ARTICLE

A mathematical model of semantic access in lexical and semantic decisions

Sergio E. Chaigneau[1] [iD] , Nicolás Marchant[1] [iD] , Enrique Canessa[2] and Nerea Aldunate[3]

[1]Center for Social and Cognitive Neuroscience, School of Psychology, Universidad Adolfo Ibáñez, Av. Presidente Errázuriz 3328, Las Condes, Santiago, Chile; [2]Faculty of Engineering and Science, Universidad Adolfo Ibáñez, Viña del Mar, Chile and [3]Centro de Investigación en Complejidad Social, Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile
Corresponding author: Nicolás Marchant; Email: nicolas.marchant@edu.uai.cl

Abstract

In this work, we use a mathematical model of the property listing task dynamics and test its ability to predict processing time in semantic and lexical decision tasks. The study aims at exploring the temporal dynamics of semantic access in these tasks and showing that the mathematical model captures essential aspects of semantic access, beyond the original task for which it was developed. In two studies using the semantic and lexical decision tasks, we used the mathematical model's coefficients to predict reaction times. Results showed that the model was able to predict processing time in both tasks, accounting for an independent portion of the total variance, relative to variance predicted by traditional psycholinguistic variables (i.e., frequency, familiarity, concreteness imageability). Overall, this study provides evidence of the mathematical model's validity and generality, and offers insights regarding the characterization of concrete and abstract words.

Keywords: lexical decision task; mathematical modeling; property listing; semantic access; semantic decision task

1. Introduction

The current work uses a mathematical model of the property listing task (PLT, Canessa & Chaigneau, 2020; Canessa et al., 2021) to predict processing time in semantic and lexical decision tasks (SDT and LDT). In the PLT, people freely list properties that are typically true of a given concept (Chaigneau et al., 2018; Hough & Ferraris, 2010; Perri et al., 2012; Walker & Hennig, 2004; Wu & Barsalou, 2009). Following an approach championed by Simon (1964), the model is formulated as a set of differential equations that characterize the PLT's listing dynamics (Canessa & Chaigneau, 2020). Note that, though, we are interested in the underlying cognitive mechanisms, our focus is on the variables' functional relations. As discussed in the

next section, several putative mechanisms could account for the empirical temporal trends in PLT listing, but our goal is not to resolve between those different mechanisms. Thus, our theoretical claim is that the model summarizes various processing mechanisms that could, in principle, account for the same trends in the data.

To provide evidence of the model's generality, we now extend its applicability beyond the original PLT task for which it was developed. The coefficients computed when applying the PLT model for each concept used in the present study were estimated from a PLT task in an independent study (Canessa et al., 2023). By using a multiple regression approach, we now compare those coefficients' ability to predict reaction time (RT) in an SDT and LDT to other generally used psycholinguistic variables that have been shown to be predictors in those tasks (*concreteness*, *imageability*, *frequency familiarity*) (Barber et al., 2013; Khanna & Cortese, 2021; Muraki et al., 2020; Yap et al., 2015).

Because the concrete/abstract distinction continues to generate controversy, we were particularly interested in differences in processing time attributable to different levels in this continuum (Canessa et al., 2021). In particular, consider that abstract concepts' referents are not spatially and physically bounded, so they pose the question of whether and how the cognitive system processes them differently from concrete concepts. This is particularly challenging for embodied approaches to cognition (see Dove, 2022; for discussions regarding varieties of abstract concepts, see also Barsalou, 2003; Borghi et al., 2018; Borghi et al., 2022; Langland-Hassan & Davis, 2023).

2. The mathematical model of the PLT

2.1. The PLT

The PLT and resulting semantic property norms (SPNs) are widely used in cognitive research to investigate conceptual content and the organization of semantic memory (SM; Canessa & Chaigneau, 2020; Chaigneau et al., 2018). Researchers apply the PLT to elicit semantic properties that are typically associated with a given concept. These properties serve as carefully controlled stimuli for experiments and can predict performance in linguistic and nonlinguistic tasks.

Following the PLT, SPNs are created from the elicited properties, providing a means of characterizing a semantic space. SPNs are matrices containing different concepts and their corresponding property frequency distributions (Canessa et al., 2020; Devereux et al., 2014; Kremer & Baroni, 2011; Lenci et al., 2013; McRae et al., 2005; Montefinese et al., 2013; Vivas et al., 2017).

The PLT and associated SPNs have found applications in both basic cognitive research and applied or field studies (Hough & Ferraris, 2010; Perri et al., 2012; Walker & Hennig, 2004; Wu & Barsalou, 2009). These norms serve as tools for testing theories, generating carefully controlled experimental stimuli and evaluating the extraction of conceptual knowledge from corpora in computational linguistics (Baroni & Lenci, 2008; Cree & McRae, 2003; Devereux et al., 2009; Fagarasan et al., 2015; Taylor et al., 2011; Vigliocco et al., 2004; Wu & Barsalou, 2009).

2.2. PLT mathematical model

The PLT mathematical model describes the property listing process as it unfolds using the listing order of properties (Canessa & Chaigneau, 2020). It has been

Table 1. Variables employed to describe the PLT listing process dynamics used in the present analyses

Variable's symbol	Definition
s	Average cumulative number of properties produced through the listing process by all participants in a PLT for a given concept
k	Number of unique properties (i.e., without considering repetitions) produced by all participants through the listing process in a PLT for a given concept
t	Time at which properties are produced by each participant in a PLT for a given concept (from cue onset)

validated with data collected in three different SPNs, spanning three countries (Italy, Argentina and Chile), two different languages (Italian and two Spanish dialects), and with 958 participants producing properties for 497 abstract and concrete concepts and verbs (Canessa & Chaigneau, 2020). The following description of the model summarizes those parts of the PLT model which were used in this paper (Canessa & Chaigneau, 2020; Canessa et al., 2023). The PLT model describes the dynamics of the listing process using six equations that relate four variables. Here, we use only two of those six equations and the variables defined in Table 1. We must note that the model's equations are somewhat redundant, in the sense that some equations are derived from the previous ones, allowing to relate all our variables in a single model of interrelated variables. The two equations we chose to keep for the analyses we report are foundational of the model, in the sense that other equations that we could have included are derived from them. Additionally, those two equations are the ones that receive stronger support from data (in terms of $R^2 > 0.96$) in the original study where the corresponding coefficients were computed (Canessa et al., 2023).

First, the model hypothesizes that as the PLT unfolds, participants' property listing rates decrease. This decreasing rate may be due to different factors, such as decreasing property availability, proactive interference (PI) and control processes such as monitoring for intrusions and repetitions (probably not an exhaustive list). We are not committed to any of the following factors, as our model attempts only to describe the time course of listing.

In listing tasks, the frequency of each property is a powerful factor in recalling from SM. In general, given a certain stimulus context, properties that are more frequent in that context are more available for processing (Balota & Spieler, 1999; Maki, 2007). Differences in availability could lead to changes in listing rates because highly available properties, which tend to be produced at the beginning of the list, may result in higher production rates than less available properties, which tend to be produced at the end of the list.

Furthermore, listing properties requires managing access to long-term memory (LTM) under PI conditions. In listing tasks, clusters of properties occur due to facilitation among items in memory that are in the same semantic field (Abwender et al., 2001; Hills et al., 2015; Rosen & Engle, 1997; Troyer, 2000). However, as a semantic field is sampled, properties in short-term memory proactively interfere with other properties in the same field (Conway & Engle, 1994; Kane & Engle, 2000; Wickens et al., 1963; 1976; 1981). PI is the interference caused by content already recalled, in the retrieval of content not yet accessed from LTM. Consequently, in listing tasks, memory performance decreases on successive trials (Keppel & Underwood, 1962). This phenomenon is called PI buildup (Fox et al., 2020; Kliegl & Bäuml,

2021). It is known that this process is influenced by the semantic field from which content is retrieved, because changes in semantic field increase performance (release from PI; Wickens et al., 1976). The PI mechanism also suggests that the rate of listing in the PLT will decrease with listing time.

Finally, when listing properties, subjects need to monitor for intrusions and repetitions (Rosen & Engle, 1997). As participants continue listing, intrusions and repetitions become increasingly likely, such that attentional resources are taxed and listing properties becomes harder. This factor also implies that the rate of listing will decrease with time.

All these issues imply that the property listing rate (i.e., how the average list size s varies with time) should be inversely proportional to time, which can be expressed by the differential equation $ds/dt = 1 / t$. Solving that equation gives eq. (1):

$$s = a_0 + a_1 \, ln|t| \qquad (1)$$

For empirical evidence for this functional form, see Canessa and Chaigneau (2020) and Canessa et al. (2023)). The constants a_0 and a_1 come from solving the differential equation and can be calculated from data through Ordinary Least Squares (OLS) regression.

Another important relation in the model is between s (average cumulative number of properties) and k (number of different or unique properties produced through the listing process). As discussed above, properties that are more available for listing should be produced earlier in subjects' lists and be of higher frequency. In contrast, less available properties should be produced later and be more idiosyncratic. Thus, given that k represents the total number of unique properties during the listing process, the rate of increase in k should be directly proportional to s. In previous research, k has been shown to increase with s at different rates depending on concept type (Canessa & Chaigneau, 2020; Canessa et al., 2023). Concepts characterized by highly available and shared properties, show a slow increase in k in the initial listing phase, and a faster rate toward the end of lists, where the less shared, more idiosyncratic properties emerge (i.e., a nonlinear increase). On the other hand, concepts characterized by low availability properties show an approximately constant rate of low-frequency properties, such that the rate of increase in k relative to s is approximately constant (i.e., a linear increase).

Thus, to summarize both alternatives, we can say that the rate of increase of k with respect to s is constant and/or directly proportional to s. This can be expressed by the differential equation $dk/ds = e_0 + e_1 s$. The solution to that equation is the following expression:

$$k = e_0 + e_1 s + e_2 s^2 \qquad (2)$$

The constants e_0, e_1 and e_2 come from solving the differential equation and can be calculated from data through regression methods such as OLS. Consequently, eq. (2) allows analyzing whether the linear and/or quadratic terms are more prevalent for each concept, according to the abovementioned arguments.

Note that eq. (2) is related to "semantic richness." In general, richer concepts are easier to process because they have a denser semantic neighborhood (i.e., a set of closely knit and interrelated concepts), more semantically related terms (i.e., the

sheer number of associates) or are characterized by more features (Mirman & Magnuson, 2008; Yap et al., 2011). To illustrate, concrete concepts are generally thought to be relatively rich, with more features, more associates and a set of interrelated concepts (e.g., all things that people know about dogs). In contrast, many abstract concepts are thought to be relatively poor, with only a few associates, and depending strongly on the specific contexts in which they occur (e.g., the concept *beauty* depends on the specific things about which *beauty* is being predicated).

In the model, coefficient e_1 indicates that the increase of k relative to s is approximately linear, due to more idiosyncratic and low-frequency properties being uniformly distributed during subjects' listing. Given that a richer concept should evoke relatively few unique or idiosyncratic properties (i.e., those not strongly associated to the cue word), this will cause k to grow at a smaller rate relative to s, and the e_1 slope representing the relation of k to s will be shallower. Relatedly, if unique and low-frequency properties tend to occur at the end of subjects' lists, then the e_2 coefficient will be significantly different from zero, showing that unique properties increase k at a higher rate at the end of listing. Here, a relatively larger slope indicates more semantic richness (i.e., people only diverge in their lists when many accessible properties have been listed). The combined effect of both coefficients allows eq. (2) to simultaneously represent the density of the unique properties during listing and their distribution throughout listing, allowing the characterization of rich concepts as low density of unique and low frequency properties and/or that accumulate unique and low frequency properties toward the end of listing.

2.3. Computing the coefficients

To predict RTs in the SDT and LDT, we use the coefficients computed in Canessa et al. (2023)). To that effect, that previous work conducted a SPN study involving 120 concepts (60 concrete and 60 abstract concepts). A total of 221 participants (all Chilean Spanish native speakers) performed the PLT and the study measured RTs from the moment the cue word was presented to the moment in which each semantic property was produced. Note that this is an independent sample from the one that was used for SDT and LDT studies reported here. Therefore, the previous data were time series produced by each subject for each concept, reflecting their listing processes (for details, see Canessa et al., 2023).

To estimate the coefficients, the same basic procedure used in Canessa and Chaigneau (2020) was followed, which involves computing the regression equations (OLS) for each concept, using the corresponding functional forms in eqs. (1) and (2) and time series data (Canessa et al., 2023). The individual coefficients for each concept used in the LDT and SDT tasks reported here can be found at the Open Science Foundation (OSF, https://osf.io/zsn4c/).

3. SDT and LDT tasks

In the SDT, people must decide whether a given word shown on a screen is concrete or abstract. It is widely assumed that in this task, semantic variables account for an important amount of RT's variance in the SDT (Pexman et al., 2017). In the LDT, subjects must decide if a string of letters is a word or a pseudoword (i.e., a letter string that can be pronounced but has no meaning). Though semantic variables play a role

in the LDT (Pexman, 2012; Pexman et al., 2002), less variance is explained by those semantic variables (Pexman, 2012), presumably because sub-lexical variables play significant roles in aiding word recognition in the LDT (Plaut, 1997).

In our study, semantic variables were the three coefficients previously computed (a_1, e_1 and e_2), and the concreteness and imageability ratings for each concept. Sub-lexical variables were text frequency and ratings of familiarity. All these variables usually appear as predictors in the SDT and LDT tasks (Barber et al., 2013). Because our materials are in Spanish, the variables of concreteness/abstractness, imageability, familiarity and frequency of use were obtained from the EsPal database (Duchon et al., 2013).

To obtain a measure of concreteness/abstractness, researchers often rely on subjective ratings collected in normative studies (Holcomb et al., 1999; Villani et al., 2019). Instructions typically used for participants to generate concreteness ratings are based on Spreen and Schulz (1966), and define concrete concepts as nouns that refer to persons, places and things that can be perceived through the senses. On the other hand, abstract concepts, which cannot be directly experienced through the senses, should receive low concreteness ratings.

A second type of rating typically collected in databases is imageability, which measures the ease with which a word can evoke mental images or sensory representations. Words with high imageability are often concrete and have strong sensory associations, while abstract words may have lower imageability (Cortese & Fugett, 2004). As expected, given the instructions described above, concreteness and imageability tend to be highly correlated (Kousta et al., 2011). The reader will find that our data show this same correlational pattern.

Familiarity is a subjective rating that assesses the degree of exposure to a particular word that individuals have. It reflects how common and recognizable a word is within a given population. Familiar words have been found to be easier to process than less familiar words, though this seems not to be a linear relation (Bridger et al., 2014).

Word frequency refers to the frequency with which a word occurs in a given language or corpus. It is often measured by counting the number of times a word appears in written or spoken texts. High-frequency words are typically easier to process and access from memory (Neville et al., 2019). Because frequency distributions are highly skewed and non-normal, a typical procedure is to take the logarithm of frequency for statistical analyses. Frequency (or the log of frequency) has been shown to predict access time (Segalowitz & Lane, 2000).

4. Predictions for the current work

In summary, we claim that the PLT mathematical model describes access to SM, and that eqs. (1) and (2) reflect different but related aspects of that access. A higher a_1 coefficient directly reflects easier access in the PLT and presumably a general easier access to a concept's semantic information. Coefficients e_1 and e_2 reflect semantic richness, which implies differences in ease of processing. Relatively smaller e_1 coefficient and/or relatively larger e_2 coefficient reflect a richer concept. Importantly, if the model describes ease of semantic access, then it should generalize beyond the PLT task for which it was originally developed.

Following our review of the SDT and LDT tasks in the previous section, we predicted that the semantic variables would account for more variance in SDT than in the LDT. If the coefficients reflect ease of access to semantic content, then they should

behave similarly to the other semantic variables (i.e., concreteness and imageability ratings). In contrast, the sub-lexical variables (i.e., frequency and familiarity ratings) should dominate in the LDT, though not exclusively of semantic variables.

Finally, we also aimed at testing whether the coefficients in the PLT model made an independent contribution to the prediction of RTs in our tasks, relative to the contribution made by their psycholinguistic counterparts. In particular, we were interested in the model's coefficients' relationship to concreteness and imageability ratings, because they have become the *de facto* operational definition of concreteness (Löhr, 2022).

5. Preliminary correlation analysis

As a preliminary analysis, we computed Pearson correlations between our variables. To linearize word frequency, we used Log frequency in all our analyses (Log10 of the word token frequency in the EsPal database). Table 2 shows those correlations. Figure 1 shows scatterplots so the reader can get a sense of the data structure. As shown in Table 2 and Figure 1, sub-lexical variables correlate with each other and semantic variables also correlate with each other. Importantly, the PLT model coefficients cluster with the semantic variables and not with the sub-lexical ones, consistent with our hypothesis that the model's coefficients reflect semantic access. Note also the positive correlations shown by the a_1 coefficient (higher coefficients indicate higher concreteness ratings), but the negative correlations for the e_1 coefficient (i.e., a larger e_1 coefficient indicates lower concreteness ratings), consistent with our analysis of Eq. (2).

6. Methods

6.1. LDT study

In this study, participants had to decide whether a letter string presented on a computer screen was a word or a pseudoword. We used concrete and abstract words to check for evidence of semantic processing but expected to find that sub-lexical variables explained most of the variance. Previous evidence showing small but

Table 2. Correlation matrix between sub-lexical and semantic variables with PLT model coefficients

	Log frequency	Familiarity	Imageability	Concreteness	a_1 coefficient	e_1 coefficient
Log frequency	—					
Familiarity	**0.20**	—				
	.03					
Imageability	−.06	**.26**	—			
	.51	*.003*				
Concreteness	−.15	.16	**.85**	—		
	.10	*.09*	*<.001*			
a_1 coefficient	−.08	.05	**.42**	**.37**	—	
	.40	*.55*	*<.001*	*<.001*		
e_1 coefficient	.17	−.03	**−.31**	**−.32**	**−.18**	—
	.07	*.74*	*<.001*	*<.001*	*.05*	
e_2 coefficient	.00	−.07	.09	.08	.02	**−.80**
	.97	*.46*	*.36*	*.36*	*.85*	*<.001*

Note: Pearson's coefficients and *p*-values (in italics). Significant coefficients in bold.
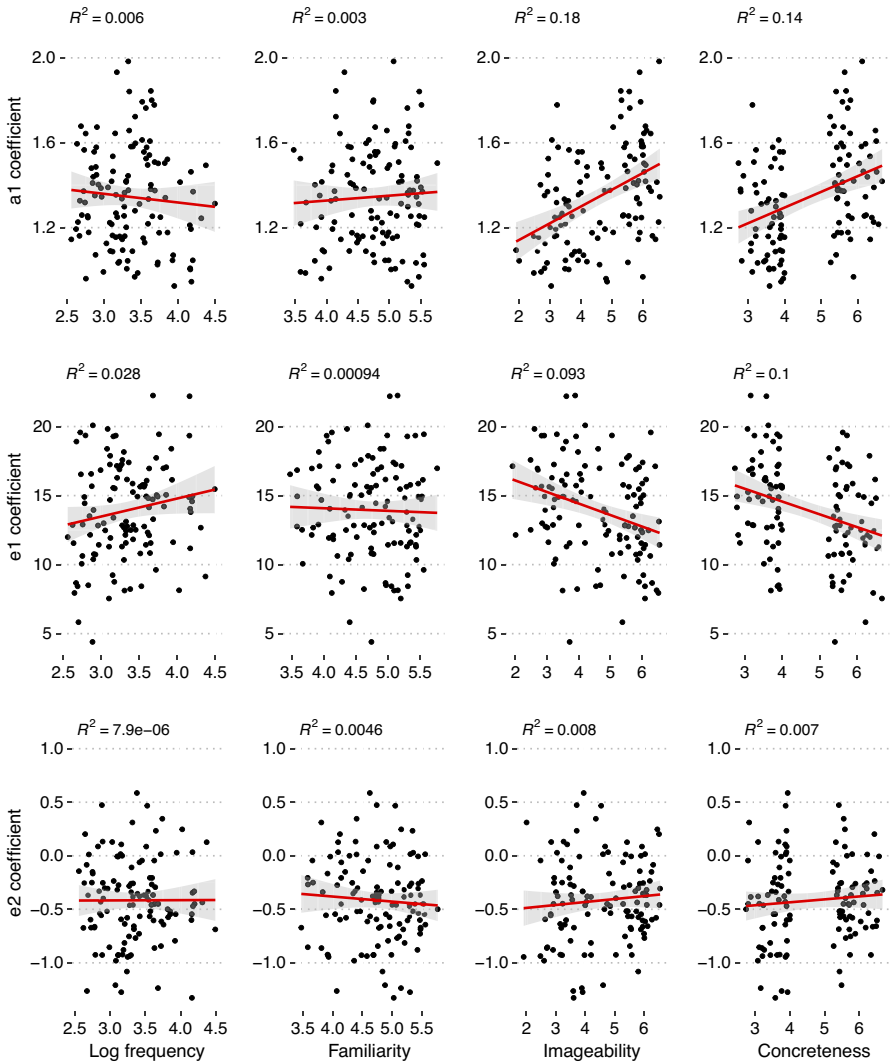
Figure 1. Scatterplots between sub-lexical and semantic variables with property listing task (PLT) model's coefficients. *Note.* Best regression line in red and standard errors filled-in-gray shape.

significant semantic effects (Pexman, 2012; Pexman et al., 2002), suggested that the coefficients might also be predictive in this task. Consequently, we predicted that a model with semantic and sub-lexical variables would be a good model, and therefore adding the coefficients should improve the regression model in terms of predicted variance (i.e., coefficients contribute independently).

6.1.1. Participants
Then, 101 undergraduate Spanish speaker students of Universidad Adolfo Ibáñez voluntarily agreed to participate in the study (M_{age} = 21.58, SD_{age} = 2.52; 23 females)

Table 3. Mean values of lexical and sub-lexical variables for concrete and abstract concepts

	Concrete	Abstract	*t*-Value	*p*-Value
Frequency of use	12.15	13.33	−.39	.70
Log frequency	3.30	3.40	−1.33	.19
Familiarity	4.83	4.66	1.66	.10
Imageability	5.66	3.50	17.74	<.001
Concreteness	5.80	3.55	32.05	<.001

and received a small no-monetary academic incentive. All participants gave their informed consent to participate, and the study was approved by the Ethics Committee of Universidad Adolfo Ibáñez.

6.1.2. Materials and procedure

Materials consisted of 120 concepts (60 concrete and 60 abstract) used in a previous study to validate the PLT mathematical model with RT (Canessa et al., 2023). Those concepts were selected so that they were not significantly different in frequency of use, familiarity and number of syllables. Also, the concrete and abstract concepts were selected so that they had a substantial difference in their concreteness ratings between those two types of concepts, so as to allow a good analysis of the difference between abstract and concrete concepts as it is shown in Table 3.

We wanted that our results be as generalizable as possible, and not restricted to a single or to a few types of abstract concepts. Thus, we selected a wide range of abstract concepts including, abstract actions (e.g., errand); behavior descriptions (e.g., clumsiness); personality traits (e.g., shyness); emotions (e.g., pity); physical (e.g., inertia) or aesthetic (e.g., ugliness).

Additionally, 120 pseudowords were derived from their corresponding word. Those pseudowords were produced from the 120 concepts controlling by relevant lexical variables (number of letters, number of syllables and sub-syllabic structure), with the multilingual pseudoword generator Wuggy (Keuleers & Brysbaert, 2010), and using methods described in Perea et al. (2016) and in Keuleers and Brysbaert (2010). While the pseudowords were derived from both concrete and abstract words, it is important to note that all of them are considered to belong to a single category (i.e., pseudowords).

The LDT was programmed using PsychoPy (Peirce et al., 2019). All words and pseudowords were sequentially presented on a computer screen in random order, controlling that trials in the same condition (i.e., concrete words, abstract words and pseudowords) were not presented more than three times in a row. Each stimulus was preceded by a fixation cross on the screen for 250 ms. After that, each stimulus was displayed for 500 ms. The total duration of the trial was 2000 ms. Participants were instructed to respond as fast as possible by pressing two different keys on a computer keyboard. To control the effect of lateralized motor responses, we implemented a counterbalancing strategy for the response keys following the completion of the first half of trials (i.e., the key to indicate that a presented word/pseudoword was a real word and the key to denote that the presented word/pseudoword was a pseudoword were switched). At the beginning of each experimental block, a training session was conducted using the same procedure (including the switching of the response keys)

employing different stimuli from those used in the actual experiment. Each participant took approximately 20 minutes to complete the task.

6.1.3. Results

To assess the performance on the LDT task, we measured accuracy (i.e., hit rate) and response time (RT) for each participant. As has been frequently reported, responses to words are faster (shorter RTs) and more accurate (higher hit rates) than responses to pseudowords (e.g., Wühr & Heuer, 2022). Thus, finding this same pattern would function as a sanity check for our data. Note, however, that our main regression analyses were performed with RT as criterion, due to our interest on access and processing speed.

As is a common practice in related studies (e.g., Yap et al., 2015), we excluded incorrect trials and trials with RTs faster than 200 ms and 2.5 standard deviations above the participant mean (12.28% of data was removed for RT analyses). To compare accuracy and RTs across the different concept types, we performed a 2×2 repeated measures ANOVA, semantic content (concrete vs. abstract) and the cue type (words vs. pseudowords) as factors. Recall that the pseudowords were derived from a corresponding word, which is why we label them concrete or abstract, even though they are not proper words. In total, we had 60 words and pseudowords for concrete and abstract concepts. Table 4 shows the descriptive statistics.

To test for evidence of semantic processing, we performed 2×2 repeated measures ANOVAs using accuracy and RTs as dependent variables. For the accuracy-dependent variable, we found a significant main effect of cue type (words vs. pseudowords), $F(1, 100) = 33.1$, $MS_e = .01$, $p < .001$, $\eta_p^2 = .25$, no significant main effect of semantic content (concrete vs. abstract), $F(1, 100) = .40$, $MS_e = .002$, $p = .53$, $\eta_p^2 = .004$, and a significant interaction between cue type and semantic content, $F(1, 100) = 4.60$, $MS_e = .001$, $p = .034$, $\eta_p^2 = .04$. As expected from previous literature (Pexman, 2012), we found no statistical differences when comparing semantic content (i.e., concrete vs. abstract words) for both types of cues (words vs. pseudowords) (see Figure 2 panel A). The overall significant difference in the ANOVA is driven by the words against pseudowords comparison.

When using RTs as dependent variable, we observed a significant main effect of cue type, $F(1, 100) = 296.3$, $MS_e = .006$, $p < .001$, $\eta_p^2 = .75$. More specifically, we found that participants' RTs were slower when dealing with pseudowords (derived either from concrete or abstract words) compared to real words. We did not find a main effect of semantic content, $F(1, 100) = .04$, $MS_e < .001$, $p = .85$, $\eta_p^2 = .001$, nor an interaction, $F(1, 100) = .21$, $MS_e < .001$, $p = .65$, $\eta_p^2 = .002$. This pattern of results suggests that semantic content is not relevant for the LDT. However, the forthcoming regression analyses offer a more nuanced view.

Table 4. Descriptive statistics on LDT performance

Type of concept	Accuracy	Response times
Concrete	.91 (.11)	.702 s (.14)
Abstract	.91 (.11)	.700 s (.14)
Concrete pseudoword	.85 (.16)	.835 s (.18)
Abstract pseudoword	.84 (.14)	.836 s (.18)

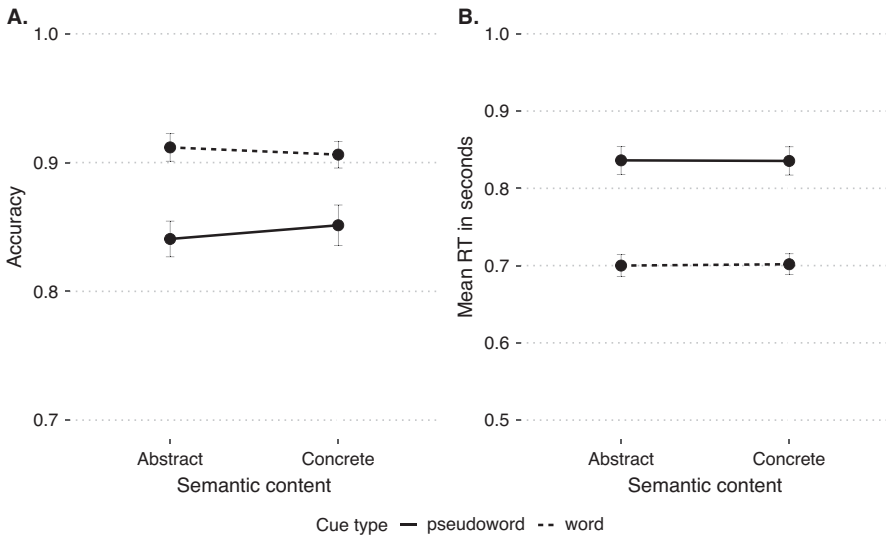*Note: Mean and SD (in parenthesis) for accuracy and RT in the LDT.*

Figure 2. Mean accuracy and RTs in the lexical decision task (LDT). *Note.* Mean responses in accuracy (panel A) and RTs (panel B) for the 2 × 2 repeated design in the LDT. Intervals show standard error of the mean.

6.1.4. *Regression analyses*

To assess the influence of the PLT mathematical model's coefficients (i.e., a_1, e_1 and e_2), other semantic variables (concreteness and imageability) and sub-lexical variables (log frequency, familiarity) we performed multiple regressions following a model comparison approach with nested models. In all our analyses, we compared regression models' fit to the data by testing differences in explained variances, and by using the Akaike information criterion (AIC; Akaike, 1973; Hurvich & Tsai, 1991). The AIC penalizes models with more parameters.

For the full data set (i.e., including concrete and abstract words), we fitted three models to the data: sub-lexical only, model coefficients only and a full model (combining sub-lexical with model coefficients). Table 5 shows the regression estimates and their corresponding *p* values for each predictor for the full dataset in the LDT. When performing analyses for the full dataset, it was not possible to use concreteness or imageability ratings as predictors. Recall that words were selected avoiding intermediate concreteness ratings and that in our data, the correlation between both ratings was high ($r = .85$). Thus, concreteness and imageability were (by design, and for the full dataset) highly correlated with RTs. Noteworthy, including the coefficients in the complete model accounts for greater RT variance (36%) in the LDT than in the other models. To test whether the models' R^2 values were significantly different, we used an *F* test for nested models (Hastie & Pregibon, 1992). Our main question regarding R^2s was whether there was a statistically significant difference between models that can be accounted for by the addition of the mathematical model's coefficients in the full regression model. Results showed that the full model was significantly better than the sub-lexical model ($F(3, 114) = 3.79, p = .012$) and then the coefficients only model ($F(2, 114) = 24.26, p < .001$). Furthermore, the full model showed a better AIC coefficient. Examining the full model shows that log frequency, familiarity and coefficients a_1 and e_2 are significant.

Table 5. Regression models on RTs for the full dataset in the LDT

	Sub-lexical		Coefficients only		Complete model	
	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value
Intercept	**1.08**	**<.001**	**.87**	**<.001**	**1.17**	**<.001**
Log frequency	**−.07**	**<.001**			**−.07**	**<.001**
Familiarity	**−.02**	**.012**			**−.03**	**.01**
a_1			**−.07**	**.01**	**−.07**	**.01**
e_1			**−.007**	**.02**	−.004	.11
e_2			**−.02**	**.02**	**−.04**	**.05**
R^2	.30		.09		.36	
Corrected AIC	−347		−313.7		−351.8	

Note: Significant results at the. 05 level are shown in bold.
Abbreviation: AIC, Akaike information criterion.

In contrast to the analysis for the entire data set, regressions within each level of the concrete/abstract factor allowed us to use concreteness and imageability ratings along with sub-lexical variables (i.e., the Sub-lexical+semantic model in Tables 6 and 7). Although the complete model achieves a better fit to the data, with log frequency, familiarity and coefficient a_1 being significant, both for concrete (Table 6) and abstract concepts (Table 7), the difference does not achieve significance. For concrete concepts, the *F* test for nested models showed that the complete model was not significantly better for predicting RTs than the sub-lexical+semantic model ($F(3, 52) = 2.72$, $p = .053$). Similarly, for the abstract concepts data, the *F* test for nested models showed that the complete model was not significantly better for predicting RTs than the sub-lexical+semantic model ($F(3, 52) = 2.77$, $p = .051$). However, for both datasets, the AIC shows that the full model that includes the equations' coefficients is the best model, and coefficient a_1 is significant in both models even in the presence of log frequency, familiarity, concreteness and imageability (see Tables 6 and 7).

6.2. SDT study

In this study, participants had to decide whether a word presented on a computer screen was abstract or concrete. We predicted that, additional to the effect of sub-

Table 6. Regression models on RTs for concrete concepts in the LDT

	Sub-lexical + semantic		Coefficients only		Complete model	
	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value
Intercept	**.99**	**<.001**	**.99**	**<.001**	**1.17**	**<.001**
Log frequency	**−.07**	**<.001**			**−.06**	**<.001**
Familiarity	−.01	.50			**−.01**	**.003**
Imageability	−.001	.96			.003	.80
Concreteness	.001	.96			−.004	.82
a_1			**−.10**	**.01**	**−.08**	**.02**
e_1			−.008	.06	−.005	.18
e_2			−.03	.54	−.02	.70
R^2	.26		.17		.36	
Corrected AIC	−163		−158.5		−163.7	

Note: Significant results at the. 05 level are shown in bold.
Abbreviation: AIC, Akaike information criterion.

Table 7. Regression models on RTs for abstract concepts in the LDT

	Sub-lexical + semantic		Coefficients only		Complete model	
	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value
Intercept	**1.09**	**<.001**	**.79**	**<.001**	**1.17**	**<.001**
Log frequency	**−.06**	**.002**			**−.07**	**<.001**
Familiarity	−.03	.06			**−.03**	**.05**
Imageability	−.02	.17			−.01	.45
Concreteness	.002	.94			−.01	.77
a_1			−.05	.26	**−.07**	**.05**
e_1			−.004	.41	.001	.83
e_2			−.07	.07	−.03	.36
R^2		.36		.10		.45
		−168.5		−150.6		−169.4
Corrected AIC						

Note: Significant results at the. 05 level are shown in bold.
Abbreviation: AIC, Akaike information criterion.

lexical variables, we would observe important effects of semantic variables, including the PLT mathematical model's coefficients. Importantly, we predicted that the coefficients should align with the semantic variables and account for an independent part of the variance, showing that they measure semantic processing, and that a full regression model including the coefficients would be the best regression model in accounting for RT variance.

6.2.1. Participants
Here, 130 undergraduate Spanish speaker students of Universidad Adolfo Ibáñez voluntarily agreed to participate in the study (M_{age} = 21.48, SD_{age} = 1.52; 31 females). Participants received a small academic incentive. All participants gave informed consent to participate. The study was approved by the Ethics Committee of Universidad Adolfo Ibáñez.

6.2.2. Materials and procedure
Materials consisted of the 120 concepts (60 concrete and 60 abstract) used in the previous LDT study. The SDT was programmed using PsychoPy (Peirce et al., 2019). Concepts were presented in random order on a computer screen (but making sure that no more than three concepts of each type were consecutively displayed), and each concept was displayed during 500 ms. Each response and its corresponding RT was recorded. Preceding each displayed word, a fixation cross appeared on screen for 250 ms. Participants were instructed to indicate whether the presented concept was concrete or abstract, by pressing two different keys on a keyboard. At the middle of the experiment, response keys were inverted (i.e., the key to indicate that a presented concept was concrete and the key to denote that the presented concept was abstract were switched). Before starting the actual experiment, a training session was conducted using the same procedure (including the switching of the response keys) but employing different words from those used in the actual experiment. The experiment for each participant lasted for about 20 minutes.

6.2.3. Results

To assess the performance on the SDT task, we measured accuracy (i.e., hit rate) and response times (RTs) for each participant. It is well documented that concrete concepts have a processing advantage over abstract concepts (i.e., higher accuracy and shorter RTs; Paivio, 1991). Thus, finding this same pattern would function as a sanity check for our data. Note, however, that our main regression analyses were done with RTs as criterion, due to our interest on access and processing speed.

As in the LDT study, we excluded incorrect trials and trials with RTs faster than 200 ms and longer than 2.5 standard deviations from the participant mean (21.4% of data was removed for RT analyses). This percentage may seem high but consider that the SDT is generally a more difficult task than the LDT, and people may genuinely differ in their judgments about whether a given concept is concrete or abstract, all of which means an increase in errors and RT relative to the LDT.

For the SDT, we followed a similar analysis approach to the LDT study. We carried out a one factor repeated measures ANOVA, using semantic content (i.e., concrete vs. abstract) as the repeated factor. Using accuracy as the dependent variable, we found a main effect of semantic content, $F(1, 129) = 75.7$, $MS_e = .025$, $p < .001$, $\eta_p^2 = .37$. As Figure 3(A) and Table 8 show, participants made more errors for abstract concepts than for concrete concepts. By examining the RTs in the SDT, we found
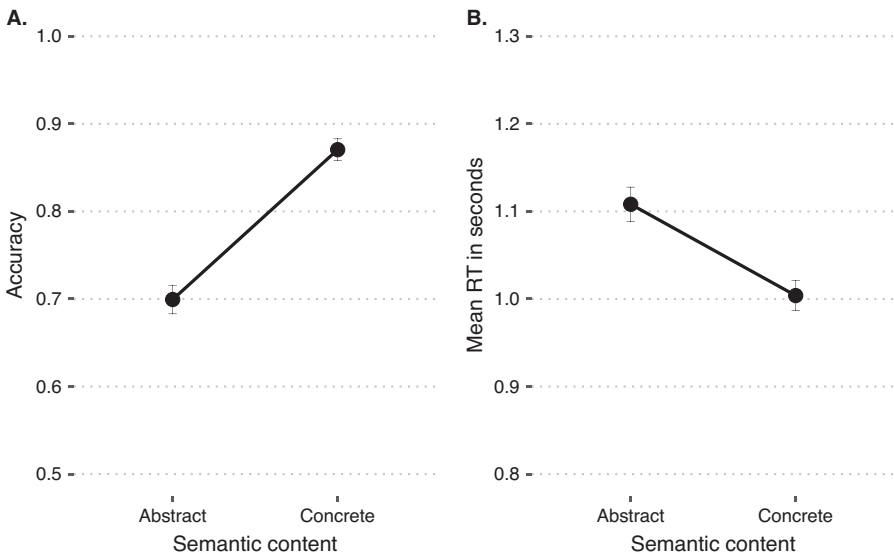

Figure 3. Mean accuracy and RTs in the semantic decision task (SDT). *Note.* Mean responses in accuracy (panel A) and RTs (panel B) for the one factor repeated design in the SDT. Intervals show the standard error of the mean.

Table 8. Descriptive statistics on SDT performance

Type of concept	Accuracy	Response times
Concrete	.87 (.14)	1.004 s (.20)
Abstract	.70 (.19)	1.108 s (.22)

Note: Mean and SD (in parenthesis) for accuracy and RT in the SDT.

again a significant main effect of semantic content, $F(1, 129) = 102.2$, $MS_e = .007$, $p < .001$, $\eta_p^2 = .44$. This shows that participants made faster responses for concrete concepts than for abstract concepts (see Figure 3(B)).

6.2.4. Regression analysis

For the full data set (i.e., concrete and abstract words), we fitted three models: sub-lexical only, model coefficients only, and a full model (combining sub-lexical with model coefficients). Table 9 shows the regression estimates and their corresponding p values for each predictor for the full dataset in the SDT. As explained above, when performing analyses for the full dataset, it was not possible to use concreteness or imageability ratings as predictors. Noteworthy, including the coefficients in the full model accounts for more RT variance (26%) in the SDT than the other nested models. To test whether the models' R^2 values were significantly different, we used the F test for nested models. Results showed that the full model was significantly better than the sub-lexical model ($F(3, 114) = 8.9$, $p < .001$) and the coefficients only model ($F(2, 114) = 4.52$, $p = .013$). Furthermore, the full model showed a better AIC coefficient. Examining the full model shows that log frequency, familiarity and coefficients a_1 and e_2 are significant.

As done for the LDT, regressions within each level of the concrete/abstract factor allowed us to use concreteness and imageability ratings along sub-lexical variables

Table 9. Regression models on RTs for the complete dataset in the SDT

	Sub-lexical		Coefficients only		Complete model	
	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value
Intercept	**1.29**	**<.001**	**1.11**	**<.001**	**1.37**	**<.001**
Log frequency	.01	.53			−.01	.60
Familiarity	**−.06**	**.002**			**−.05**	**.008**
a_1			**−.14**	**.001**	**−.14**	**<.001**
e_1			**.01**	**.005**	**.01**	**.01**
e_2			**.09**	**.03**	.08	.07
R^2	.09		.20		.26	
Corrected AIC	−191.4		−205.3		−210	

Note: Significant results at the. 05 level are shown in bold.
Abbreviation: AIC, Akaike information criterion.

Table 10. Regression models on RTs for concrete concepts in the SDT

	Sub-lexical + semantic		Coefficients only		Complete model	
	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value
Intercept	**1.66**	**<.001**	**1.2**	**<.001**	**1.73**	**<.001**
Log frequency	**−.07**	**.009**			**−.07**	**.007**
Familiarity	.003	.92			.01	.67
Imageability	**−.04**	**.05**			−.03	.15
Concreteness	−0.4	.17			−.05	.09
a_1			**−.16**	**.002**	**−.14**	**.006**
e_1			.01	.31	.01	.19
e_2			.08	.14	.09	.11
R^2	.22		.21		.38	
Corrected AIC	−116.8		−118.2		−122.6	

Note: Significant results at the. 05 level are shown in bold.
Abbreviation: AIC, Akaike information criterion.

Table 11. Regression models on RTs for abstract concepts in the SDT

	Sub-lexical + semantic		Coefficients only		Complete model	
	Estimate	*p*-Value	Estimate	*p*-Value	Estimate	*p*-Value
Intercept	**1.33**	**<.001**	**1.04**	**<.001**	**1.22**	**<.001**
Log frequency	**.08**	**.005**			**.09**	**.008**
Familiarity	**−.06**	**.004**			**−.06**	**.005**
Imageability	**−.05**	**.01**			**−.05**	**.01**
Concreteness	.01	.65			−.01	.82
a_1			−.01	.85	.05	.42
e_1			.01	.19	.001	.91
e_2			.04	.44	−.003	.95
R^2		.34		.04		.35
Corrected AIC		−115.9		−95.7		−108.9

Note: Significant results at the .05 level are shown in bold.
Abbreviation: AIC, Akaike information criterion.

(i.e., the Sub-lexical+semantic model in Tables 10 and 11). For concrete concepts, the full model achieved a better fit to the data, with log frequency, familiarity and a_1 coefficient being significant ($F(3, 52) = 4.48$, $p = .007$). Furthermore, the full model also outperformed the Sub-lexical+semantic model according to the AIC. However, for abstract concepts, the difference in R^2 did not achieve significance ($F(3, 52) = .28$, $p = .84$), and the AIC favored the partial model including only Sub-lexical+semantic variables.

7. Discussion

The mathematical model described in the current paper aims at understanding how SM is accessed during language processing tasks. The model includes coefficients that represent different aspects of semantic accessibility. These coefficients are derived from regression analyses and are used to predict RTs in LDTs and SDTs. The model provides quantitative measures for the fact that concepts with denser semantic neighborhoods and more features are easier to process (e.g., Pexman et al., 2002; 2003). The coefficients in the model provide insights into the relationship between different variables and the speed of access to semantic information.

In the current work, we tested the model's ability to predict RTs in the LDT and SDT tasks, extending its applicability. In the LDT, participants are presented with a string of letters and are asked to determine whether the string forms a real word or a pseudoword. The SDT requires participants to decide whether a given word shown on a screen is concrete or abstract. Concrete words refer to objects or things that can be perceived through the senses, while abstract words represent ideas, concepts or emotions. The task examines how semantic variables, such as concreteness, influence participants' RTs and accuracy in categorizing words.

Our results are consistent with the main trends that characterize the performance in LDT and SDT tasks according to prior literature. On the LDT, the main variables that predict RTs are sub-lexical variables. In previous research, participants tend to have slower RTs and lower accuracies when dealing with pseudowords compared to real words, a pattern we replicate in our results. On the SDT, though both sub-lexical and semantic variables play a role, semantic variables have a more significant impact on performance. Also, concrete concepts tend to be associated with higher accuracy

and shorter RTs compared with abstract concepts, which is the pattern we replicate in our results (Schwanenflugel et al., 1988; Plaut & Shallice, 1993; though see Kousta et al., 2011; Barber et al., 2013).

In this context, the current work aimed at comparing the PLT mathematical model against more traditional psycholinguistic variables. If the model in fact offers a measure of semantic access for a concept, then it should predict RTs in the SDT. As a more stringent test, we also tested it in the LDT, given that there is evidence that semantic variables do allow predicting performance in the LDT (Pexman et al., 2017).

Consistent with our overarching hypothesis, the coefficients clustered with the semantic and not with the sub-lexical variables (see Table 2). Noteworthy, we found positive correlations for the a_1 coefficient (higher coefficients indicate higher concreteness ratings), and negative correlations for the e_1 coefficient (i.e., a larger e_1 coefficient indicates lower concreteness ratings), a pattern that is consistent with our theoretical analysis of Eq. (2).

In the LDT, semantic variables (i.e., concreteness and imageability) were not found to be predictive of RTs. When the same nested regression models were tested within concrete and within abstract concepts, though the full model including the mathematical model coefficients outperformed the simpler models according to the AIC, the F test comparisons did not achieve significance. Adding the PLT mathematical model's coefficients did not significantly improve the prediction of the RTs. However, when coefficients were included in the full model, the a_1 coefficient achieved significance in both concrete and abstract concepts, even in the presence of the sub-lexical variables. This suggests that the semantic accessibility of a concept, as measured by the coefficients, influences performance in the LDT. Interestingly, concreteness and imageability were not predictive of RTs.

In the SDT, when the full data set and concrete concepts were considered, the coefficients from the PLT mathematical model increased the prediction of the RTs. For the full data set, including a_1 and e_1 significantly improved the regression model. When only concrete concepts were considered and a_1 was included in the full model, the effect of imageability was lost, which is consistent with their shared variance (see Table 2), and adding the PLT model's coefficients significantly improved explained variance. This suggests that coefficients, in particular a_1, reflect semantic information that affects processing speed. We will discuss the contradictory results we obtained for abstract concepts in the SDT in the next subsection.

In conclusion, the current work offers evidence that the PLT model's coefficients (a_1, e_1 and e_2) reflect different aspects of semantic accessibility. The a_1 coefficient represents the ease of access to a concept's semantic information. A higher a_1 coefficient indicates easier access to semantic content. This suggests that the concept is more readily available in memory and that its semantic information can be accessed more easily. The e_1 and e_2 coefficients reflect semantic richness. A relatively smaller e_1 coefficient and/or a relatively larger e_2 coefficient indicate a richer concept. These measures provide insights into how easily semantic information can be accessed and processed. Furthermore, though the PLT mathematical model's coefficients correlate with concreteness and imageability ratings, they make an independent contribution to the prediction of RTs in LDT and SDT tasks, suggesting that they tap onto a different dimension of the concrete/abstract dimension from that which is captured by the psycholinguistic ratings.

7.1. Consequences for the concrete/abstract distinction

There is extensive literature discussing the disparities between abstract and concrete concepts. Consistently with our results, concrete concepts are generally considered easier to learn and process compared with abstract concepts (the *concreteness effect*, e.g., Jones, 1985; Paivio, 1991; Vigliocco et al., 2018; Walker & Hulme, 1999). Some researchers propose that the distinction between the two can be qualitatively characterized by the type of features associated with each type of concept (Barsalou et al., 2008; Paivio, 1986; Wiemer-Hastings & Xu, 2005). Concrete concepts are best described by perceptible physical properties, whereas abstract concepts are linked to other types of features. The specific characterization of these other features varies among authors (Barsalou et al., 2008; Borghi & Cimatti, 2009; Borghi et al., 2017; Breedin et al., 1994; Paivio, 1986; Wiemer-Hastings & Xu, 2005). The qualitative view has wide appeal, as evidenced from the *de facto* definition of concreteness implicit in instructions given to subjects that provide concreteness ratings (Löhr, 2022).

While qualitative distinctions are likely to be critical to our understanding of the concrete/abstract distinction, our current work takes a processing perspective that emphasizes quantitative differences between abstract and concrete concepts. From a quantitative point of view, concrete concepts are characterized by a larger number of conceptual features and exhibit stronger contextual associations (Plaut & Shallice, 1991; 1993; Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983). Consistently with these differences, research suggests that SM is more densely structured for concrete concepts, facilitating easier access (Jones, 1985; Plaut & Shallice, 1993; Recchia & Jones, 2012; Reilly & Desai, 2017; Yap & Pexman, 2016). This richer semantic structure of concrete concepts relative to abstract ones aligns with abstract concepts having multiple senses compared to concrete concepts (Hoffman et al., 2013).

Because the PLT model's coefficients correlate with concreteness and imageability ratings and are able to predict processing speed in ways consistent with the concreteness effect, we believe that the PLT mathematical model offers measures of the concrete/abstract dimension. As such, we believe that this model can be considered a summary of previous literature that adopts a quantitative perspective on the topic. However, we acknowledge that when only abstract concepts were considered, the PLT mathematical model's coefficients did not follow the expected pattern in the SDT. Only log frequency, familiarity and imageability contributed to the prediction of RTs. Because all our independent variables were the same across all our analyses for the LDT and SDT, the only possible culprit of this null result is the distribution of RTs in the SDT for abstract concepts. Visual inspection of this distribution showed that not only were RTs higher on average than in the other conditions (see Tables 4 and 8), but that a large proportion of data points were well above the 1 *s* mark (see Supplementary Material, Figure 1, at https://osf.io/zsn4c/). Thus, it is possible that subjects experienced more difficulty when having to classify abstract concepts, perhaps engaging more reflective and relatively slower processes when responding.

Interestingly, a similar pattern was previously reported. In a study using the PLT mathematical model to classify concrete and abstract concepts (Canessa et al., 2021), better classifications were achieved for concrete than for abstract concepts. The explanation provided for those results was that concrete concepts in that study were more homogeneously concrete and that abstract concepts were more graded in that

same factor. This is consistent with the observation that concreteness ratings for abstract concepts are often bimodal, with some subjects judging them relatively low in concreteness and others judging them relatively higher in concreteness (Gary Lupyan, personal communication, 08-29-2023), which again suggests that judging abstract concepts is harder because they require subjects to reflect in order to decide on a specific sense that is relevant to their current task. In the future, it will be interesting to control for number of different senses. There is evidence that words with multiple meanings (e.g., *bark*, which has different and unrelated senses) are slower to process in lexical decision tasks, but words with multiple but related senses are relatively faster (e.g., *twist* which has different but related senses) (Rodd et al., 2002, 2004). This has been explained as an effect of competition (different meanings compete and make access difficult). In our model, abstract concepts are precisely those with multiple senses, not those with multiple meanings. Thus, we would predict that, for example, Equation 1's coefficient would vary with the number of senses (but not with the number of meanings).

Also, as previously discussed, it may be that abstract concepts are a heterogeneous bunch (e.g., actions, emotions and values, cognitive processes, social institutions moral terms; see Dove, 2022), and cannot be easily placed on a single univariate concreteness dimension. An anonymous reviewer was concerned about whether emotional processing (such as valence and arousal) could explain differences in response times between concrete and abstract concepts. We performed a brief correlational analysis of our empirical distributions of RTs in the SDT with valence and arousal values obtained for the 14031 Spanish word database from Stadthagen-Gonzalez et al. (2017). Our results (shown in Supplementary Material, Figure 2) show that these variables do not account for differences in RT in SDT, suggesting that emotional variables do not account for our findings. However, we believe that designing studies using different types of abstract concepts (such as emotional abstractions) is something we would like to explore in the future.

With relative independence of the actual cognitive or neural mechanisms that account for the concreteness effect, our contribution to this literature is that we offer a model that is able to measure the ease of semantic access and processing that depends on very general assumptions about semantic access, that generalizes across at least two different tasks, is correlated with traditional measures of concreteness, but that taps on a source of variance in the LDT and SDT tasks that is independent from variance accounted for by traditional psycholinguistic variables.

References

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, 8(3), 323–338.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In Petrov, B.N., Csaki, F. (Eds.), *Second international symposium on information theory* (pp. 267–281). Akademiai Kiado.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology. General*, 128(1), 32–55.

Barber, H. A., Otten, L. J., Kousta, S., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, 125, 47–53.

Baroni, M., & Lenci, A. (2008). Concepts and properties in word spaces. *Italian Journal of Linguistics*, 20(1), 55–88.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London. B: Biological Sciences*, 358(1435), 1177–1187. https://doi.org/10.1098/rstb.2003.1319.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning*. Oxford University Press.

Borghi, A, M, Barca, L., Binkofski, F., & Tummolini, L. (2018). Varieties of abstract concepts: Development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170121. https://doi.org/10.1098/rstb.2017.0121

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, 143(3), 263–292.

Borghi, A. M., & Cimatti, F. (2009). Words as tools and the problem of abstract word meanings. In Taatgen, N., & van Rijn, H. (Eds). *Proceedings of the 31st annual conference of the cognitive science society, 2304-2309*. Amsterdam: Cognitive Science Society. Retrieved from https://escholarship.org/uc/item/58m9n8rp.

Borghi, A. M., Shaki, S., & Fischer, M. H. (2022). Abstract concepts: External influences, internal constraints, and methodological issues. *Psychological Research*, 86(8), 2370–2388. https://doi.org/10.1007/s00426-022-01698-4

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.

Bridger, E.K., Bader, R., & Mecklinger, A. (2014). More ways than one: ERPs reveal multiple familiarity signals in the word frequency mirror effect. *Neuropsychologia*, 57, 179–190.

Canessa, E., & Chaigneau, S. E. (2020). Mathematical regularities of data from the property listing task. *Journal of Mathematical Psychology*, 97, 102376. https://doi.org/10.1016/j.jmp.2020.102376

Canessa, E., Chaigneau, S. E., & Moreno, S. (2021). Language processing differences between blind and sighted individuals and the abstract versus concrete concept difference. *Cognitive Science*, 45(10), e13044. https://doi.org/10.1111/cogs.13044

Canessa, E., Chaigneau, S. E., & Moreno, S. (2023). Describing and understanding the time course of the property listing task. *Cognitive Processing*, 25, 61–74. https://doi.org/10.1007/s10339-023-01160-2

Canessa, E., Chaigneau, S.E., Moreno, S., & Lagos, R. (2020). Informational content of cosine and other similarities calculated from high-dimensional conceptual property norm data. *Cognitive Processing*, 21, 601–614. https://doi.org/10.1007/s10339-020-00985-5

Chaigneau, S. E., Canessa, E., Barra, C., & Lagos, R. (2018). The role of variability in the property listing task. *Behavior Research Methods*, 50(3), 972–988.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology. General*, 123(4), 354–373.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods*, 36(3), 384–387.

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.

Devereux, B., Pilkington, N., Poibeau, T., & Korhonen, A. (2009). Towards unrestricted, large-scale acquisition of feature-based conceptual representations from corpus data. *Research on Language and Computation*, 7(2), 137–170.

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.

Dove, G. (2022). *Abstract concepts and the embodied mind: Rethinking grounded cognition*. Oxford Academic.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, 45(4), 1246–1258.

Fagarasan, L., Vecchi, E., & Clark, S. (2015). From distributional semantics to feature norms: Grounding semantic models in human perceptual data. In *Proceedings of the 11th international conference on computational semantics* (pp. 52–57). Association for Computational Linguistics.

Fox, J., Dennis, S., & Osth, A. F. (2020). Accounting for the build-up of proactive interference across lists in a list length paradigm reveals a dominance of item-noise in recognition memory. *Journal of Memory and Language*, 110, 104065.

Hastie, T. J. and Pregibon, D. (1992) Generalized linear models. In: Chambers, J. M. and Hastie, T. J. (Eds.), *Statistical models in S*. Wadsworth & Brooks/Cole, Chapter 6.

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: How we search through memory. *Topics in Cognitive Science*, 7(3), 513–534.

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(3), 718–730.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 25(3), 721–742.

Hough, G., & Ferraris, D. (2010). Free listing: A method to gain initial insight of a food category. *Food Quality and Preference*, 21(3), 295–301.

Hurvich, C. M., & Tsai, C.-L. (1991) Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, 78, 499–509.

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language*, 24(1), 1–19. https://doi.org/10.1016/0093-934X(85)90094-X

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 26(2), 336–358.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1(3), 153–161.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627–633.

Khanna, M. M., & Cortese, M. J. (2021). How well imageability, concreteness, perceptual strength, and action strength predict recognition memory, lexical decision, and reading aloud performance. *Memory*, 29(5), 622–636. https://doi.org/10.1080/09658211.2021.1924789

Kliegl, O., & Bäuml, K. H. T. (2021). Buildup and release from proactive interference–Cognitive and neural mechanisms. *Neuroscience & Biobehavioral Reviews*, 120, 264–278.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, 140(1), 14–34.

Kremer, G., & Baroni, M. (2011). A set of semantic norms for German and Italian. *Behavior Research Methods*, 43(1), 97–109.

Langland-Hassan, P., &Davis, C. P. (2023). A context-sensitive and non-linguistic approach to abstract concepts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 378, 20210355. https://doi.org/10.1098/rstb.2021.0355.

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.

Löhr, G. (2022). What are abstract concepts? On lexical ambiguity and concreteness ratings. *Review of Philosophy and Psychology*, 13(3), 549–566.

Maki, W. S. (2007). Judgments of associative memory. *Cognitive Psychology*, 54(4), 319–353.

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.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighbourhood density: Processing is slowed by near neighbours and speeded by distant neighbours. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 34, 65–79.

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.

Muraki, E. J., Sidhu, D. M., & Pexman, P. M. (2020). Mapping semantic space: Property norms and semantic richness. *Cognitive Processing*, 21(4), 637–649.

Neville, D. A., Raaijmakers, J. G., & van Maanen, L. (2019). Modulation of the word frequency effect in recognition memory after an unrelated lexical decision task. *Journal of Memory and Language*, 108, 104026.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 45(3), 255–287. https://doi.org/10.1037/h0084295

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perea, M., Marcet, A., Vergara-Martínez, M., & Gomez, P. (2016). On the dissociation of word/nonword repetition effects in Lexical decision: An evidence accumulation account. *Frontiers in Psychology*, 7, 215.

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.

Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition: Meaning and context, individuals and development* (pp. 24–43). Psychology Press.

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

Pexman, P. M., Holyk, G. G., & Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, 31(6), 842–855.

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, 9(3), 542–549. https://doi.org/10.3758/bf03196311

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language & Cognitive Processes*, 12, 765–805.

Plaut, D. C., & Shallice, T. (1991). Effects of word abstractness in a connectionist model of deep dyslexia. In *Proceedings of the 13th annual meeting of the cognitive science society* (pp. 73–78). Erlbaum.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10(5), 377–500.

Recchia, G., & Jones, M. N. (2012). The semantic richness of abstract concepts. *Frontiers in Human Neuroscience*, 6(15). https://doi.org/10.3389/fnhum.2012.00315

Reilly, M., & Desai, R. (2017). Effects of semantic neighborhood density in abstract and concrete words. *Cognition*, 169, 46–53. https://doi.org/10.1016/j.cognition.2017.08.004

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 245–266. https://doi.org/10.1006/jmla.2001.2810

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, 28(1), 89–104. https://doi.org/10.1016/j.cogsci.2003.08.002

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology. General*, 126, 211–227. https://doi.org/10.1037/0096-3445.126.3.211

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.

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.

Segalowitz, S., & Lane, K. (2000). Lexical access of function versus content words. *Brain and Language*, 75, 376–389. https://doi.org/10.1006/brln.2000.2361.

Simon, H. A. (1964). The construction of social science models. In G. A. Miller (Ed.), *Mathematics and psychology* (pp. 137–146). John Wiley & Sons, Inc.

Spreen, O., & Schulz, R. W. (1966). Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior*, 5(5), 459–468.

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods*, 49(1), 111–123.

Taylor, K. I., Devereux, B. J., & Tyler, L. K. (2011). Conceptual structure: Towards an integrated neurocognitive account. *Language and Cognitive Processes*, 26(9), 1368–1401.

Troyer A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, 22(3), 370–378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V; FT370

Vigliocco, G., Ponari, M., & Norbury, C. (2018). Learning and processing abstract words and concepts: Insights from typical and atypical development. *Topics in Cognitive Science*, 10(3), 533–549.

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(4), 422–488.

Villani, C., Lugli, L., Liuzza, M., & Borghi, A. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition*, 11(3), 403–430. https://doi.org/10.1017/langcog.2019.23

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.

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.

Walker, L. J., & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology*, 86(4), 629–647.

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 2(5), 440–445.

Wickens, D. D., Dalezman, R. E., & Eggemeier, F. T. (1976). Multiple encoding of word attributes in memory. *Memory & Cognition*, 4(3), 307–310.

Wickens, D. D., Moody, M. J., & Dow, R. (1981). The nature and timing of the retrieval process and of interference effects. *Journal of Experimental Psychology. General*, 110(1), 1–20.

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, 29(5), 719–736.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, 132(2), 173–189.

Wühr, P., & Heuer, H. (2022). Mapping effects in choice-response and go/no-go variants of the lexical decision task: A case for polarity correspondence. *Quarterly Journal of Experimental Psychology*, 75(3), 491–507. https://doi.org/10.1177/17470218211043860

Yap, M. J., & Pexman, P. M. (2016). Semantic richness effects in syntactic classification: The role of feedback. *Frontiers in Psychology*, 7, 1394.

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon project. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 41(3), 597–613.

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, 18(4), 742–750. https://doi.org/10.3758/s13423-011-0092-y