ARTICLE

The semantic content of concrete, abstract, specific, and generic concepts

Caterina Villani[1] (iD), Adele Loia[1] and Marianna M. Bolognesi[1,2] (iD)

[1]Department of Modern Languages, Literatures, and Cultures, University of Bologna, Bologna, Italy; [2]Faculty of Medieval and Modern Languages, University of Oxford, Oxford, United Kingdom of Great Britain and Northern Ireland
Corresponding author: Caterina Villani; Email: caterina.villani6@unibo.it

Abstract

Abstraction processes involve two variables that are often confused with one another: concreteness (*banana* versus *belief*) and specificity (*chair* versus *furniture* or *Buddhism* versus *religion*). Researchers are investigating the relationship between them, but many questions remain open, such as: What type of semantics characterizes words with varying degrees of concreteness and specificity? We tackle this topic through an in-depth semantic analysis of 1049 Italian words for which human-generated concreteness and specificity ratings are available. Our findings show that (as expected) the semantics of concrete and abstract concepts differs, but most interestingly when specificity is considered, the variance in concreteness ratings explained by semantic types increases substantially, suggesting the need to carefully control word specificity in future research. For instance, mathematical concepts (*phase*) are on average abstract and generic, while behavioral qualities (*arrogant*) are on average abstract but specific. Moreover, through cluster analyses based on concreteness and specificity ratings, we observe the bottom-up emergence of four subgroups of semantically coherent words. Overall, this study provides empirical evidence and theoretical insight into the interplay of concreteness and specificity in shaping semantic categorization.

Keywords: conceptual knowledge; abstraction; specificity; concreteness; semantic analysis

1. Introduction

Language is a powerful communication system, which gives us tools to refer to both concrete and abstract concepts. Concrete concepts are defined as concepts that designate referents in the world that can be experienced directly through our sensory modalities. Conversely, abstract concepts do not directly designate referents that can be experienced through our sensory modalities. Concrete and abstract concepts behave differently in our minds. Research shows that, overall, concrete concepts are processed faster and more accurately than abstract concepts in lexical decision

and word naming tasks (e.g., Binder et al., 2005; James, 1975; Kroll and Merves, 1986; Paivio, 1986; Schwanenflugel et al., 1988; Schwanenflugel and Stowe, 1989). Moreover, words expressing concrete concepts on average are acquired earlier (Gleitman et al., 2005; Ponari et al., 2018; Wauters et al., 2003), are more imageable (Paivio, 1971), and are more easily associated with contextual information (Davis et al., 2020; Schwanenflugel, 1991) than abstract concepts. While this general phenomenon, often dubbed as 'the concreteness effect' (Paivio, 1991), seems to be quite established, researchers also report reversed effects, showing a processing advantage of abstract over concrete words when various psycholinguistic variables are controlled (Barber et al., 2013; Kousta et al., 2011). These findings suggest that something has been overlooked and that further analyses are required to unpack the complex notion of conceptual concreteness.

Recently, it has been shown that both concrete and abstract concepts vary in specificity (Bolognesi et al., 2020; Bolognesi and Caselli, 2023). In other words, both concrete and abstract concepts can define wide, highly inclusive conceptual categories or highly precise and constrained conceptual categories. For instance, FRUIT is a conceptual category that includes a variety of concrete referents: bananas, apples, strawberries, and so on. As such, the conceptual category FRUIT is more inclusive than, for instance, the conceptual category BANANA. Both categories are concrete, or better, they apply to concrete referents: all instances of bananas are typically concrete (if we are using the word 'banana' in its most literal sense), and the same goes with all instances of FRUIT. Similarly, we have words denoting abstract concepts at different levels of specificity. For instance, RELIGION is an abstract concept that denotes a category that includes various subtypes: HINDUISM, CHRISTIANITY, and so forth. Each of these subcategories is less inclusive and therefore more specific than RELIGION. While it is quite uncontroversial that the semantics of concrete and abstract concepts is qualitatively different (see the next section for a deeper elaboration on this issue), the reader may infer that categorical specificity does not affect the semantic type of a concept: concepts on the same taxonomic ladder belong to the same semantic type. For example, BEAGLE and DOG are both mammals, both referring to the same semantic type.[1] Therefore, it could be expected to see semantically different types of concepts placed along the axis of concreteness (from abstract to concrete concepts), but semantically similar types of concepts placed along the axis of specificity (from generic to specific concepts). However, this may not be the case. As a matter of fact, for some types of concepts, it may be easier to observe a variation in specificity than for others. Consider the Linnaean taxonomy of the animal kingdom: here, a concept like ANIMAL is perceived to be highly generic and highly inclusive, while conceptual categories like MAMMAL, DOG, and DALMATIAN are perceived to be increasingly more specific (and therefore less inclusive, because they denote an increasingly smaller range of individuals). All concepts on this ladder (namely ANIMAL, MAMMAL, DOG, and DALMATIAN) belong to the same semantic type, and therefore, this semantic type, which we can label as 'living beings', is distributed along the specificity axis, with datapoints toward both highly generic

---

[1]Note that for the purpose of this set of studies we are not going to take a standpoint about how 'semantics' can be operationalized in semantic memory (i.e., whether we support a featural view of semantics, a view based on how the sensory modalities contribute to shaping the perceptual strength of a concept's semantics, a distributional view of semantics). For further elaborations on different definitions of semantics, we refer the reader to the consensus paper 'What We Mean When We Say Semantic: A Consensus Statement on the Nomenclature of Semantic Memory', by Reilly et al. (under review)

and highly specific poles. Other conceptual categories including many abstract concepts do not seem to allow for long taxonomic ladders. Consider the abstract concept ESSENCE. This concept is arguably not only abstract but also quite generic, and it appears to be quite difficult to think of a taxonomic ladder that includes types of essence or more generic categories linked to this concept. This abstract concept, therefore, belongs to a semantic type that may be found only toward the highly generic pole of the specificity axis. In other words, some concepts allow us to develop long ladders of taxonomic relationships, from highly generic to highly specific, while others may develop extremely short ladders that remain quite generic or quite specific. It follows that we can expect to find concepts belonging to different semantic types associated with, respectively, high and low degrees of concreteness, as well as with high and low degrees of specificity.

This gap in the scientific literature was partially addressed in a preliminary study, where Bolognesi et al. (2020) reported a positive and significant correlation between concreteness and specificity. In particular, concepts that are concrete and specific are prototypical instances of concrete concepts (e.g., PARROT, CARROT). Concepts that are abstract and generic are typical instances of abstract concepts (e.g., FREE-DOM, DEMOCRACY). Yet, such correlation is mild (around .03), suggesting that there are also concepts that are concrete but generic, as well as concepts that are quite abstract but specific. The former type seems to include several mass nouns, such as SUBSTANCE or CROWD. The latter type seems to include human-born concepts belonging to the social reality, such as types of religions (e.g., BUDDHISM) or specific aspects or arguments involved in trials (e.g., SUMMONS).

However, while these first qualitative intuitions seem promising, the analysis on which they are based was conducted on human-generated concreteness ratings (Brysbaert et al., 2014) and specificity scores extracted automatically from WordNet, which is a lexical database constructed by lexicographers that used all sorts of external knowledge bases. Such specificity ratings do not reflect speakers' intuitions, as discussed by the authors themselves. A more systematic and fine-grained analysis of the semantic differences between concepts varying in concreteness *and* in specificity is needed, if we want to further the research on the general phenomenon of abstraction and unpack the complex relationship between these two variables: concreteness and specificity.

Thus, the aim of this study is to investigate the semantic differences between concepts that vary in concreteness and in specificity. We tackle this objective by addressing the following research questions:

RQ1: To what extent does conceptual concreteness depend on the semantics that characterizes different concepts, and how does this relationship change when specificity is controlled?

RQ2: Considering the four quadrants that we obtain by crossing concreteness and specificity, which semantic types characterize each quadrant?

RQ3: What clusters emerge from the data when we analyze concepts considering their concreteness and specificity scores? What are the semantic types associated with each cluster? To what degree do the identified clusters align with the four quadrants?

Given the scientific literature illustrated above, we hypothesize that:

Hp1: The semantics of a concept is a good predictor of its concreteness. Moreover, we expect to find the semantics to be an even better predictor of concreteness when specificity is included as a covariate, namely, when specificity is controlled.

Hp2: When crossing specificity and concreteness data, we obtain four quadrants in which we can observe words characterized by: high Specificity + high Concreteness; high Specificity + low Concreteness; low Specificity + high Concreteness; and low Specificity + low Concreteness. These words correspond to different semantic types, in line with preliminary, qualitative research (Bolognesi et al., 2020).

Hp3: When a word is characterized by its concreteness and its specificity, we expect to see emerging from the data clusters that have good internal coherence. However, given the exploratory character of the analysis, we formulated only highly general predictions. We predict that the degree of specificity and concreteness influences the organization of clusters, reflecting in different semantic types.

We report three studies based on a manual semantic annotation (conducted in a formal content analysis with inter-rater reliability tests) of a dataset of words for which concreteness and specificity scores are available. Each of the three studies hereby reported addresses one of the three research questions.

2. Theoretical background
2.1. Unpacking conceptual concreteness

Several scholars have recently raised significant criticisms about the 'concreteness' construct (Langland-Hassan and Davis, 2023; Löhr, 2023), which is typically operationalized through concreteness ratings that measure the degree of perceptibility of word referent toward five senses (Brysbaert et al., 2014). Defining concepts purely based on whether they are perceivable or not (i.e., as concrete or abstract) falls short in capturing differences between concepts as it is not informative about their semantic contents and can even lead to misinterpretations of empirical evidence. For example, it has been demonstrated that multidimensional measures of individual perceptual and action modalities associated with words are a better predictor of many cognitive tasks than concreteness ratings (Connell and Lynott, 2012; Lynott et al., 2020). A growing number of studies are working to unravel other experiential dimensions that might significantly contribute to language processing. Among these dimensions are emotional valence and dominance (Kousta et al., 2011), modality of acquisition (Della Rosa et al., 2010), body–object interaction (Tillotson et al., 2008), interoception (Connell and Lynott, 2012; Villani et al., 2021a), socialness (Diveica et al., 2023; see also Fini and Borghi, 2019; Borghi, 2022), time to perceive a concept (Davis and Yee, 2023), and personal relevance (Westbury and Wurm, 2022).

In the present study, as previously mentioned, we focus on word specificity, a variable often overlooked and sometimes confused with concreteness. Specificity indicates the degree of precision of a word meaning in terms of category inclusiveness. Highly specific concepts refer to a particular instance or individual within a category (e.g., OWL, HONEYMOON), whereas highly generic and more inclusive concepts denote a group of entities or events (e.g., BIRD, VOYAGE).

The relationship between concreteness and specificity has been addressed by prior studies but mainly to deal with the issue of abstract concept representation. For

example, Borghi and Binkofski (2014) emphasized the distinction between 'abstract-ness' and 'abstraction'. In their perspective, abstraction is a common feature of concepts overall and refers to the fact that concepts serve the function of generalizing across the multitude of instances that share similar properties. So, every concept can potentially vary in degree of abstraction (e.g., CAT requires a higher degree of abstraction than SIAMESE CAT). Abstractness refers instead to the higher degree of detachment from sensorial experiences of some concepts, like abstract concepts that typically 'lack a single bounded and clearly perceivable referent'. Similarly, Langland-Hassan and Davis (2023) have recently discussed the current definition of concreteness as either diversity of concept's referents or imperceptibility, arguing that those definitions are insufficient to account for categorization process in non-verbal tasks, which is better explained by considering instances of conceptualization in a context-relative way (i.e., Trial Concreteness). In other words, the authors pointed out that the degree of abstraction required for the application of a particular concept depends critically on the relationships among a set of distraction items. For instance, in a pictorial semantic categorization task, a canonical concrete word, such as 'cow', requires a high degree of abstraction when is operationalized as a link between two concepts of 'a piece of leather' and 'bottle of milk', which have low perceptual similarity and share minimal common setting associations, and the distractor images are other beverages, leading to a low average accuracy among participants on that trial.

We do not aim to eschew the notion of concreteness, replacing it with or drawing attention to specificity. Instead, we aim to disentangle the variable of specificity from concreteness. We do so quantitatively, using specificity norming data. Previous studies showed that specificity and concreteness, two theoretically distinct aspects of word meaning and conceptual representation, are positively correlated variables: the more a concept is concrete, the more it tends to be specific, when the relationship is calculated based on human judgments of concreteness and specificity and also when specificity is extracted automatically from the WordNet taxonomy (Bolognesi et al., 2020). However, the correlation is mild, suggesting that there are also words and concepts that are concrete but generic, or abstract but specific.

The analyses hereby reported allow us to understand the relationship between these two variables, concreteness and specificity, in terms of the semantics of the concepts that they characterize.

Disentangling concreteness and specificity is crucial if we are to understand the difference between concrete and abstract concepts in fair terms. As a matter of fact, it could be the case that experimental evidence in support of the concreteness effect may have selected stimuli in the two conditions (concrete concepts versus abstract concepts) that vary not only in concreteness but also in specificity. If that were the case, one could argue that specificity may play a role in the reported 'concreteness effects'.

2.2. The semantics of abstraction

Research on the semantics of concrete concepts has been addressed in a variety of ways, by tapping into how such concepts are processed by healthy subjects and clinical patients. Neuropsychological and neuroimaging evidence strongly sup-ports the subdivision of concrete meanings into distinct subcategories. Studies

examining brain-damaged patients revealed deficits in specific domains of knowledge (e.g., animals, fruits, tools, musical instruments, body parts). In particular, since the seminal works of Warrington and Shallice (1984), numerous studies have been dedicated to the investigation of the double dissociations between living and non-living entities (for review, see Capitani et al., 2003), as well as the distinction between natural kinds and artifacts (Keil, 1989). Recently, significant attention has been given to the domain of food, which can be seen as belonging to both natural objects and artifacts (Chen et al., 2016; Rumiati and Foroni, 2016). Additionally, compelling evidence has highlighted the role of perceptual and motor information in the processing of concrete concepts and action-related words (e.g., Glenberg and Gallese, 2012; Hauk et al., 2004; Martin, 2007), which elicit sensory modality-specific brain activation depending on their semantic content. Consistently, behavioral studies conducted on healthy subjects who rated the contribution of sensory–motor modalities to different categories of concrete concepts confirmed that the proportion of visual and auditory properties is higher in the representation of living entities like animals than for artifacts, which instead showed a great proportion of tactile and motor features (Gainotti et al., 2009; McRae and Cree, 2002; Vigliocco et al., 2004).

Overall, until the early 2000s, research focused mainly on the domain of concrete concepts, probably due to their characteristics such as the possession of purely physical and spatially bound referents, which facilitates the setup of many empirical tasks (Barsalou and Wiemer-Hastings, 2005). The semantic classification of abstract concepts, conversely, has been systematically addressed only more recently. A recent approach suggests that abstract concepts cannot be treated as a unitary whole as opposed to concrete ones. Rather, they are a multidimensional and heterogeneous class that, in analogy to concrete ones, can be differentiated into various sub-kinds that may describe their semantics. For instance, we can differentiate between emotional concepts, numbers, mental states, institutional concepts; religious, spiritual, philosophical, aesthetic and evaluative; and social concepts (for overviews, see Borghi et al., 2018a, 2018b; Villani, 2018; Villani et al., 2019; for a meta-analysis and a systematic review on abstract concept kinds and their brain representation, see Conca et al., 2021; Desai et al., 2018). Therefore, concrete and abstract concepts differ in semantics: they denote different semantic types. When considering categorical specificity, one may think that this variable does not hold an interesting relationship with the semantic types that characterize conceptual content. One may think that the same semantic types may cover various levels of abstraction, and therefore, for each semantic class, it may be possible to list highly generic as well as highly specific concepts. For instance, a semantic class of entities labeled as 'artifacts' may include both highly generic concepts like TOOL, DECORATION, and INSTRUMENT, as well as highly specific concepts like MUFFLER, CHRISTMAS TREE, and ELECTRIC GUITAR. However, as described in the introduction, some conceptual categories seem to remain only quite generic and they tend to not have a long cascade of more fine-grained specifications and subtypes lexicalized in language. This seems to be typically the case for abstract concepts, and such intuition is backed up by empirical evidence showing that there is a mild but significant correlation between concreteness and specificity (Bolognesi et al., 2020; Bolognesi and Caselli, 2023). Overall, concrete concepts tend to be perceived as more specific than abstract concepts, and vice versa, abstract concepts tend to be perceived as more generic than concrete concepts. Hence our first

hypothesis, in which we claim that we expect to find the semantics of a concept to be, first of all, a good predictor of its concreteness, and secondly an even better predictor of concreteness when specificity is included as a covariate, namely, when specificity is controlled.

When crossing the two variables involved in abstraction (concreteness and specificity), we obtain four qualitatively different types of concepts: generic–concrete, generic–abstract, specific–concrete, and specific–abstract. As mentioned above, previous studies have provided some preliminary observations on the different semantic content that characterizes these four quadrants. However, a more fine-grained semantic analysis of the concepts that belong to these four quadrants is missing. Hence our second analysis, which aims to investigate which semantic classes are particularly representative of which quadrants, and which semantic classes are not going to be found in which quadrants. We tackle this question without specific predictions, in a rather exploratory manner. Our approach to the semantic analysis is based on a coding scheme of semantic types (described in the next section) that was derived from theoretically motivated semantic taxonomies. We then explored how the semantic types included in the coding scheme were distributed across the four quadrants, discussing strong relationships emerged between specific quadrants and specific semantic types. Finally, if concreteness and specificity interact in such a way that they afford the emergence of different semantic classes, we expect to see such classes emerging also in a bottom-up manner, when running a hierarchical cluster analysis on concreteness and specificity ratings. This type of analysis would complement the previous analyses, with a different approach. While the previous analysis is based on cutting in a top-down manner the four quadrants generated by concrete and specificity scores, and observing the semantic types included in each quadrant, the cluster analysis aims at observing the semantics of clusters emerging in a bottom-up manner when crossing specificity and concreteness scores.

3. Content analysis: The semantic annotation of the dataset

3.1. Method

All analyses were performed with R version 4.2.3 (R Core Team, 2023). Data and scripts are made publicly available in the following Open Science Framework repository: https://osf.io/v76bz/

3.2. Materials

The stimuli used in this study consisted of approximately one thousand Italian words, for which concreteness and specificity human-generated ratings are available. The dataset was obtained by combining specificity norms collected by Bolognesi and Caselli (2023) with the Italian Affective Norms for English Words (ANEW) concreteness scores (Montefinese et al., 2014, retrieved from https://osf.io/eu7a3/). The Italian ANEW dataset was developed from translations of the 1,034 English words present in the ANEW (Bradley and Lang, 1999) and from words taken from Italian semantic norms (Montefinese et al., 2013). Both specificity and concreteness in the dataset used for these analyses are expressed on a 5-point Likert scale. In particular, specificity is measured on a scale from 1 = highly generic (i.e., concepts referring to wide and varied groups or classes) to 5 = highly specific (i.e., concepts referring to

individuals), while concreteness is measured on a scale from 1 = highly abstract (i.e., low sensory-based information) to 5 = highly concrete (i.e., high sensory-based information). For each word in the dataset, the part-of-speech tags were extracted from the 'la Repubblica' corpus (a huge corpus of Italian newspaper text, Baroni et al., 2004; see also Montefinese et al., 2014), resulting in 771 nouns, 220 adjectives, and 58 verbs. Although analysis of syntactic word class is beyond the aims of this study, it is noteworthy that part of speech is typically associated with varying degrees of concreteness (e.g., Van Loon-Vervoorn, 1984). Consistently, within our dataset, we observed that nouns are the most concrete ($M = 3.71$, $SD = 0.91$), followed by verbs ($M = 2.93$, $SD = 0.64$) and adjectives ($M = 2.74$, $SD = 0.52$). An omnibus test reveals that there is an overall effect of part of speech (F (2, 1046) = 129.11, $p < .001$) that describes a considerable 20% of the variation in concreteness ratings across words ($R^2 = 0.1965$) (see Open Science Framework (OSF) repository file 'Study1_linear_models.html'). These findings align with those of a prior study which demonstrated that part of speech is one of the significant lexical predictors of concreteness ratings ($R^2 = 0.26$; Strik Lievers et al., 2021). Given the replication of these patterns and the absence of additional predictions related to part of speech, in our study we primarily focus on word semantics.

3.3. Procedure

The semantic annotation of the dataset was performed based on a coding scheme that was developed from existing semantic taxonomies, previously used to annotate concrete and abstract concepts. In particular, the authors consulted the work conducted by Barsalou and Wiemer-Hastings (2005) and by Wu and Barsalou (2009) in which the authors provide a fine-grained taxonomy of semantic types, which was adopted as a general template for the current analysis. Moreover, the coding scheme was enriched after consulting other semantic schemes provided in the literature: McRae et al. (2005), which led us to adopt their general distinction between living and non-living entities; Warrington and Shallice (1984), from which we adopted the distinction between foods and living beings; Altarriba et al. (1999), from which we adopted the distinction between emotions and other types of abstract concepts; Setti and Caramelli (2005), from which we adopted the distinction between cognitive processes, emotions, and internal states; Ghio et al. (2013), which led us to the identification of mathematical concepts as a subclass of abstract concepts; Roversi et al. (2013), which helped us in a fine-grained classification of subtypes of artifacts; and Villani et al. (2021b), which guided us in the identification of specific subtypes of abstract concepts.

The coding scheme was developed with a hierarchical structure consisting of three levels of semantic granularity. As a first level, we identified five macro-categories (labeled as *Code 1*). Each macro-category was further divided into nested categories that define more granular semantic types (*Code 2* and *Code 3*). Table 1 provides an overview of the coding scheme. This taxonomy with increasingly higher-resolution labels allows for a detailed description of the semantics of a concept and multiple reliability tests, at each of the three levels of annotations.

To ensure consistent and accurate coding, we developed a training session prior to the actual annotation. During this session, we drafted detailed descriptions of each category, including examples and counterexamples, so that the categories were

Table 1. Coding scheme used for the analysis

Code 1	Code 2	Code 3
Macro-category	Category	Subcategory
ENTITY	Living	Animals
		Flowers and plants
		People
		Supernatural
	Bodily	Disease and trauma
		Part and product of body
	Non-living	Clothing
		Food and drinks
		Furnishing
		Natural elements
		Social artifacts
		Tools
		Vehicles
	Inner	Emotions and cognitive states
	Perceptual	
QUALITY	Non-perceptual	
SPATIO-TEMPORAL RELATIONS	Events	Actions
		Social actions
		Natural phenomena
	Temporal	
	Space	Real space
		Imaginary space
MATH-LOGIC RELATIONS		
UNDECIDED/MISCELLANEOUS		

mutually exclusive. After the training, two independent researchers, one of whom was a novice rater, coded the words listed in the dataset applying the coding scheme. Note that for each category included in the coding scheme, we developed a dedicated label. For instance, the concept FLOWER was labeled as EVF (Italian label), which corresponds to the three levels: Entity, Living, and Flower and plants. The complete coding scheme guide is stored in the OSF repository file 'Coding Scheme.docx'.

To assess the quality of the content analysis, we used *DescTools* package [version 0.99.48] (Signorell et al., 2020) in R to perform Krippendorff inter-reliability tests (Hayes and Krippendorff, 2007) on the annotations provided by the two independent coders (see OSF repository file 'Content_Analysis_ Inter_rater_reliability. html').

The inter-coder agreement was calculated for each hierarchical code separately and for the combination of all three codes (labeled as *Id_code*). The results showed high agreement between coders at each hierarchical code: for *Code 1*, based on five macro-categories the Krippendorff α was .80; for *Code 2*, based on nine nested categories, the Krippendorff α was .81; and for *Code 3*, based on 19 nested subcategories, the Krippendorff α was .84. Overall, the inter-coder agreement on the concatenation of all codes (*Id_code*) was Krippendorff α = .77. Any cases of disagreement were then resolved through consensus after discussion with a third judge.

4. Study 1: To what extent does conceptual concreteness depend on the semantics that characterizes different concepts, and how does this relationship change when specificity is controlled?

4.1. Methods

Through regression analyses, we explored the relationship between concreteness and semantic classification, using the *lm* () function in R. First, we ran individual regression analyses using concreteness as the dependent variable and the semantic classification (respectively, *Code 1*, *Code 2*, *Code 3*, and *Id_code*, namely the combination of the three levels of annotations) as independent variables. Subsequently, we ran again the regressions including specificity as a covariate.

4.2. Results

Table 2 reports the results of individual and simultaneous regression models, showing the amount of variance (R^2) explained by each predictor for concreteness ratings with and without the addition of specificity to the model. The results indicate that when specificity is added, there is an increase in the percentage of variance in concreteness explained by the semantics of the concepts. Specifically, *Code 1* explains 23.68% of the variance in concreteness ratings, which increases to 31.50% with the addition of specificity. Similarly, *Code 2* explains 39.91% of the variance, increasing to 43.55% with specificity. *Code 3* explains 57.44% of the variance, with a slight increase to 57.82% when specificity is added. *Id_code* explains 62.13% of the variance, with a slight increase to 62.46% with specificity. We hereby report the details of the latter regressions, in which semantic types and specificity are used as predictors for concreteness ratings. Regression coefficients for all models are available in the OSF repository file 'Study1_linear_models.html'.

4.2.1. Code 1 and specificity over concreteness

Semantic types included in *Code 1* and their specificity scores have a significant effect on concreteness. Math–logic, Spatio–temporal, and Quality categories show negative effects on concreteness, indicating that those semantic types are less concrete. Interestingly, specificity shows a positive and significant effect on concreteness,

Table 2. Summary of the regression output

Predictors	Concreteness scores (N = 1,049 words)				
	R^2	Adjusted R^2	Residual St. Err	F-statistic	p-value
Code 1	0.2368	0.2339	0.8122	80.99 (*df* = 4.1044)	<.001
Code 1 + Specificity	0.315	0.3117	0.7698	95.94 (*df* = 5.1043)	<.001
Code 2	0.3991	0.3944	0.7191	83.53 (*df* = 8.1006)	<.001
Code 2 + Specificity	0.4355	0.4304	0.6973	86.15 (*df* = 9.1005)	<.001
Code 3	0.5744	0.5643	0.5959	57.13 (*df* = 18.762)	<.001
Code 3 + Specificity	0.5782	0.5677	0.5936	54.9 (*df* = 19.761)	<.001
Id_code	0.6213	0.6124	0.5777	69.99 (*df* = 24.1024)	<.001
Id_code + Specificity	0.6246	0.6155	0.5754	68.09 (df = 25.1023)	<.001

indicating that higher levels of specificity are associated with higher concreteness ratings.

4.2.2. Code 2 and specificity over concreteness
Semantic types included in *Code 2* and their specificity scores have a statistically significant effect on concreteness. Events, Inner entities, Non-perceptual quality, Perceptual quality, and Temporal relations showed negative effects on concreteness. In contrast, Space and Living entities had positive effects, although the latter was only marginally significant. Non-living entities instead have a positive but no significant effect on concreteness. Furthermore, the results showed that higher specificity was positively associated with concreteness.

4.2.3. Code 3 and specificity over concreteness
Semantic types included in *Code 3* and its specificity scores yielded several significant effects on concreteness. Specifically, Clothing and Furnishing had positive effects, while Social artifacts, Actions, Emotions, Cognitive states, Natural elements, Social events, Natural phenomena, Imaginary places, Diseases, People, and Supernatural entities had negative effects. Only Body parts, Food and drinks, Flowers and plants, Real places, Vehicles, and Tools were not significant predictors of concreteness. In addition, the model confirms that specificity has a significant positive predictor of concreteness.

4.2.4. Id_code and specificity over concreteness
Semantic types of words (*Id_code*) and the level of specificity significantly predict concreteness ratings. Words that describe Body parts, Clothing, Furnishing, Food and drinks, Vehicles, Tools, Animals, Flowers and plants, People, and Real places are all positively associated with concreteness and therefore tend to be more concrete. On the other hand, words that describe Inner entities Emotions, Cognitive states, Social events, Supernatural entities, Math–logic relations, Non-perceptual quality, Perceptual quality, and Imaginary places are negatively associated with concreteness, so they tend to be more abstract. While Living entities, Social artifacts, and Actions have no strong associations with concreteness. In addition, some predictors show mixed associations. For example, Natural elements and Natural phenomena have, respectively, positive and negative associations with concreteness, and Temporal has a negative association with concreteness. Finally, specificity is positively associated with concreteness, indicating that more specific words tend to be more concrete. Fig. 1 illustrates, for *Code 1*, the coarsest level of semantic annotation, how the semantic categories distribute across specificity and concreteness.

4.3. Discussion
Overall, as expected, the semantic types included in our coding scheme (based on existing semantic taxonomies) explain a substantial amount of variance in concreteness ratings, ranging from roughly 23% when the semantic types are coarse (*Code 1*) to a staggering 62% when they are fine-grained and on multiple levels (*Id_code*). In other words, concrete and abstract concepts are indeed different in their semantic content, and the more the semantic annotation is fine-grained, the more such
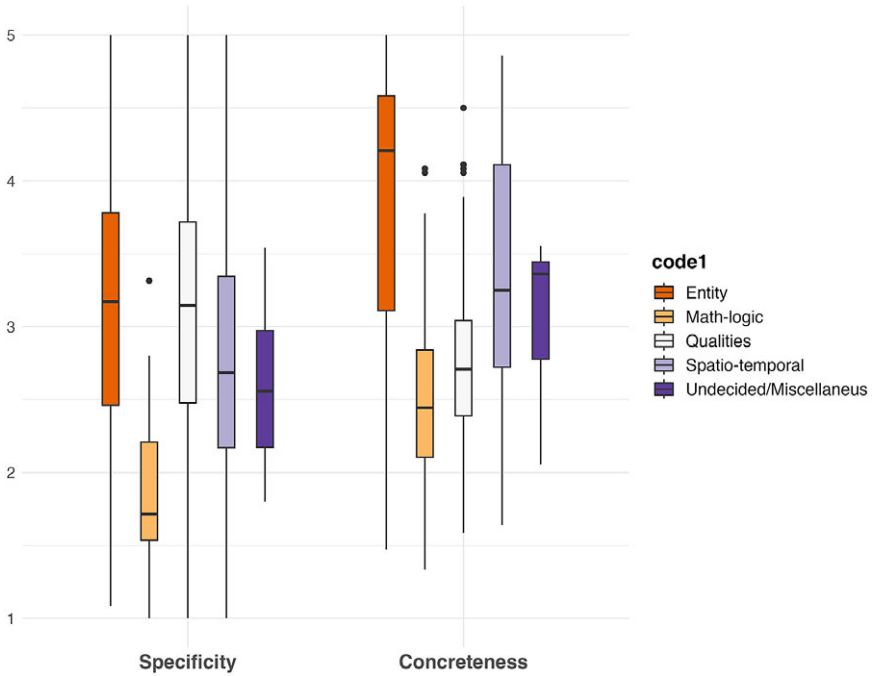
Figure 1. Semantic categories of Code 1 distributed across specificity and concreteness.

differences emerge. Adding specificity to the model, as we predicted, generally improves the model fit, which ranges roughly from 23% when the semantic types are coarse (*Code 1*) to 62% when they are fine-grained and on multiple levels (*Id_code*). Interestingly, however, the contribution of Specificity to the explanation of the variance in concreteness ratings is higher when the semantic annotation is coarser. When, conversely, the semantic annotation is fine-grained (*Code 3* and *Id_code*), Specificity adds a negligible contribution to the model, while the semantic classification explains most of the variance in concreteness. This can be observed by calculating the delta between the model that uses a semantic annotation alone as a predictor (e.g., row 1) and the model that uses the same annotation plus specificity (e.g., row 2). When the coarse semantic classification is considered (*Code 1*), Specificity alone explains 8% of the variance in concreteness ratings. When a slightly more fine-grained semantic classification is considered (*Code 2*), the individual contribution of Specificity in the model drops to 4%, to basically disappearing for more fine-grained semantic classifications. This suggests that the macro-types of concepts identified in our coding scheme (Entities, Qualities, Spatio-temporal relations, and Math–logic relations) differ substantially not only in concreteness but also in specificity, while intra-category distinctions are not quite relevant for specificity. As Fig. 1 shows, there are some interesting differences in specificity between types of abstract concepts. For instance, Mathematical and logical concepts in the dataset tend to be not only quite abstract but also quite generic, while Qualities tend to be quite abstract but more specific.

5. Study 2: Considering the four quadrants that we obtain by crossing concreteness and specificity, which semantic types characterize each quadrant?

5.1. Methods

Upon testing the distribution of specificity and concreteness ratings for bimodality, we established criteria for the classification of the words in our dataset into four quadrants: generic and abstract words; generic and concrete words; specific and abstract words; and specific and concrete words.

To investigate the potential bimodality of the concreteness and specificity rating distributions, we used *is.bimodal()* function in *LaplacesDemon* package [version 16.1.6] (Statisticat LLC, 2021) and *bimodality_coefficient()* function in *mousetrap* package [version 3.2.1] (Wulff et al., 2021) in R. The results were positive, confirming that both concreteness and specificity data can be modeled by a function that shows two humps. This allowed us to cross the two variables using the middle point of scales as a threshold criterion, to generate four distinct types of words: generic–abstract (*N* = 263), generic–concrete (*N* = 254), specific–abstract (*N* = 135), and specific–concrete (*N* = 397). The four quadrants are displayed in Fig. 2.

To test the association between the quadrants and the semantic types included in our coding scheme (both categorical variables), we ran a series of chi-square analyses. By comparing the observed and expected frequencies of occurrence of each semantic
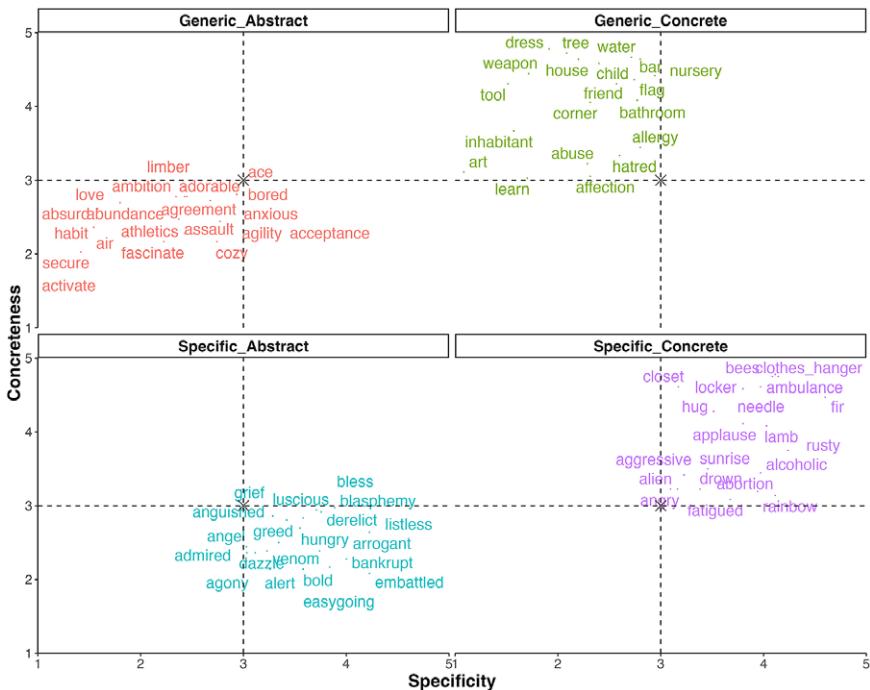

Figure 2. Four quadrants resulting from the intersection of the two variables, specificity and concreteness (only the first 20 words for each quadrant are displayed, for readability).

type in each quadrant, we tested and measured the strength of association between semantic types and quadrants using Cramer's V^2 (Kotrlik and Williams, 2003) from *rcompanion* package [version 2.4.0] (Mangiafico, 2021) and the adjusted standardized residuals. All analyses and results are available in the OSF repository files 'Study2_Chi-square.html' and 'Study2_chi-square_results.docx'.

5.2. Results

5.2.1. Code 1

This level of semantic annotation holds a strong significant relationship with the four quadrants ($\chi2(12) = 296.55$, $p < .001$, Cramer's $V = .30$). Inspection of the adjusted standardized residuals, visualized in Fig. 3, indicates that Entities typically appear in the quadrant featuring specific–concrete concepts; Qualities are typically associated with specific–abstract concepts; and Math–logic relations are often associated with generic–abstract concepts.
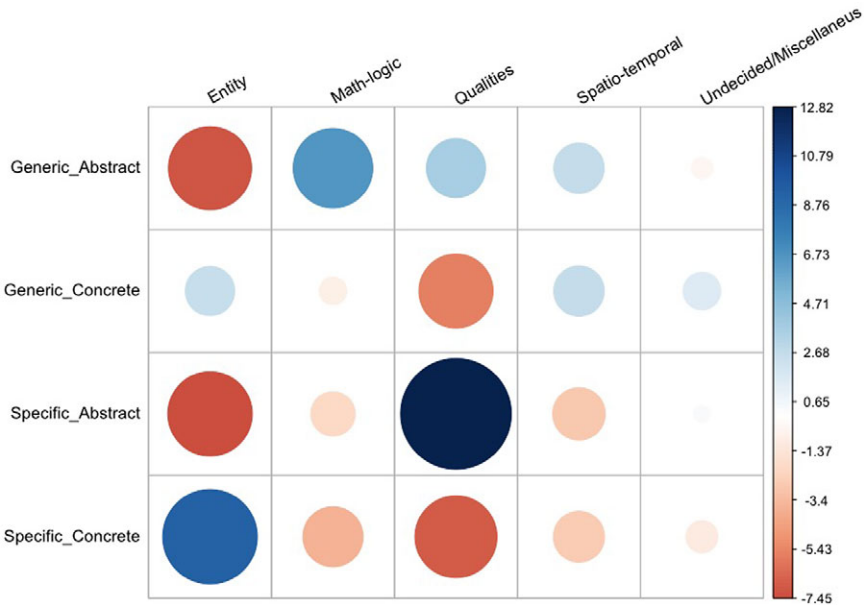

Figure 3. Correlation plot showing the standardized residuals for the chi-square analysis involving *Code 1* and the four quadrants. Bright red areas show strong negative relationships between categories, while bright blue areas show strong positive relationships. The lighter the hue (blue or red), the weaker the association between the two categories.

[2]Cramer's V is an effect size measurement for the chi-square test of independence. It is used to determine the strength of association between two categorical variables, and it ranges from 0 to 1. A value of 0 indicates no association, while a value of 1 indicates a perfect association. Typically, Cramer's V values from 0.3 to 0.5 are taken as indicators of a strong association.

5.2.2. Code 2

This level of semantic annotation holds a strong significant relationship with the four quadrants ($\chi2(24) = 432.36$, $p < .001$, Cramer's $V = .37$). Inspection of the adjusted standardized residuals, visualized in Fig. 4, reveals that concepts associated with Bodily, Living, and Non-living entities tend to appear in the quadrant featuring specific–concrete concepts. Inner entities are strongly associated with generic–abstract concepts. Among the qualities, Non-perceptual qualities are strongly associated with specific abstract concepts, while Perceptual qualities are moderately associated with specific abstract concepts. Finally, Events are typically associated with generic–abstract concepts.

5.2.3. Code 3

This level of semantic annotation holds a strong significant relationship with the four quadrants ($\chi2(54) = 426.44$, $p < .001$, Cramer's $V = .42$). Inspection of the adjusted standardized residuals, visualized in Fig. 5, indicates that among the Living entities, Animals, Food and drinks, Flowers and plants, and People tend to appear in the quadrant featuring specific–concrete concepts. Among the Non-living entities, Furnishing, Vehicles, and Tools tend to appear in the quadrant featuring
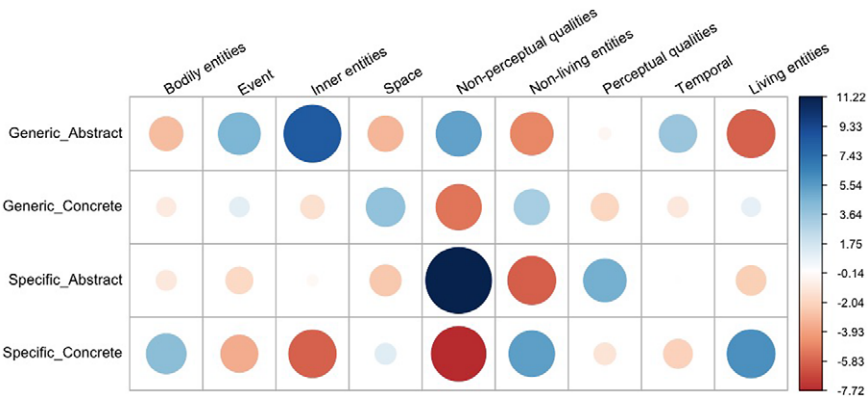

Figure 4. Correlation plot showing the standardized residuals for the chi-square analysis involving *Code 2* and the four quadrants.
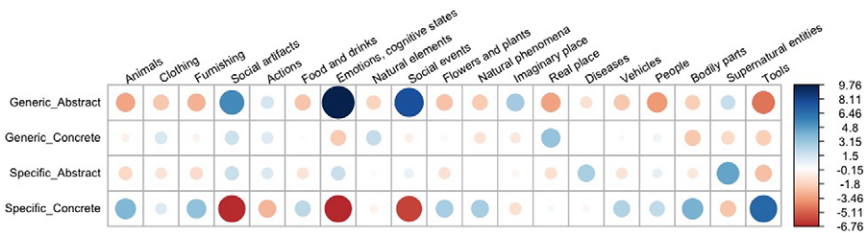

Figure 5. Correlation plot showing the standardized residuals for the chi-square analysis involving *Code 3* and the four quadrants.

specific–concrete concepts. Likewise, Bodily parts and products are often associated with specific–concrete concepts. Conversely, Supernatural entities and Diseases tend to appear in the quadrant featuring specific–abstract concepts, whereas Social artifacts are often associated with generic–abstract concepts. Among the different types of space, Imaginary places are associated with generic–abstract concepts, while Real places are associated with generic–concrete concepts. Finally, Social events, Emotions, and Cognitive states are strongly associated with generic–abstract concepts.

5.2.4. Id_code

This level of semantic annotation holds a strong significant relationship with the four quadrants ($\chi2(72) = 688.44$, p < .001, Cramer's V = .46). Inspection of the adjusted standardized residuals resembles the distribution illustrated in *Code 3*, with some minor differences, such as Social artifact, often associated with generic–abstract and generic–concrete concepts. Moreover, as shown in *Code 2*, Math–logic relations are often associated with general-abstract concepts, while Non-perceptual and Perceptual qualities are strongly associated with specific–abstract concepts.

5.3. Discussion

The results of Study 2 show that when looking at the four quadrants obtained by crossing specificity and concreteness ratings, the words within each quadrant tend to have different semantic contents.

First, as expected, we found that concrete words typically refer to physical entities and objects, while abstract words mainly denote events, social relations, and internal states (in line with Barsalou and Wiemer-Hastings, 2005). Among different kinds of concrete concepts, we further observed the well-established distinction between living versus non-living entities (e.g., FISH versus SHOES) and natural kinds versus artifacts (e.g., IRON versus KEY). Furthermore, we observed the distinction between emotions, mental states, quantitative concepts, and social artifacts among kinds of abstract concepts (e.g., JOY, INDIGNATION, PART, PRESTIGE). However, a fine-grained semantic analysis of words characterized by the intersection of their concreteness and specificity scores allowed us to frame concepts in a novel way.

When comparing the distribution of semantic types across the four quadrants, a major opposition emerges between generic–abstract concepts and specific–concrete concepts. On the one hand, Social artifacts, Emotions, Cognitive states, Math–logic relations, Non-perceptual qualities, and Social events have a strong positive association with generic–abstract concepts and a strong negative association with specific–concrete concepts. On the other hand, Living entities (i.e., Animals, Flowers and plants, People), Non-living entities (i.e., Clothing, Furnishing, Food and drinks, Vehicles, and Tools), and Bodily entities (i.e., Bodily parts and Diseases) are all positively associated with specific–concrete concepts and negatively associated with generic–abstract concepts. These results confirm the main differences between abstract and concrete domains but also suggest that these differences tend to emerge when their levels of specificity are taken into consideration.

Finally, within concrete concepts, the most specific are strongly associated with Living and Non-living entities, while the generic ones are only weakly or not at all linked to these semantic types. Rather, they tend to refer to Social artifacts and

Natural phenomena. Within abstract concepts, categories of Emotions, Cognitive states, Social artifacts, Math–logic relations, Social events, and Temporal relations are frequently associated with generic–abstract but not with specific–abstract concepts, which instead include Non-perceptual qualities and Supernatural entities. Notably, semantic types are distributed differently across concreteness and specificity axes: specific– concrete concepts seem to reflect classic taxonomic categories, whereas generic–concrete concepts describe wider social and natural categories. Likewise, specific–abstract concepts indicate precise religious or fictional entities and behavioral qualities, while generic–abstract concepts denote a vast range of internal and social phenomena, likely applied to wider contexts.

6. Study 3: What clusters emerge from the data when we analyze concepts considering their concreteness and specificity scores? What are the semantic types associated with each cluster? To what degree do the identified clusters align with the four quadrants?

6.1. Methods

To identify groups of datapoints with similar characteristics in terms of concreteness and specificity, we performed a cluster analysis. First, we fitted a Gaussian mixture model using the expectation–maximization (EM) algorithm, implemented via *Mclust* package [version 6.0.0] (Scrucca et al., 2016). The matrix of concreteness and specificity ratings (z-scored) was used as input for the analysis. Bayesian information criterion (BIC) scores were used to assess which number of clusters provided the optimal fit to the data. To enhance the interpretability of cluster analysis in the context of previous findings, we examined how the distribution of words within each cluster matched with a-priori determined quadrants. Additionally, we conducted a chi-square test of independence to explore the relationship between semantic types (*Id_code*) and words in the clusters. We finally calculated the strength of association using Cramer's V effect size coefficients (Kotrlik and Williams, 2003) and assessed the significance of association using adjusted standardized residuals, as in Study 2. All analyses and results are available in the OSF repository file 'Study3_cluster.html'.

6.2. Results

The model indicated that the data best support a VVI (i.e., diagonal, varying volume, and shape) model with four components (log-likelihood = -2687.21; Bayesian information criterion $BIC = -5506.577$; Integrated Complete-data Likelihood $ICL = -6007.393$). The clustering table indicated that the four components included 254, 290, 341, and 164 observations, respectively. This means that there is statistical evidence for four distinct subgroups of words that have highly related concreteness and specificity ratings. The complete list of the 1049 words distributed across the four clusters is stored in the OSF repository file 'data_cluster.csv'. Fig. 6 displays the 20 most representative words for each cluster (i.e., the ones that are the most certain for each cluster). The descriptive statistics of the clusters are reported in Table 3.

The chi-square test of independence showed a significant association between the semantic types and the four emerged clusters ($\chi2(72) = 932.53$, $p < .001$, Cramer's $V = .54$). Inspection of the adjusted standardized residuals indicates that words in different clusters tend to be associated with distinct semantic types. For each cluster,
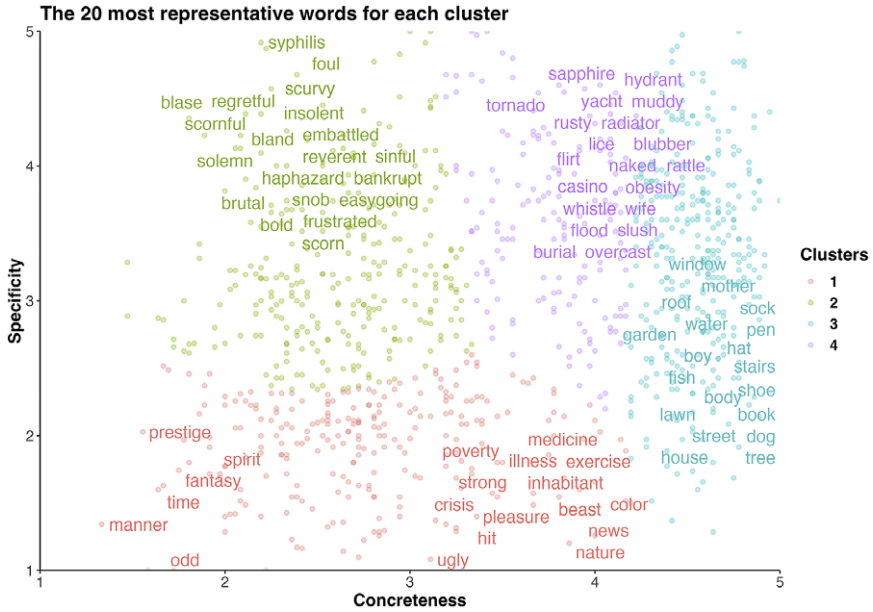
Figure 6. Scatterplot showing the clustering of our data. The 20 most representative words in each cluster (i.e., the ones that are the most certain for that cluster) are selected.

Table 3. Descriptive statistics of the clusters in terms of concreteness and specificity ratings

		Concreteness	Specificity
Cluster 1	Mean	2.78	1.92
	Max	4.19	2.60
	Min	1.33	1.00
	Median	2.74	2.00
	Std	0.60	0.36
Cluster 2	Mean	2.61	3.28
	Max	3.33	5.00
	Min	1.47	2.34
	Median	2.65	3.17
	Std	0.40	0.60
Cluster 3	Mean	4.54	3.32
	Max	5.00	5.00
	Min	4.17	1.29
	Median	4.56	3.37
	Std	0.19	0.75
Cluster 4	Mean	3.78	3.61
	Max	4.22	5.00
	Min	3.19	2.20
	Median	3.81	3.60
	Std	0.28	0.64

the frequency of the semantic types occurring therein is displayed in Fig. 7. Descriptive statistics of observed, expected, and standardized residuals are available in the OSF repository file 'Study3_chi-square_results.docx'.

Figure 7. Frequency of different semantic types occurring within each cluster.

The four clusters were then labeled, with the help of ChatGPT. We prompted ChatGPT multiple times, with the following string: 'Provide a label to the semantic category of this list of words…' The verbal labels indicated by ChatGPT served to summarize the essential characteristics of each group.

6.2.1. Cluster 1: Human experiences and emotions (N = 254 words)

The first cluster is characterized by the lowest scores in specificity ratings. It is composed of 166 generic–abstract and 88 generic–concrete concepts. This cluster includes a significant portion of words referring to Inner entities: Emotions,

Cognitive states (e.g., FANTASY, PLEASURE, FUN, IDEA, OPINION), Social artifacts (e.g., NEWS, PRESTIGE, POVERTY, MEDICINE, ART), Math−logic relations (e.g., MANNER, EXERCISE, UNIT, PHASE), and Social events (e.g., CRISIS, GAME, CHAOS, DISASTER).

6.2.2. Cluster 2: Attitudes and behaviors (N = 290 words)
The second cluster is characterized by medium ratings of concreteness and specificity. It comprises 97 generic–abstract, 19 generic–concrete, 135 specific–abstract, and 39 specific–concrete concepts. The words included in this cluster tend to be mainly associated with Non-perceptual qualities (e.g., BLASE, BRUTAL, EASYGOING, BOLD, INSOLENT, SHY) and Perceptual qualities (e.g., SOLEMN, BLAND, FOUL, CRUDE, QUIET, BRIGHT, RIGID).

6.2.3. Cluster 3: Common nouns (N = 341 words)
The third cluster is characterized by the highest concreteness ratings. It consisted of 111 generic–concrete and 230 specific–concrete concepts. This cluster includes words denoting Living and Non-living entities, mainly Tools (e.g., PEN, KEY, RUBBER, COPYBOOK, SPRAY, VASE, POT), followed by Animals (e.g., DOG, FISH, MONKEY, HORSE), Furnishings (e.g., STAIRS, WINDOW, ROOF, LAMP, DOOR), Flower and plants (e.g., TREE, PLANT, TRUNK, FLOWER, PINE, OLIVE TREE, OAK), Clothing (e.g., SHOE, HAT, SOCK, DRESS), and Vehicles (e.g., BOAT, MOTORCYCLE, SCOOTER, TRAIN). It also comprises words that denote Real places (e.g., GARDEN, LAWN, HOUSE, STREET, BAR), Bodily parts (e.g., BODY, FACE, HEAD, LEG), and Foods and drinks (e.g., PASTA, MILK, SALAD, CAKE).

6.2.4. Cluster 4: Natural world and negative things (N = 164 words)
The fourth cluster is characterized by high concreteness and high specificity ratings. It is composed of 36 generic–concrete and 128 specific–concrete concepts. The words included here are more likely to refer to Natural elements (e.g., SAPPHIRE, SLUSH, SLIME, GOLD, LAKE) and Natural phenomena (e.g., FLOOD, TORNADO, OVERCAST, HURRICANE, SUNRISE). This cluster also includes words of different semantic types, negatively connoted (e.g., Tools: REVOLVER, BULLET, GUN; Diseases: OBESITY, GANGRENE, INJURY; Real places: CASINO, SLUM, MORGUE; Social artifacts: BURIAL, FUNERAL).

6.3. Discussion
The results of Study 3 showed that clustering words according to their concreteness and specificity ratings led to the emergence of groups that have a fairly coherent semantic content.

The 'human experiences and emotions' cluster encompasses categories that describe both individual and collective experiences. The former includes feelings and emotional states, and bodily and sensory perceptions. The latter consisted of semantic categories associated with social interaction and interpersonal relationships, such as actions, events, and social and theoretical constructs.

The 'attitude and behaviors' cluster includes semantic categories of emotions, character traits, individual experiences, as well as qualities that describe human

behaviors. These categories cover a broad range of positive and negative terms that capture the diverse and complex nature of human beings.

The 'common nouns' cluster consisted of semantic categories of words that represent ordinary, tangible items, or entities that we encounter in our daily lives. These categories include objects, people, places, and animals that are familiar and commonly used in our language and conversations.

Lastly, the 'natural world and negative terms' cluster includes semantic categories that share a predominantly negative connotation. This cluster brings together words denoting a wide range of negative or unpleasant phenomena and objects, ranging from natural catastrophes to physical and mental diseases, from terms associated with discomfort and violence to social taboo and sexuality. In short, these semantic categories describe aspects of human suffering, danger, and social discomfort.

Although this analysis was conducted with a limited number of words, the identification of distinct semantic domains supports the idea that considering both word concreteness and specificity can offer valuable insights into the study of semantic concepts.

7. General discussion and conclusion

The main aim of this paper is to provide empirical evidence to support the claim that when investigating conceptual concreteness, attention should be paid to a neglected variable, namely categorical specificity. As anticipated in the introduction of this work, specificity is a variable that characterizes both concrete and abstract concepts alike. A word is highly specific when it specifies a concept in detail, zooming in on the semantics of precise referents or situations. In this sense, a word that scores high in specificity is a word that pinpoints a high-resolution semantic content. Conversely, a word that scores low in specificity is a word that denotes a generic conceptual category. Conceptual categories are generic when they include a variety of subcategories, and therefore, they designate many different concepts, and they can be used to refer to many different entities and situations. These categories are therefore highly inclusive and are characterized by low-resolution semantics. The categories MAMMAL, ART, and ACTIVATE refer to many concepts and activities, while MUFFIN, OIL PAINTING, and STABBING are more specific because they refer to particular types of objects or actions.

Previous research has shown that there is a positive, though mild, correlation between specificity and concreteness, suggesting that words that are perceived to describe more concrete concepts are also perceived to be more specific and vice versa. In the first study hereby presented, we ran a series of regressions to investigate to what extent the semantic content of a concept explains its perceived concreteness and how does such relation change when specificity is considered. We reported that overall, as expected, semantic content is a good predictor of conceptual concreteness, especially when such content is described in a fine-grained manner, by means of high-resolution semantic characterizations. The information about categorical specificity added to the information about the semantic content of a concept improves the power of such predictions. In other words, when we have not only information about the semantics of a concept but also information about the level of specificity/ generality at which such a concept is expressed, we can better predict its concreteness. Information about categorical specificity seems to improve the prediction of

conceptual concreteness particularly when the semantic content is described coarsely. For instance, when we only know that AMBULANCE and WEAPON are Entities, rather than Qualities or Spatio-temporal relations, and we do not have more precise information about the semantics of these concepts, we can predict that they are concrete concepts. But knowing that they score, respectively, high in specificity (AMBULANCE) and low in specificity (WEAPON) helps us make a better prediction on their degree of concreteness, which in this case is quite high. Similarly, if we have limited information about the semantic content of AGONY and HABIT, knowing their level of specificity helps us predict how concrete these concepts are, and in this case, they are characterized by low concreteness because they are both abstract concepts.

The positive but low correlation observed by Bolognesi and Caselli (2023) between concreteness and specificity suggested that there would be not only concrete+specific and abstract+generic concepts but also concrete+generic and abstract+specific ones. Building on this preliminary study, we cut the distribution of concreteness and specificity scores into four quadrants, upon testing the two individual distributions for bimodality. The (in total) four humps obtained corresponded to the four types of concepts hypothesized: concrete+generic, concrete+specific, abstract+generic, and abstract+specific. Each type is typically characterized by a different semantic content. However, the strongest distinction seems to characterize the different semantics between the quadrants generic–abstract and specific–concrete. Here, we observed Social artifacts, Emotions, Cognitive states, Math−logic relations, Non-perceptual qualities, and Social events to have a strong positive association with generic–abstract concepts, while Living entities (i.e., Animals, Flowers and plants, People), Non-living entities (i.e., Clothing, Furnishing, Food and drinks, Vehicles, and Tools), and Bodily entities (i.e., Body parts and Diseases) are all positively associated with specific–concrete concepts. As mentioned above, these results support the main differences between abstract and concrete domains, but crucially suggest that these differences tend to emerge when specificity is taken into consideration.

Finally, we reversed the approach adopted in the second study, from a top-down approach to a bottom-up approach: we explored the semantics of the clusters emerging from the aggregation of concepts based on their concreteness and specificity scores. In other words, we momentarily put aside our own semantic classification and generated automatically clusters of concepts based on concreteness and specificity. Then, we interpreted the content of these clusters with the aid of our semantic classification scheme. We reported the emergence of four clusters that feature relatively different semantic content: a cluster including concepts related to human experiences and emotions; a cluster including concepts describing human attitudes and behaviors, represented mainly by non-perceptual and perceptual qualities; a cluster including concepts that define common nouns and therefore tools, places, furniture, animals, and so forth; and a cluster that includes natural elements and phenomena, including many concepts that seem to have a particularly negative valence. This last explorative analysis provides further evidence that adds to the evidence reported in the second study, in support of the idea that when we cross specificity and concreteness, we obtain groups of concepts characterized by different semantics.

In line with recent findings, we acknowledge that the organization of conceptual knowledge is neither centered around the classical dichotomy between concrete and abstract concepts nor distributed along with a linear continuum of concreteness

(Banks et al., 2023; Barsalou et al., 2018). Rather, concepts are better represented as points in a multidimensional semantic space with various dimensions assuming different weights. Recent studies analyzing the distribution of words in a large semantic space using ratings across multiple semantic dimensions (e.g., Hollis & Westbury, 2016; Troche et al., 2017; Villani et al., 2019) report that the concrete-abstract distinction becomes less clear and different subgroups of concepts emerge based on other latent factors. We contribute to this discussion by investigating the variable Specificity.

Overall, the three studies contribute to deepening our understanding of how abstract, concrete, generic, and specific concepts compare with one another in terms of their semantic content. The studies proposed are based on a dataset of roughly one thousand words in Italian, for which concreteness and specificity ratings collected from humans were already available, and semantic annotations are provided as an additional resource, thanks to this paper. There are some limitations to the current study, namely the relatively small size of the dataset and the fact that it includes only Italian words. Yet, we believe that these insights and the additional resource hereby provided (the semantic annotation of these words) may prompt new research pathways, with the general goal of understanding in greater detail how to approach the intuitive distinction between concrete and abstract concepts. One of the pathways could be that of understanding how different semantic types span along the generic-specific continuum. In the Introduction to this paper, as a matter of fact, we suggested that for some types of concepts, it may be easier to observe a variation in specificity than for others. We compared concepts belonging to the animal kingdom (ANIMAL, MAMMAL, DOG, and DALMATIAN) to abstract concepts that do not seem to allow for long taxonomic ladders (ESSENCE). We suggested that some concepts may allow us to develop long ladders of taxonomic relationships, from highly generic to highly specific, while others may develop extremely short ladders that remain skewed toward the highly generic or highly specific poles. To visualize this idea, refer to Fig. 8, which illustrates it with three semantic categories used in our coding scheme. As the figure shows, Social events are a semantic type that spans over all levels of specificity and encompasses both highly specific concepts (e.g., BANKRUPTCY, BLACKMAIL, UPRISING) and highly generic ones (e.g., AGREEMENT, DAMAGE, DANGER). Other semantic types, instead, do not have such a wide range of options, from highly specific to highly generic. For instance, Math−logic relations appear to include only concepts that are quite generic (e.g., CAUSE, ELEMENT, METHOD, PHASE), while Supernatural entities include only concepts that are quite specific (e.g., VAMPIRE, DEVIL, ANGEL).

This finding suggests that (the Italian) language gives us words to talk about concepts with different degrees of granularity (specificity) and that in different semantic fields, for different semantic types, the words we have vary. For some semantic fields, we have more levels of granularity, and for others, we have less. The more granularity there is, the better we can effectively discuss that semantic field, with representations being more or less detailed depending on the context.

The idea that different types of concepts may cover different steps on the ladder of abstraction could be investigated in further studies in a different way. For instance, they could be investigated through language production tasks in which word ladders are produced by speakers, for both concrete and abstract concepts. Based on speakers' productions, it should be possible to observe potential differences in how the
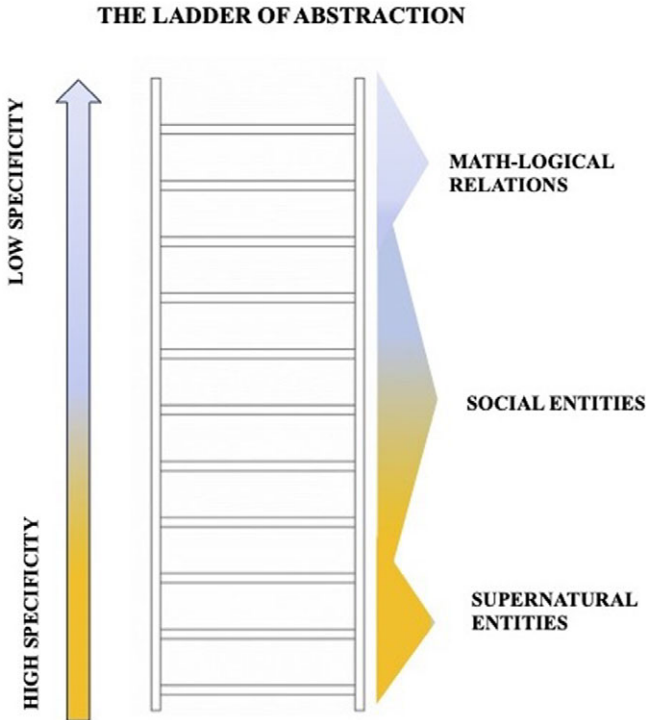
THE LADDER OF ABSTRACTION

Figure 8. Visualization of the abstraction ladder, showing how different semantic types span across the ladder in different ways.

taxonomic relationship of category inclusion, hereby labeled as specificity, characterizes different types of concepts, such as concrete and abstract concepts.

To conclude, we renew our warning to colleagues and researchers working on the general phenomenon of abstraction and on the difference between concrete and abstract concepts in particular: when working on conceptual abstraction, it is important that the very notion of 'abstraction' is preliminarily clarified, in terms of the two dimensions that often are conflated therein: concreteness and specificity. And when working on the difference between concrete and abstract concepts, aiming for instance at capturing concreteness effects (and therefore processing advantages of concrete over abstract words), it is important to control for specificity because this variable may otherwise work as a confound, hence leading to experimental effects that may be erroneously attributed to concreteness alone.

dataset separately, solving the ambiguities together with M.B. C.V. performed the statistical analyses and data visualization. C.V. and M.B. wrote the main manuscript text. All authors reviewed and approved the final version of the manuscript before submission. For the specific concerns of the Italian academic attribution system, C.V. is responsible for sections Content Analysis, Study 1, Study 2, and Study 3; M.B. is responsible for sections Introduction, Theoretical Background, and General Discussion.

References

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments & Computers*, 31(4), 578–602. https://doi.org/10.3758/bf03200738

Banks, B., Borghi, A. M., Fargier, R., Fini, C., Jonauskaite, D., Mazzuca, C., Montalti, M., Villani, C., & Woodin, G. (2023). Consensus paper: Current perspectives on abstract concepts and future research directions. *Journal of Cognition*, 6(1):62. https://doi.org/10.5334/joc.238

Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, 125(1), 47–53. https://doi.org/10.1016/j.bandl.2013.01.005

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., & Mazzoleni, M. (2004). Introducing the La Repubblica corpus: A large, annotated, TEI (XML)-compliant corpus of newspaper Italian. In *Proceedings of the 4th international conference on language resources and evaluation LREC*, ELRA European Language Resources Association, Parigi: 1771–1774.

Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170144.

Barsalou, L., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 129–163). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.007

Binder, J. R., Westbury, C. F., Mckiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract words. *Journal of Cognitive Neuroscience*, 17, 905–917. https://doi.org/10.1162/0898929054021102

Bolognesi, M., Burgers, C., & Caselli, T. (2020). On abstraction: Decoupling conceptual concreteness and categorical specificity. *Cognitive Processing*, 21(3), 365–381.

Bolognesi, M. M., & Caselli, T. (2023). Specificity ratings for Italian data. *Behavior Research Methods*, 55(7), 3531–3548.

Borghi, A. M., Barca, L., Binkofski, F., & Tummolini, L. (2018a). Abstract concepts, language and sociality: From acquisition to inner speech. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 373(1752), 20170134. https://doi.org/10.1098/rstb.2017.0134

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

Borghi, A. M. (2022). Concepts for which we need others more: The case of abstract concepts. *Current Directions in Psychological Science*, 31(3), 238–246.

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts* (Vol. 2). Springer.

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings (Vol. 30, No. 1, 25–36). Technical Report C-1, the Center for Research in Psychophysiology, University of Florida.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness rating for 40 thousand generally known English words lemmas. *Behavior Research Methods*, 46(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003) What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, 20(3), 213–261. https://doi.org/10.1080/02643290244000266

Chen, J., Papies, E. K., & Barsalou, L. W. (2016). A core eating network and its modulations underlie diverse eating phenomena. *Brain and Cognition*, 110, 20–42. https://doi.org/10.1016/j.bandc.2016.04.004

Conca, F., Borsa, V. M., Cappa, S. F., & Catricalà, E. (2021). The multidimensionality of abstract concepts: A systematic review. *Neuroscience & Biobehavioral Reviews*, 127, 474–491. https://doi.org/10.1016/j.neubiorev.2021.05.004

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

Davis, C. P., Altmann, G. T. M., & Yee, E. (2020). Situational systematicity: A role for schema in understanding the differences between abstract and concrete concepts. *Cognitive Neuropsychology*, 37 (1–2), 142–153. https://doi.org/10.1080/02643294.2019.1710124

Davis, C. P., & Yee, E. (2023). Is time an embodied property of concepts? *Plos One*, 18(9), e0290997.

Desai, R. H., Reilly, M., & van Dam, W. (2018). The multifaceted abstract brain. *Philosophical Transactions of the Royal Society B* 373, 20170122. https://doi.org/10.1098/rstb.2017.0122

Diveica, V., Pexman, P. M., & Binney, R. J. (2023). Quantifying social semantics: An inclusive definition of socialness and ratings for 8388 English words. *Behavior Research Methods*, 55(2), 461–473.

Fini, C., & Borghi, A. M. (2019). Sociality to reach objects and to catch meaning. *Frontiers in Psychology*, 10, 838.

Gainotti, G., Ciaraffa, F., Silveri, M.C., & Marra, C. (2009). Mental representation of normal subjects about the sources of knowledge in different semantic categories and unique entities. *Neuropsychology*, 23(6), 803–812. https://doi.org/10.1037/a0016352

Ghio, M., Vaghi, M. M. S., & Tettamanti, M. (2013). Fine-grained semantic categorization across the abstract and concrete domains. *PLoS ONE*, 8(6):e67090. https://doi.org/10.1371/journal.pone.0067090.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Glenberg, A. M., & Gallese, V. (2012). Action based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48, 905–922. https://doi.org/10.1016/j.cortex.2011.04.010

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301–307. https://doi.org/10.1016/s0896-6273(03)00838-9

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89. https://doi.org/10.1080/19312450709336664

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 1(2), 130–136. https://doi.org/10.1037/0096-1523.1.2.130

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. The MIT Press.

Kotrlik, J. W., & Williams, H. A. (2003). The incorporation of effect size in information technology, learning, and performance research. *Information Technology, Learning, and Performance Journal*, 21, 1–7.

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. https://doi.org/10.1037/a0021446

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(1), 92–107. https://doi.org/10.1037/0278-7393.12.1.92

Langland-Hassan, P., & Davis, C. P. (2023). A context-sensitive and non-linguistic approach to abstract concepts. *Philosophical Transactions of the Royal Society B*, 378(1870), 20210355.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, 52, 1271–1291.

Löhr, G. (2023). Does the mind care about whether a word is abstract or concrete? Why concreteness is probably not a natural kind. *Mind and Language*. Advance online publication. https://doi.org/10.1111/mila.12473

Mangiafico, S. (2021). rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.4.0. https://CRAN.R-project.org/package=rcompanion

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, 58, 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

McRae, K., & Cree, G. S. (2002). Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category-specificity in brain and mind* (pp. 211–249). Psychology Press.

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, 547–559. https://doi.org/10.3758/BF03192726

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). Semantic memory: A feature-based analysis and new norms for Italian. *Behavior Research Methods*, 45(2), 440–461. https://doi.org/10.3758/s13428-012-0263-4

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the Affective Norms for English Words (ANEW) for Italian. *Behavior Research Methods*, 46(3), 887–903. https://doi.org/10.3758/s13428-013-0405-3

Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart & Winston

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

Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, 21(2), e12549. https://doi.org/10.1111/desc.12549

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reilly, J., Diaz, M., Pylkkänen, L., Jefferies, E., Poeppel, D., Zubicaray, G., Connell, L., Mirman, D., Sandberg, C., Brysbaert, M., Borghesani, V., Kemmerer, D., Cantlon, J., Hickock, G., Mahon, B., Vigliocco, G., Bi, Y., Hartwigsen, G., Peelle, J. E., …, Rodd, J. (under review). What we mean when we say semantic: A consensus statement on the nomenclature of semantic memory. PsyArXiv preprint: https://osf.io/preprints/psyarxiv/xrnb2

Roversi, C., Borghi, A. M., & Tummolini, L. (2013). A marriage is an artefact and not a walk that we take together: An experimental study on the categorization of artefacts. *Review of Philosophy and Psychology*, 4 (3), 527–542.

Rumiati, R. I, & Foroni, F. (2016). We are what we eat: How food is represented in our mind/brain. *Psychonomic Bulletin and Review*, 23(4), 1043–1054. https://doi.org/10.3758/s13423-015-0908-2

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27(5), 499–520. https://doi.org/10.1016/0749-596X(88)90022-8

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, 24(1), 114–126. https://doi.org/10.2307/748013

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In *The psychology of word meanings*. Psychology Press.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1), 205–233. https://doi.org/10.32614/RJ-2016-021

Setti, A., & Caramelli, N. (2005). Different domains in abstract concepts. In B. Bara, B. Barsalau & M. Bucciarelli (Eds.), *Proceedings of the XXVII annual conference of the cognitive science, 1997–2002*. Erlbaum.

Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arppe, A., …, Wilkins-Diehr, N. (2020). DescTools: Tools for descriptive statistics. R package version 0.99.41. https://CRAN.R-project.org/package=DescTools

Statisticat, LLC (2021). LaplacesDemon: Complete Environment for Bayesian Inference Bayesian-Inference. com. R package version 16.1.6. https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software

Strik Lievers, F., Bolognesi, M., & Winter, B. (2021). The linguistic dimensions of concrete and abstract concepts: Lexical category, morphological structure, countability, and etymology. *Cognitive Linguistics*, 32 (4), 641–670. https://doi.org/10.1515/cog-2021-0007.

Tillotson, S. M., Siakaluk, P. D., & Pexman, P. M. (2008). Body—object interaction ratings for 1,618 monosyllabic nouns. *Behavior Research Methods*, 40(4), 1075–1078.

Troche, J., Crutch, S. J., & Reilly, J. (2017). Defining a conceptual topography of word concreteness: Clustering properties of emotion, sensation, and magnitude among 750 English words. *Frontiers in Psychology*, 8, 1787.

Van Loon-Vervoorn, W. A. (1984). Afbeeldbaarheidsnormen van 1087 Nederlandse woorden [Imageability norms for 1,087 Dutch words]. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, 39(8), 462–470.

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. https://doi.org/10.1016/j.cogpsych.2003.09.001

Villani, C. (2018). L'Embodied Cognition e la sfida dei concetti astratti. Un approccio multidimensionale. *Rivista internazionale di Filosofia e Psicologia*, 9(3), 239–253.

Villani, C., Lugli, L., Liuzza, M. T., & Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition*, 11(3), 403–430.

Villani, C., Lugli, L., Liuzza, M. T., Nicoletti, R., & Borghi, A. M. (2021a). Sensorimotor and interoceptive dimensions in concrete and abstract concepts. *Journal of Memory and Language*, 116, 104173.

Villani, C., D'Ascenzo, S., Borghi, A. M., Roversi, C., Benassi, M., & Lugli, L. (2021b). *Is justice grounded? How expertise shapes conceptual representation of institutional concepts* (pp. 1–17). Psychological Research.

Wauters, L. N., Telling, A. E. J. M., Van Bon, W. H. J., & Van Haaften, A. W. (2003). Mode of acquisition of word meanings: The viability of a theoretical construct. *Applied Psycholinguistics*, 24(3), 385–406. https://doi.org/10.1017/S0142716403000201

Warrington, E.K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107(Pt 3), 829–854. https://doi.org/10.1093/brain/107.3.829

Westbury, C., & Wurm, L. H. (2022). Is it you you're looking for? Personal relevance as a principal component of semantics. *The Mental Lexicon*, 17(1), 1–33.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica (Amsterdam)*, 132(2), 173–189. https://doi.org/10.1016/j.actpsy.2009.02.002

Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2021). Movement tracking of cognitive processes: A tutorial using mousetrap. *PsyArXiv*, 10. https://doi.org/10.31234/osf.io/v685r