLEX-UK word–frequency norms. It is labeled “exposure” because we use it as a proxy for how often someone may be exposed to the concept in natural language (Brysbaert & New, 2009). Because of this variable’s distribution, we always submitted it to a base 10 logarithmic transformation (see text for further detail). Only 603 concepts in the SUBTLEX-UK norms were unambiguously present in the CSLB norms. Therefore, analyses on this variable were done on those 603 concepts

asked to produce at least five properties for each. The task was performed online, and subjects could participate in more than one session. Interestingly for our purposes, the CSLB norms contain more detailed information than other comparable norms. Not only does it provide the frequency distributions for each of the 638 concepts, but also the specific properties produced by each subject (allowing us to compute the s_l parameter). Just as importantly, the CSLB norms retain all properties produced by at least two subjects. This allowed us to compute a $p(al)$ that is representative of the original intersubject variability in the norms.

The SUBTLEX-UK word frequency norms were obtained by collecting words from the subtitles of nine British Broadcasting Corporation channels during a 3-year period (over 45,000 broadcasts). This produced a register of 159,235 word types, and a total number of approximately 201.3 million word tokens. To combine both sets of norms, we searched in the SUBTLEX-UK norms for the 638 concrete nouns in the CSLB norms. Of those, we were able to

unambiguously identify 603 which were in both sets of norms. Thus, we supplemented the CSLB norms with data for frequency of occurrence in natural language for 603 concepts.

We inspected all our variables to check for departures from normality. As Table 3 illustrates, our variables are in general slightly positively skewed and leptokurtic. However, visual inspection showed they do not deviate noticeably from normality.

Analyses

Note that our theory views all our different variables as inter-related (as is evident from Eqs. (1), (B3), and (3)). However, for practical purposes, we fractioned these relations and used multiple regression as our fundamental tool. Because we had reasons to predict non-linear relations between some of our variables (as per our discussion of the k_l to s_l relation), we

Table 3 Descriptives for variables in the analyses, including skewness (skew.) and kurtosis (kurt.) statistics

	Min.	Max.	Mean	S.D.	Skew.	Skew. Std. Err.	Kurt.	Kurt. Std. Err.
s_l	4.1	12.5	6.9	1.2	.599	.097	1.089	.193
k_l	13	71	33	7	.739	.097	1.666	.193
$p(al)$.21	.56	.33	.05	.492	.097	.616	.193
Log10(exposure)	.70	4.7	3.1	.66	-.048	.100	-.085	.199
Avg. Generality	7.5	103.7	52.1	17.9	.168	.097	-.250	.193

For variable definitions, see Table 2

Min. minimum, Max. maximum, SD standard deviation, Std. Err standard error

Table 4 Regression models for the question of how the number of properties affects the PLT

Model	Model <i>F</i> test	Linear coeff. <i>t</i> test	Quadratic coeff. <i>t</i> test
1. $p(aI) = 0.602 - 0.012 (-1.614) k_I + 1.113 \times 10^{-4} (1.064) k_I^2$	$p < .001$	$p < .001$	$p < .001$
2. $p(aI) = 0.101 + 0.060 (1.310) s_I - 0.004 (-1.231) s_I^2$	$p < .001$	$p < .001$	$p < .001$
3. $p(aI) = 0.298 + 0.044 (0.089) s_I$	$p < .05$	$p < .05$	
4. $p(aI) = 0.316 + 2.11 \times 10^{-4} (0.068) s_I^2$	$p = .088$		$p = .088$
5. $k_I = 20.412 - 0.436 (-0.072) s_I + 0.317 (0.767) s_I^2$	$p < .001$	$p = .743$	$p < .001$
6. $k_I = 18.849 + 0.288 (0.696) s_I^2$	$p < .001$		$p < .001$

Note: Models in bold type are those judged best according to our criteria (see text for details). For each model, values in parenthesis are standardized coefficients (β s)

used linear and quadratic terms and then chose the best model according to the following criteria. We first built a complete model with linear and quadratic terms. A Variance Inflation Factor (VIF) greater than 10 suggested a problem of multicollinearity among predictor variables (Hair, Anderson, Tatham, & Black, 1992). We handled this multicollinearity by building separate linear and quadratic models. If the predictors in the combined model were non-significant, but became significant when considered separately, then this corroborated the multicollinearity problem. In this latter case, the complete model was never preferred over its alternatives. When two or more models remained possible, they were evaluated depending on statistical significance, magnitude of their unstandardized coefficients, overall explained variance, and

parsimony (i.e., all other things being equal, we preferred linear to quadratic models).

How does the number of properties affect the PLT?

Does increasing k_I decrease $p(aI)$? Recall that Eqs. (1) and (3) imply that variability in conceptual content should increase (i.e., a decrease in $p(aI)$) as a function of the total number of properties produced by a group for a given concept (i.e., k_I). This is our validity check. An additional issue we wanted to address was whether the relation between k_I and $p(aI)$ is linear or not. Thus, we calculated k_I and $p(aI)$ for each of the 638 concepts in the CSLB norms, and computed a regression equation with k_I as predictor and $p(aI)$ as criterion variable.

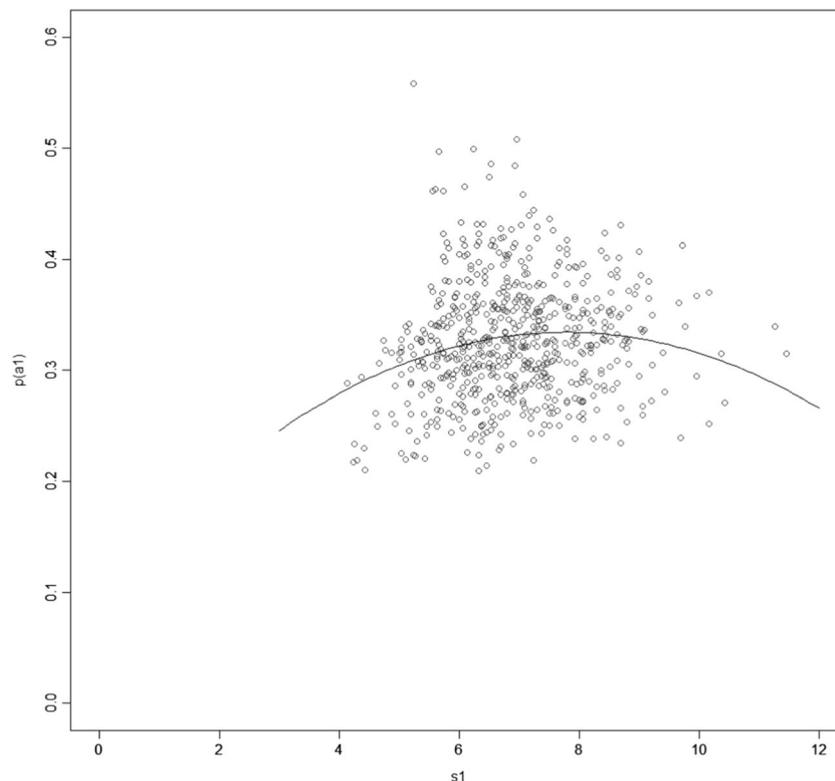


Fig. 2 Graph of model 2 in Table 4, showing the relation between s_I (mean list length) and $p(aI)$ (true agreement probability) for 638 concepts

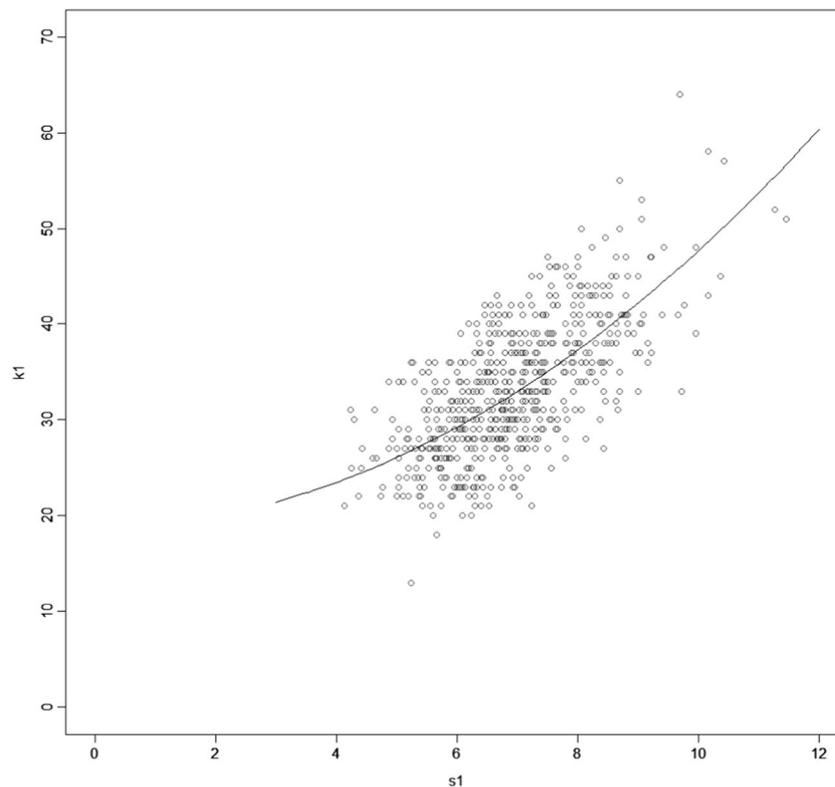


Fig. 3 Graph of model 6 in Table 4, showing the relation between s_l (mean list length) and k_l (total number of properties) for 638 concepts generally, but the quadratic term suggests that

To test for nonlinearity, we added a quadratic term to the equation (i.e., k_l^2).

As Table 4 (model 1) shows, the complete model is statistically significant, as are both its linear and quadratic coefficients. Furthermore, the model has an R^2 of 0.353, and its VIF is lower than 10 (thus, we had no concerns of multicollinearity). Hence, we can say that indeed as k_l increases, $p(al)$ decreases, but that this decrement tends to stabilize at higher k_l values.

Does increasing s_l increase $p(al)$? This is also part of our validity check. Equations (1) and (3) imply that as s_l increases, so should $p(al)$. In words, all other things being equal, as people individually produce more content for a given concept, this content should become less variable. Just as above, an additional issue we wanted to address was whether the relation between s_l and $p(al)$ is linear or not. Thus, we calculated s_l and $p(al)$ for each of the 638 concepts in the CSLB norms, and computed a regression equation with s_l as predictor and $p(al)$ as criterion variable. To test for nonlinearity, we added a quadratic term to the equation (i.e., s_l^2).

Results showed that the complete model is statistically significant, but with a low R^2 of 0.034. As shown in Table 4 (model 2), the linear and quadratic coefficients are both significant. Notably, the model's VIF was 59.13, which led us to suspect multicollinearity. Thus, we built two additional models, each with one term (either the linear or the quadratic

one; see Table 4, models 3 and 4). The linear model is significant, with a very low R^2 of 0.008. The linear coefficient is also significant. The quadratic model is non-significant, with a very low R^2 of 0.005. Because there is no clear evidence of multicollinearity (i.e., both coefficients in model 3 are significant), we decided between models by comparing R^2 s. Here, the complete model's R^2 (model 2) was significantly greater than the single predictor models' (for the linear model, $F(1,635) = 16.9$; $p < .001$; for the quadratic model, $F(1,635) = 19.14$; $p < .001$). Therefore, the complete model is the overall best. Thus, as expected, $p(al)$ increases as s_l also increases. Note that the negative quadratic coefficient in model 2 in Table 4 indicates that as s_l increases, its positive effect on $p(al)$ decreases, such that $p(al)$ exhibits a maximum (see Fig. 2).

One further analysis we needed to perform was to compare the correlation of s_l and $p(al)$ versus that of k_l and $p(al)$. Recall here that our Hypothesis 1b was that s_l is bound to show a smaller covariance with $p(al)$ than k_l , due to k_l growing from an increase in sample size and from the inter-individual differences in sample content. This analysis showed that the r_{xy} between $p(al)$ and k_l is -0.566 ($p < .001$) and the one between $p(al)$ and s_l is 0.089 ($p = 0.024$). A z test for dependent correlations with a shared term (Steiger, 1980) showed that this is a significant difference ($z = -23.5$; $p < .001$).

Table 5 Regression models for the question of how exposure affects the PLT

Model	Model <i>F</i> test	Linear coeff. <i>t</i> test	Quadratic coeff. <i>t</i> test
1. $s_I = 5.171 + 0.480 (0.276) \log_{10}(\text{exposure}) + 0.027 (0.096) \log_{10}(\text{exposure})^2$	$p < .001$	$p = .282$	$p = .707$
2. $s_I = 4.930 + .647 (0.371) \log_{10}(\text{exposure})$	$p < .001$	$p < .001$	
3. $s_I = 5.882 + 0.105 (0.369) \log_{10}(\text{exposure})^2$	$p < .001$		$p < .001$
4. $k_I = 26.147 + 1.458 (0.138) \log_{10}(\text{exposure}) + 0.234 (0.136) \log_{10}(\text{exposure})^2$	$p < .001$	$p = .602$	$p = .608$
5. $k_I = 24.096 + 2.877 (0.273) \log_{10}(\text{exposure})$	$p < .001$	$p < .001$	
6. $k_I = 28.305 + 0.469 (0.273) \log_{10}(\text{exposure})^2$	$p < .001$		$p < .001$
7. $p(aI) = 0.319 + 0.003 (0.031) \log_{10}(\text{exposure})$	$p = .441$	$p = .441$	
8. $p(aI) = 0.323 + 3.69 \times 10^{-4} (0.028) \log_{10}(\text{exposure})^2$	$p = .491$		$p = .491$

Note: Models in bold type are those judged best according to our criteria (see text for details). For each model, values in parenthesis are standardized coefficients (β s)

Because s_I and k_I have divergent effects on $p(aI)$ (i.e., as per Eqs. (1) and (3), and as the analyses above confirm), we performed a partial correlation analysis between s_I and $p(aI)$ controlling for k_I . If our discussion about the relation between s_I and k_I is correct, then controlling for k_I should increase the R^2 between s_I and $p(aI)$, relative to when no control was performed (i.e., in the regression analysis above). As we predicted, the partial correlation between s_I and $p(aI)$ controlling for k_I is 0.801 ($p < .001$). This is an 81-fold increase in explained variance relative to the corresponding correlation with no control (i.e., .089).

Does increasing s_I increase k_I ? Recall that we expected that an increase in individual content being produced (i.e., s_I) should increase the group-level number of properties (k_I). This is our Hypothesis 1a. Furthermore, recall that we expected this relation to be nonlinear. To this end, we computed a regression equation with s_I and s_I^2 as predictors and k_I as criterion.

Results showed that the complete model is statistically significant, with an R^2 of 0.484. As shown in Table 4 (model 5), the linear term is non-significant, while the quadratic term is significant. However, because the VIF is 59.13 (i.e., greater than 10), we suspected multicollinearity and computed a regression equation with only the quadratic term (shown in Table 4, model 6). This model is also significant and exhibits the same R^2 as before (0.484), with a significant quadratic coefficient. Thus, our data corroborates our hypothesized non-linear relation between s_I and k_I (see Fig. 3).

How does exposure affect the PLT?

Does exposure affect s_I ? Recall that we hypothesized that, assuming that being exposed to a concept leads to learning more about it, exposure would lead to people having more conceptual content to inform (i.e., increasing s_I). This is our Hypothesis 2. We used word-frequency as a proxy for exposure. The SUBTLEX-UK norms provided us with a measure of how often someone might encounter a linguistically

encoded concept in natural language settings. A total of 603 concepts in the CSLB norms were also found in the SUBTLEX-UK word-frequency norms, and so the following analyses are based on them. Because of the long-tailed distribution of word frequency and the clustering of data points in the low-frequency values, we computed the base 10 logarithm of the frequency reported in the SUBTLEX-UK norms. This is a usual practice when handling this type of distribution (Kleinbaum, Kupper, & Muller, 1988). Henceforth, we label this variable $\log_{10}(\text{exposure})$.

We submitted the s_I and $\log_{10}(\text{exposure})$ values to a regression analysis with $\log_{10}(\text{exposure})$ and its squared value as predictors (see Table 5, model 1). Results showed that the complete model is statistically significant, with an R^2 of 0.138. However, as shown in Table 5 (model 1), neither the linear term nor the quadratic term are significant. Furthermore, the model's VIF is 45.57, suggesting that variables exhibit multicollinearity. Thus, we performed two separate regressions, one with $\log_{10}(\text{exposure})$ as predictor and another with squared $\log_{10}(\text{exposure})$ (see Table 5, models 2 and 3). The linear model yields a significant R^2 of 0.138, with a significant coefficient. The quadratic model is also significant, with an R^2 of 0.136, and a significant coefficient. Given that the R^2 s for models 2 and 3 are similar, based on parsimony we preferred the linear model. However, recall that the predictor is a logarithmic transformation, meaning that the relation between s_I and the original exposure values is actually non-linear. That is, relative to the original untransformed variable, an increase in exposure at the small exposure range increases s_I much more than the same increase at the large exposure range. In other words, there is a saturation effect between exposure and s_I , where the increase in s_I due to an increase in exposure is much bigger for small values of exposure than for larger ones.

Does exposure affect k_I ? Consistent with the analysis above, we expected that the $\log_{10}(\text{exposure})$ values also predicted k_I . Thus, we computed a regression equation with $\log_{10}(\text{exposure})$ and squared $\log_{10}(\text{exposure})$ as predictors, and k_I as criterion. As shown in Table 5 (model 4), the

complete model is significant, with an R^2 of 0.075 and a VIF of 45.57. Furthermore, neither the linear nor the quadratic coefficients are significant. Thus, due to multicollinearity concerns, and as done before, we computed two separate models. The linear model (Table 5, model 5) yields a significant R^2 of 0.075, and a significant coefficient. The quadratic model (Table 5, model 6) also shows a significant R^2 of 0.075, and a significant coefficient. Given the equivalent R^2 s, for parsimony we chose the linear model. This model shows that exposure indeed increases k_I . Here again, since the predictor is $\log_{10}(\text{exposure})$, the relation between exposure and k_I is actually non-linear, with the same saturation effect explained before.

Does exposure affect $p(aI)$? Recall that part of our Hypothesis 2 holds that exposure should affect $p(aI)$. Thus, as we have shown, exposure has a positive relation with s_I and k_I , and both these variables relate to $p(aI)$. Because from our two previous analyses we know there is a multicollinearity problem with $\log(\text{exposure})$ and $\log(\text{exposure})^2$, we did not compute the complete model. Instead, we went on to directly compute two separate models, regressing $p(aI)$ on $\log_{10}(\text{exposure})$ in the first and on $\log_{10}(\text{exposure})^2$ in the second (respectively, models 7 and 8 in Table 5). Both models are not significant, suggesting an absence of effect of exposure on $p(aI)$. As a way of testing for the independent (i.e., mutually controlled) effects of s_I , k_I and $\log_{10}(\text{exposure})$ on $p(aI)$, we built a third model in which we regressed $p(aI)$ on s_I , k_I and $\log_{10}(\text{exposure})$. This model is significant ($F(3,599) = 619.6$, $MSe = .001$; $p < .001$) and has an R^2 of 0.756. The coefficient for s_I is 0.042 ($t(599) = 31.3$; $p < .001$), for k_I is -0.009 ($t(599) = -42.884$; $p < .001$) and for $\log_{10}(\text{exposure})$ is 0.002 ($t(599) = 1.02$; $p = .306$). Thus, the effects on $p(aI)$ are positive for s_I , negative for k_I , and null for exposure.

In hindsight, though contrary to our hypothesis, this result does not appear surprising. From our analysis in the Theory section, we know that k_I and s_I are related. The only way in which k_I can increase is if subjects in a CPN study produce longer lists that have the same content, or by an increase in the diversity from one list to another (i.e., if lists contain low frequency properties that tend not to be repeated in other lists). Our results suggest that in the CSLB norms, exposure increases both components of lists, leading the s_I/k_I proportion to remain approximately constant. From Eqs. (1) and (3), this means that such a factor would be unable to correlate with $p(aI)$.

How does generality affect the PLT?

Are generality and s_I related? Recall here that generality is the number of categories in the CSLB norms in which a property appears. As discussed in the section on Theory, if generality is to have an effect on conceptual variability, it should relate to s_I (this is our Hypothesis 3). For this analysis, we tested whether

s_I was able to predict the average generality of properties produced for each of the 638 concepts in the CSLB norms. To this end, we computed a regression equation with average generality as criterion, and s_I as predictor. Again, as done before, we added an s_I^2 quadratic term as predictor (see Table 5).

The complete model (see Table 5, model 1) is significant, and with a small R^2 of 0.061. However, neither the linear nor the quadratic coefficients are significant and the VIF is 59.13, indicating multicollinearity. Thus, we computed two separate regressions, one with s_I as predictor and the other with s_I^2 . The model with s_I (see Table 5, model 2) is significant as a whole, has a significant coefficient, and its R^2 is 0.059. The model with s_I^2 (see Table 5, model 3) is also significant as a whole, has a significant coefficient, and its R^2 is 0.06. As the two R^2 s are very similar, we selected the simpler linear model. Consequently, our results show that the more individuals can report about a concept, the more distinctive (less general) are those properties, though this is a small effect.

Are generality and k_I related? Given that generality is related to s_I , and k_I is affected by s_I , it is reasonable to expect that generality affects k_I . Moreover, since the relation between generality and s_I is negative, and between s_I and k_I is positive, we should expect a negative relation between generality and k_I . To test that, we regressed mean generality on k_I and k_I^2 . The complete model (see Table 5, model 4) and its two coefficients (linear and quadratic) are significant, and it shows an R^2 of 0.139. Because the complete model's VIF is lower than 10 and both coefficients are significant, we retained this as our preferred model. Thus, a higher k_I entails a lower mean this negative effect saturates as k_I increases (see Fig. 4).

Notably, our data show that k_I is able to predict much more of generality than s_I (more than twice the explained variance). This dovetails nicely with our previous results, and suggests that low generality properties (i.e., distinctive properties) tend to be low frequency properties (i.e., properties reported by relatively few subjects). If this is the case, a decrease in the generality of properties being produced by a group of subjects, must imply a relatively large increase in k_I (due to more distinctive properties having a greater chance of being low frequency, non-shared properties, and thus contributing more to an increase in k_I). We will offer a test of this explanation later.

Are generality and exposure related? We know from our previous results that s_I and exposure are related. Thus, it is likely that average generality and exposure are also related. Given that we already know that $\log_{10}(\text{exposure})$ exhibits multicollinearity with $\log_{10}(\text{exposure})^2$, we estimated two separate models for predicting average generality (see Table 5, models 5 and 6). The model for $\log_{10}(\text{exposure})$ and its coefficient are significant, and shows a small R^2 of 0.036. The corresponding quadratic model and its coefficient are also significant, with a similarly small R^2 of 0.036. As done before, for parsimony we retain the linear model for

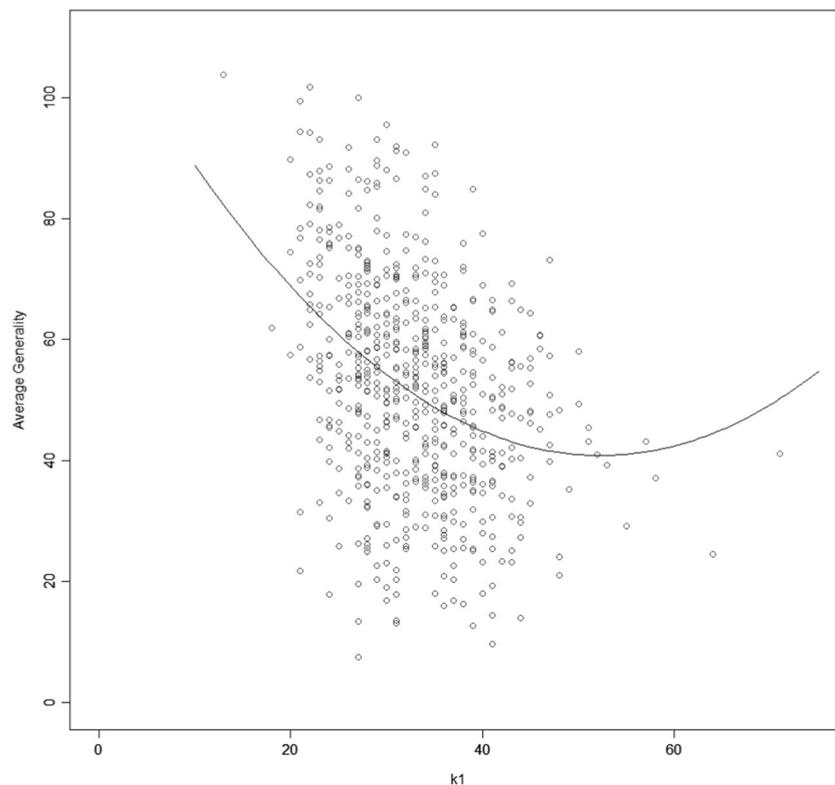


Fig. 4 Graph of model 4 in Table 5, showing the relation between k_I (total number of properties) and Average Generality for 638 concepts

interpretation. Note that though the relation is linear, the predictor is a logarithmic transformation, and thus the relation between exposure and average generality is non-linear, with a much more prominent decrease in average generality for increases in exposure when exposure is small than when it is large (i.e., the saturation effect already explained for exposure and k_I). Note also, that the overall effect is small (3.6%), which seems consistent with the null effect we found between exposure and $p(aI)$.

Are generality and $p(aI)$ related? Given that less average generality in the properties produced in a PLT associates with an increased k_I , and that k_I has a negative effect on $p(aI)$, we expected a positive relation of average generality and $p(aI)$ (i.e., concepts with a greater average generality of their properties, should show a higher $p(aI)$). This is our Hypothesis 3 again. To assess this, we built a model with average generality and squared average generality as predictors of $p(aI)$ (see Table 5, model 7). This model is significant, with an R^2 of 0.117. Both, the linear and the quadratic coefficients are also significant. However, given that the VIF is 21.87, bigger than 10, we suspected multicollinearity and built two separate models. The linear model and its coefficient are significant (see Table 5, model 8), with an R^2 of 0.089. The quadratic model and its coefficient are also significant (see Table 5, model 9), and it has an R^2 of 0.107. To decide if the linear component should be kept in the complete model, we compared models 7 and 9 regarding their R^2 values. The test

showed that the complete model explained significantly more variance than the one with only the quadratic term ($F(1, 635) = 6.832; p < .01$). Thus, as generality increases, there is a slight positively accelerated tendency to increase $p(aI)$.

Are property generality and property frequency related? Recall that when we analyzed how average generality and k_I are related, we reasoned that our pattern of results implied that properties that are more distinctive must tend to be low frequency properties (i.e., non-shared properties that, in the PLT, are produced by fewer subjects), in contrast to general properties, which must have a relatively higher frequency. This is a novel prediction, not anticipated in our theoretical analysis. To test this explanation more directly, we correlated generality across the 638 concepts in the CSLB norms, with property frequency. Just as happens with exposure, the distribution of generality at the property level is highly skewed, and thus we applied a base 10 logarithmic transformation to it (henceforth, $\log_{10}(\text{generality})$). Because $\log_{10}(\text{generality})$ is a property level variable, and property frequency is a property by concept level variable, we could not accumulate these variables across concepts to perform a single analysis. Consequently, we computed a separate linear correlation between these two variables for each of the 638 concepts in the CSLB norms, and then submitted these correlation values to a t -test with the null hypothesis that $\text{Rho} = 0$. This test showed that the mean correlation value (mean $r = .229$; S.D. = .158) was significantly different from zero ($t(637) = 36.6; p < .001$).

To provide further evidence, we counted the number of concepts that exhibited the same correlation sign. Of the 638 concepts, the correlation between $\log_{10}(\text{generality})$ and property frequency was positive in 588 concepts (more than 92%), and significant (at $\alpha = .05$) in 149 of them. Importantly, only 50 concepts showed a negative correlation (less than 8%), none of which was significant. Thus, for most concepts there is a tendency for general properties to be listed more frequently than distinctive properties. It is important to note here that because $\log_{10}(\text{generality})$ is a logarithmic transformation, the relation between the original generality measure and property frequency is really nonlinear, such that at low values of generality, a small increase in generality implies a relatively large increase in property frequency; whereas at higher values of generality, the same small increase in generality implies a relatively smaller increase in property frequency. Just as we have shown for earlier analyses, the effect of generality on property frequency saturates as generality increases.

Discussion

In this work we aimed at contributing to a better understanding of the PLT. This task is widely used in psychology, but little is known about its characteristics. An often overlooked aspect of the PLT, but that we used as our starting-point, is that people typically produce different sets of properties. Here, instead of dealing with this inter-subject variability in conceptual content by treating it as measurement error, we used it as the driving force in our analyses. To that effect, we used a formula that includes the average size of lists provided by subjects in a CPN study, the total number of properties listed at the group level, and the frequency distribution generated by those lists, to compute a measure of inter-subject variability ($p(aI)$). Aided by this mathematical machinery, and by combining two sets of publicly available norms (the CSLB and the SUBTLEX-UK norms), we investigated how variability relates to concept exposure, property generality, and property production frequency.

Rather than offering a summary of our results in a piecemeal fashion, in this section we integrate them to offer a characterization of the PLT and its relation to inter-subject variability. To favor readability, in this discussion we will mostly use the verbal descriptions of the relevant variables, rather than the corresponding letters or abbreviations used in data analysis. The basic element of this characterization is that lists of conceptual properties produced in the PLT are mixtures of properties of different levels of generality. Low generality properties are less likely to appear in individual conceptual properties lists than higher generality properties. This effect is stronger at the lower end of the generality scale. A simple explanation for this is that properties that describe more concepts

are more cognitively accessible (cf., Grondin, Lupker, & McRae, 2009).

The nature of these mixtures of properties relates to the length of individual property lists, and, consequently, to the total number of properties informed for a given concept by a group of individuals. Our results show that this latter quantity will be sensitive to any factor capable of changing the mixture of general and distinctive properties, even when the length of individual lists does not change. Because generality correlates with property frequency, less generality (i.e., more distinctiveness) also implies lower frequencies. These lower frequency properties increase the total count of properties much more than the relatively higher frequency properties. To clarify this, consider the extreme case where everyone produced the same properties. In that case, individual list length and total number of properties would covary linearly. In contrast, due to low frequency properties, there is an exponential relation between average list length and total number of properties. Consequently, the total number of properties for a concept contains information about the distribution of general and distinctive properties in lists. When this distribution skews towards low generality properties, the total number of properties listed for a concept grows, and so does intersubject variability in the PLT. In fact, we found a strong relation between total number of properties and $p(aI)$ (35% shared variance).

A related issue that our results address is what type of factor might affect intersubject variability. A result we did not predict is that though exposure to concepts in naturally occurring language affected mean list length and total number of properties, it did not affect intersubject variability. Apparently, more exposure to a concept in natural language increases the availability of its properties (i.e., more exposure increases mean list length and total number of properties). However, this seems to increase general and distinctive properties to a similar extent. In consequence, the relation between mean list length and total number of properties remains stable and so does $p(aI)$. This suggests that exposure as defined in the current work does not change the distribution of general and distinctive properties that are experienced (i.e., it has no effect on semantic structure). If we were to perform an experiment where people were put in high and low exposure conditions in a natural language environment, we would presumably see that with both conditions they would experience the same proportions of general and distinctive properties. Note that this does not mean that exposure in general should have this same effect. Other conditions, such as an increase in expertise, might increase exposure to distinctive more than to general properties, and could therefore change the distributions of general and distinctive properties, thus increasing intersubject variability (i.e., decreasing $p(aI)$). In general, our data suggest that to have an effect on inter-subject variability, a manipulation has to be able to affect the distribution of generality in individual subjects' property lists.

What we have learned from the current work has methodological consequences for researchers using the PLT. Importantly, we can now understand the effect that weeding out low frequency responses from CPNs may have. Recall that this procedure is reasonable if we assume that low frequency properties reflect measurement error. In contrast, in our approach, low frequency properties reflect bona fide conceptual variability that needs to be taken into account. What weeding out achieves is to decrease k_I much more than s_I , thus increasing $p(aI)$. We know now that this procedure has several other consequences. Because production frequency relates to generality in a non-linear fashion, excluding low frequency properties will change the true relation between these variables in the data, such that the relation may now appear linear or perhaps tend to disappear depending on how severe weeding out is. Furthermore, because generality and the total number of properties (k_I) are also related, and because distinctive/low frequency properties make an important contribution to the total number of properties (i.e., as per our discussion of the non-linear relation between s_I and k_I), weeding out will portrait data as being higher in generality and less variable than they really are. Because data in CPNs are used with the goal of testing theories and controlling variables, it is apparent that weeding out will change data structure in important ways and perhaps defeat these goals.

A similar analysis applies to property list elicitation procedures. These procedures vary across studies in such a way that will probably affect the number of properties being produced. However, researchers generally fail to explain why they choose those specific procedures, or how they may or may not influence their results. Examples of different procedures abound. A request that asks for typically or generally true properties (e.g., Wu & Barsalou, 2009) seems to ask for more general properties, and therefore might produce shorter lists with fewer low frequency properties and higher average generality. In contrast, a request to provide properties that could help someone to correctly guess the concept (e.g., Rechia & Jones, 2012), might produce property lists with a lower average generality and relatively more intersubject variability. Similarly, procedures also vary in whether properties are elicited in written or spoken format (e.g., respectively, Chang, Mitchel & Just, 2011; Perri, Zannino, Caltagirone, & Carlesimo, 2012), whether participants are asked or not for a minimum number of properties (e.g., Devereux, Tyler, Geertzen & Randall, 2014), and whether they are subject to time limits (i.e., how long participants are allowed to list properties). To the extent that these alternatives promote or limit the property production process, they should have effects on collected data in the lines of what we describe above, making comparisons between studies difficult.

Another issue that our results speak to is the way data from PLT studies are interpreted. Our newly gained understanding of the PLT offers different explanations for prior results.

Consider the following example. Perri et al. (2012) set out to use the PLT to test Alzheimer disease (AD) patients. An important comparison in which Perri et al. were interested was the frequency of high versus low generality properties. Their findings showed that AD patients produced significantly fewer low than high generality properties, which they attributed to a greater effect of the disease on distinctive than on high generality properties (i.e., a semantic deterioration). Furthermore, their results also showed that AD patients produced fewer overall properties than the control group. Perri et al. discussed the latter result as consequence of the former. In their analysis, AD patients produced fewer properties because of a greater impairment of distinctive over general properties. However, as our results suggest, an alternative explanation is that producing fewer properties leads to properties that are on average more general. Because AD is typically characterized by a progressive decay of working memory (see Jahn, 2013), and assuming that working memory is necessary to perform the PLT, Perri et al.'s results could really reflect greater working memory limitations, rather than a primary semantic deterioration.

Another example of how our analyses may help interpreting results is the following. The total number of properties produced by a group of subjects for a given concept (i.e., our k_I variable), is taken to be a measure of the concept's "semantic richness" (Pexman, Holyk, & MonFils, 2003; Pexman, Lupker, & Hino, 2002). This variable has been shown to have a facilitatory effect in different cognitive tasks (e.g., faster reaction times in lexical and semantic decision tasks for concepts with a greater total number of properties; Goh et al. 2016; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Pexman et al. 2013). Presumably, words with more properties receive stronger word activation. Consistent with the semantic richness effect, Grondin, Lupker, and McRae (Study 3, 2009) found faster reaction times for concepts with a high rather than a low number of properties (approximately 15 vs. 9). However, they also claim to have found that this effect was strongest when concepts' k_I properties differed only in their absolute number of shared properties while keeping the number of distinctive properties constant (shared+distinctive, 3+6=9 properties versus 9+6=15 properties), than when they differed only in their absolute number of distinctive properties while keeping the number of shared properties constant (shared+distinctive, 6+3=9 properties versus 6+9=15 properties) (see Grondin et al., 2009, Tables 5 and 6). Remarkably, the interaction was carried largely by concepts with a low total number of properties and half the number of shared relative to distinctive properties (i.e., the 3+6 condition), which showed slower reaction times than the three other conditions. Grondin et al. interpreted their results as a kind of moderation effect where the semantic richness effect is stronger when shared

Table 6 Regression models for the question of how generality affects the PLT

Model	Model <i>F</i> test	Linear coeff. <i>t</i> test	Quadratic coeff. <i>t</i> test
1. <i>Average Generality</i> = 61.201 + 1.039 (0.067) s_I - 0.331 (-0.312) s_I^2	<i>p</i> <.001	<i>p</i> =.821	<i>p</i> =.291
2. <i>Average Generality</i> = 78.157 - 3.764 (-0.243) s_I	<i>p</i> <.001	<i>p</i> <.001	
3. <i>Average Generality</i> = 64.929 - 0.260 (-0.246) s_I^2	<i>p</i> <.001		<i>p</i> <.001
4. <i>Average Generality</i> = 114.181 - 2.806 (-1.096) k_I + 0.027 (0.757) k_I^2	<i>p</i> <.001	<i>p</i> <.001	<i>p</i> <.001
5. <i>Average Generality</i> = 67.987 - 5.233 (-0.191) $\log_{10}(\text{exposure})$	<i>p</i> <.001	<i>p</i> <.001	
6. <i>Average Generality</i> = 60.271 - 0.847 (-0.189) $\log_{10}(\text{exposure})^2$	<i>p</i> <.001		<i>p</i> <.001
7. $p(aI) = 0.333 - 0.001 (-0.459) \text{Average Generality} + 2.085 \times 10^{-5} (0.775) \text{Average Generality}^2$	<i>p</i> <.001	<i>p</i> <.01	<i>p</i> <.001
8. $p(aI) = 0.280 + 0.001 (0.299) \text{Average Generality}$	<i>p</i> <.001	<i>p</i> <.001	
9. $p(aI) = 0.299 + 8.797 \times 10^{-6} (0.327) \text{Average Generality}^2$	<i>p</i> <.001		<i>p</i> <.001

Note: Models in bold type are those judged best according to our criteria (see text for details). For each model, values in parenthesis are standardized coefficients (β s)

properties rather than distinctive properties are varied. Our current work suggests the following caveat regarding this interpretation.

Throughout our analyses, we used property generality as a continuous dimension, in contrast to arbitrarily distinguishing between shared and distinctive properties. Because Grondin et al. dichotomized this dimension and separately varied the absolute number of shared or distinctive properties, average property generality may have been a confounding factor. In fact, assuming that individual properties are shared or are distinctive to approximately the same degree across conditions (which is a necessary condition for their design to be correct), then average generality will be a linear function of shared over total number of properties (the higher this proportion, the more general in average properties are). For the four conditions discussed above, this gives, respectively, $3/9=.33$; $9/15=.6$; $6/9=.67$; $6/15=.4$. Obviously, the estimated average property generality was not constant across conditions. Furthermore, though estimated average property generality appears to be constant across the number of properties manipulation (respectively, $(.33+.67)/2=.5$; $(.6+.4)/2=.5$), it might not have been so across the shared versus distinctive manipulation (respectively, $(.33+.6)/2=.47$; $(.67+.4)/2=.54$). This would explain why Grondin et al., somewhat puzzlingly, obtained faster reaction times when distinctive properties were manipulated (i.e., consistently with our results, higher generality produces a processing advantage). Interestingly, the condition with the lowest estimated average property generality (i.e., the $3+6=9$ condition) is the one that produced the slowest reaction times, which may suffice to explain the purported interaction.

Conclusion

Throughout the current work, we have analyzed the PLT and its relation to intersubject variability. Our results

have shown the PLT's complexities. Rather than reiterating them, we want to close by noting that intersubject variability seems to be an unavoidable feature of conceptual content. Instead of treating variability as problematic measurement error, we offer a way to describe it and to characterize factors that could systematically affect it. Our results suggest that a researcher using the PLT should carefully consider the effects that limiting the number of properties will have on their data (either because of elicitation procedures or because of weeding out); that they should also consider that variables of interest (such as property generality and list length) are not independent but relate in complex and often non-linear ways; and that they should not arbitrarily dichotomize variables such as those discussed in the current work. By pursuing our strategy, we hope not only to have gained a greater understanding of the PLT, but also to contribute to a better comprehension of CPNs and the nature of the uncovered conceptual structures.

Acknowledgements We thank Gorka Navarrete and two anonymous reviewers for their useful comments on earlier versions of this manuscript. This research was carried out with funds provided by grant 1150074 from the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) of the Chilean government to the first and second authors.

Appendix A: Example of calculation of $p(aI)$ when properties are non-equiprobable

Note that in order to use Eq. (1) we must previously calculate the probabilities of obtaining each of the different possible samples drawn from the concept (i.e. the p_i and p_j). Although this is simple to do, that calculation involves many summations and multiplications, making it cumbersome to do by hand. To show it here for the example corresponding to the

concept depicted in Fig. 1 for the non-equiprobable case, we will only calculate the probabilities of some of the samples. Recall that the possible samples of size two for the concept $\{a, b, c\}$ are (ab) , (ac) and (bc) . Thus:

$$p(ab) = p(ab) + p(ba) = p(a) p(b/a) + p(b) p(a/b) \tag{A1}$$

$$p(ab) = p(a) \frac{p(b)}{p(b) + p(c)} + p(b) \frac{p(a)}{p(a) + p(c)} \tag{A2}$$

In (A1) and (A2), remember that the order in which properties a and b are sampled is irrelevant, thus the two permutations of properties (ab) and (ba) are equivalent. Thus, to calculate the probability of combination (ab) , we calculate the probability of permutation (ab) $(p(ab))$ and of permutation (ba) $(p(ba))$, and given that they are mutually exclusive, we can add them. Also, assume the following frequencies for each

property: $f(a) = 2, f(b) = 3, f(c) = 5$, and hence, $p(a) = 2/(2 + 3 + 5) = 2/10 = 1/5$ and for $p(b) = 3/10$ and $p(c) = 5/10 = 1/2$. In the case of combination $p(ab)$, and putting the corresponding values of $p(a)$, $p(b)$ and $p(c)$ in (A2), we obtain:

$$p(ab) = \frac{1}{5} \frac{\frac{3}{10}}{\frac{3}{10} + \frac{1}{2}} + \frac{3}{10} \frac{\frac{1}{5}}{\frac{1}{5} + \frac{1}{2}} = \frac{45}{280} = \frac{9}{56} \tag{A3}$$

Using expressions similar to (A2), the value for $p(ac)$ is $91/280 = 13/40$ and for $p(bc) = 144/280 = 18/35$. Recalling that the number of common properties for all independent samples drawn from the concept is $\{2, 1, 1, 1, 2, 1, 1, 1, 2\}$ and applying Eq. (1), we obtain:

$$p(a1) = \frac{1}{2} \left(\left(\frac{9}{56} \right)^2 \cdot 2 + \frac{9}{56} \frac{13}{40} \cdot 1 + \frac{9}{56} \frac{18}{35} \cdot 1 + \dots + \left(\frac{18}{35} \right)^2 \cdot 2 \right) = 0.69797 \tag{A4}$$

Note that $p(a1)$ for non-equiprobable properties (0.69797) is bigger than the one for equiprobable ones (0.66667, see computation in section Theory) as expected.

Appendix B: Demonstration that $p(a1)$ for equiprobable properties is a lower bound on $p(a1)$ for non-equiprobable properties

Let us consider one set:

$$C1 = \{a_1, a_2, \dots, a_{k_1}\} \tag{B1}$$

In this case, the cardinalities of $C1$ is k_1 . Let's also denote the size of a sample extracted from $C1$ as s_1 . Then, the number of such possible independent samples will be: n_1 : total number of possible samples of size s_1 obtained from $C1 = n_1 = \binom{k_1}{s_1}$.

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

$$M_1 = \{S \subseteq C1 : \#S = s_1\} = \{S_1^1, S_2^1, \dots, S_{n_1}^1\} \tag{B2}$$

So, we have the following probability:

$$p(a1) = \frac{1}{s_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \#(S_i^1 \cap S_j^1) p_i p_j, \sum_{i=1}^{n_1} p_i = 1, p_i \geq 0, \sum_{j=1}^{n_1} p_j = 1, p_j \geq 0. \tag{B3}$$

where p_i is the probability of obtaining sample S_i^1 and p_j is the probability of obtaining sample S_j^1 .

The probability in Eq. (B3) may be seen as a quadratic form by defining the following matrices:

$$A = (a_{ij}) \text{ where } a_{ij} = \#(S_i^1 \cap S_j^1) \tag{B4}$$

$$p = (p_1, p_2, \dots, p_{n_1}) \tag{B5}$$

$$p(a1) = \frac{1}{s_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} p_i a_{ij} p_j = \frac{1}{s_1} p^t A p \tag{B6}$$

To calculate a lower bound on $p(a1)$, we must minimize expression (B6). This minimization problem expressed in matrix notation is:

$$\frac{2}{s_1} \min \frac{1}{2} p^t A p, \text{ s.t. } p^t 1 = 1, p \geq 0. \tag{B7}$$

The properties of A are:

1. Symmetric, i.e., $A^t = A$.
2. $a_{ii} = s_j \forall i$.
3. $s_j = a_{ii} > a_{ij} \geq 0 \forall i, j$
4. $AI = \alpha I$ or $I^t A = \alpha I^t$, where α is the sum of a row $S(L)$. $S(L)$ is a matrix that contains the sum of r common properties between independent samples S_i^l and S_j^l multiplied by the number of samples that contain those r common properties. $S(L)$ is important for other demonstrations but here we can just take the sum of its rows as a constant α .

The first property indicates that all the eigen-values 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 α is an eigen-value of A and the vector I is its respective eigen-vector.

Before solving (B7), we must note that $\frac{2}{s_1}$ is a constant, and thus 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 $D = \{p \in \mathfrak{R}^{n_1} : p^t 1 = 1, p \geq 0\}$, it is easy to see that the problem has 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 $p = 0$. Weierstrass theorem guarantees the existence of the minimizer in D .

To solve problem (B7), let us consider the lagrangian:

$$L(p, \lambda, \mu) = \frac{1}{2}p^t A p + \lambda(p^t 1 - 1) - \mu^t p. \tag{B8}$$

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:

$$\begin{aligned} \nabla_p L(p^*, \lambda^*, \mu^*) &= 0, \\ (p^t 1 - 1) &= 0 \\ -p &\leq 0, \\ \mu &\geq 0, \\ \mu_i p_i &= 0, \quad i = 1, \dots, n_1 \end{aligned} \tag{B9}$$

Remembering matrix derivatives, we have:

$$\nabla_p L(p, \lambda, \mu) = \begin{pmatrix} \frac{\partial L}{\partial p_1} \\ \vdots \\ \frac{\partial L}{\partial p_{n_1}} \end{pmatrix} = A p + \lambda 1 - \mu \tag{B10}$$

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,

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

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

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

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

$$p^* = \frac{1}{n_1} 1 = \begin{pmatrix} 1/n_1 \\ \vdots \\ 1/n_1 \end{pmatrix} \tag{B11}$$

If $\det(A) = 0$, there exist infinite solutions and one of them is the one already found in (B11). The rest of the solutions are of the form:

$$q^* = p^* + b v, \text{ where } v \in \text{Ker}(A) = \{v \in \mathfrak{R}^{n_1} : A v = 0\} \tag{B12}$$

The constant $b \in \mathfrak{R}$, is any that gives $q^* \geq 0$. The vector v , is a linear combination of the eigen-vectors of matrix A associated with the zero eigen-values. 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 (B11) means that all samples drawn from CI have the same probability ($1/n_1$) and, thus all the properties of CI are equiprobable. Hence, a lower bound on $p(a1)$ occurs when all properties of CI are equiprobable.

References

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, 39(2), 216–233.

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: 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). London: Lawrence Erlbaum Associates.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990.

- Canessa, E., & Chaigneau, S. E. (2016). When are concepts comparable across minds? *Quality & Quantity*, *50*(3), 1367–1384. doi:10.1007/s11135-015-0210-4
- Chaigneau, S. E., Canessa, E., & Gaete, J. (2012). Conceptual agreement theory. *New Ideas in Psychology*, *30*(2), 179–189.
- Chang, K. K., Mitchell, T., & Just, M. A. (2011). Quantitative modeling of the neural representation of objects: How semantic feature norms can account for fMRI activation. *NeuroImage*, *56*, 716–727.
- Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and discontent* (pp. 206–261). New York: The Free Press.
- 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*, 163–201.
- 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. doi:10.3758/s13428-013-0420-4
- Frege, G. (1893). On sense and reference. In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). Oxford: Blackwell.
- Gabora, L., Rosch, E., & Aerts, D. (2008). Toward an ecological theory of concepts. *Ecological Psychology*, *20*(1), 84–116.
- Glock, H. J. (2009). Concepts: Where subjectivism goes wrong. *Philosophy*, *84*(1), 5–29.
- Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M. R., & Tan, L.-C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology*, *7*, 976.
- Griffiths, T. L., Sanborn, A. N., Canini, K. R., Navarro, D. J., & Tenenbaum, J. B. (2011). Nonparametric Bayesian models of categorization. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 173–198). Cambridge: Cambridge University Press.
- Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language*, *60*(1), 1–19. doi:10.1016/j.jml.2008.09.001
- Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441–461.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1992). *Multivariate data analysis* (3rd ed.). New York: Macmillan Publishing Company.
- Jahn, H. (2013). Memory loss in Alzheimer's disease. *Dialogues in Clinical Neuroscience*, *15*(4), 445–454.
- Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariable methods* (2nd ed.). Boston: PWS-Kent Publishing Company.
- McNorgan, C., Kotack, R. A., Meehan, D. C., & McRae, K. (2007). Feature–feature causal relations and statistical co-occurrences in object concepts. *Memory & Cognition*, *35*(3), 418–431.
- 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.
- McRae, K., Cree, G. S., Westmacott, R., & de Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology*, *53*, 360–373.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge: MIT Press.
- Perri, R., Zannino, G., Caltagirone, C., & Carlesimo, G. A. (2012). Alzheimer's disease and semantic deficits: A feature-listing study. *Neuropsychology*, *26*(5), 652–663. doi:10.1037/a0029302
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin Review*, *15*, 161–167. doi:10.3758/PBR.15.1.161
- Pexman, P. M., Holyk, G. G., & MonFils, M. H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, *31*, 842–855.
- Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feed-back semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, *9*, 542–549.
- Pexman, P. M., Siakaluk, P. D., & Yap, M. J. (2013). Introduction to the research topic meaning in mind: Semantic richness effects in language processing. *Frontiers in Human Neuroscience*, *7*, 723. doi:10.3389/fnhum.2013.00723
- Pothos, E. M., & Wills, A. J. (Eds.). (2011). *Formal approaches in categorization*. Cambridge: Cambridge University Press.
- Prinz, J. J. (2005). The return of concept empiricism. In H. Cohen & C. Lefebvre (Eds.), *Categorization and cognitive science* (pp. 679–694). Amsterdam: Elsevier.
- Recchia, G., & Jones, M. N. (2012). The semantic richness of abstract concepts. *Frontiers in Human Neuroscience*, *6*, 315.
- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive Development and the acquisition of Language* (pp. 111–144). New York: Academic Press.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.
- Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502.
- Russell, B. (1997). *The problems of philosophy*. Oxford: 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*(1), 83–119.
- Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Brain and Behavioral Sciences*, *21*, 1–17.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.
- Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. doi:10.1080/17470218.2013.850521
- Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422–488.
- Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*, 173–189.