



OPEN ACCESS

EDITED BY

Lara L. Jones,
Wayne State University, United States

REVIEWED BY

Charles P. Davis,
Duke University, United States
Anna M. Borghi,
Sapienza University of Rome, Italy

*CORRESPONDENCE

Andrew S. Persichetti
✉ persichettias@nih.gov

RECEIVED 16 August 2023

ACCEPTED 08 December 2023

PUBLISHED 04 January 2024

CITATION

Persichetti AS, Shao J, Denning JM,
Gotts SJ and Martin A (2024) Taxonomic
structure in a set of abstract concepts.
Front. Psychol. 14:1278744.
doi: 10.3389/fpsyg.2023.1278744

COPYRIGHT

© 2024 Persichetti, Shao, Denning, Gotts and Martin. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Taxonomic structure in a set of abstract concepts

Andrew S. Persichetti^{1*}, Jiayu Shao¹, Joseph M. Denning²,
Stephen J. Gotts¹ and Alex Martin¹

¹Section on Cognitive Neuropsychology, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, United States, ²Department of Psychology, University of California Los Angeles, Los Angeles, CA, United States

A large portion of human knowledge comprises “abstract” concepts that lack readily perceivable properties (e.g., “love” and “justice”). Since abstract concepts lack such properties, they have historically been treated as an undifferentiated category of knowledge in the psychology and neuropsychology literatures. More recently, the categorical structure of abstract concepts is often explored using paradigms that ask participants to make explicit judgments about a set of concepts along dimensions that are predetermined by the experimenter. Such methods require the experimenter to select dimensions that are relevant to the concepts and further that people make explicit judgments that accurately reflect their mental representations. We bypassed these requirements by collecting two large sets of non-verbal and implicit judgments about which dimensions are relevant to the similarity between pairs of 50 abstract nouns to determine the representational space of the concepts. We then identified categories within the representational space using a clustering procedure that required categories to replicate across two independent data sets. In a separate experiment, we used automatic semantic priming to further validate the categories and to show that they are an improvement over categories that were defined within the same set of abstract concepts using explicit ratings along predetermined dimensions. These results demonstrate that abstract concepts can be characterized beyond their negative relation to concrete concepts and that categories of abstract concepts can be defined without using *a priori* dimensions for the concepts or explicit judgments from participants.

KEYWORDS

categories, abstract words, semantic knowledge, concepts, automatic semantic priming

Introduction

Storing knowledge about the world as concepts and organizing those concepts into categories is a core feature of human cognition (Rosch, 1973, 1975). Categorization allows us to generalize knowledge across different experiences and to retain seemingly infinite facts about things in the world. A common framework for studying how categories are represented in the mind is to assume that they are comprised of concrete (e.g., “dog,” “hammer,” “carrot”) and abstract (e.g., “love,” “justice,” “intelligence”) concepts (e.g., Schwanenflugel and Shoben, 1983; Paivio, 1991; but also see, Troche et al., 2017; Löhr, 2023). Whether concrete and abstract concepts form a dichotomy or are represented along

a unified and continuous space, there is a wealth of evidence that concrete concepts are acquired and processed more easily than abstract concepts – the so-called “concreteness effect” (Gilhooly and Logie, 1980; Bleasdale, 1987; Schwaneflugel et al., 1988; Walker and Hulme, 1999; Allen and Hulme, 2006). Furthermore, prior behavioral studies have demonstrated that concrete categories like “animals,” “tools,” and “food” are represented based on shared physical and functional properties among the category members (Cree and McRae, 2003) – i.e., properties that can be easily referenced by ostension. Since concrete concepts share perceivable and functional properties that can be easily referenced, their category structure is often intuitive and apparent. By contrast, it is thought that the category structure of abstract concepts is not so intuitive because, by definition, they lack readily perceivable properties (Paivio, 1991; Borghi et al., 2017; Desai et al., 2018) and therefore the domain of abstract concepts has often been treated as an undifferentiated category in the both the psychology and neuropsychology literatures (Warrington, 1975; Schwaneflugel and Shoben, 1983; Riddoch et al., 1988; Paivio, 1991; Plaut and Shallice, 1993; Breedin et al., 1994; Franklin et al., 1995).

More recently, many experiments have mapped the representational space of abstract concepts using methods that rely on participants making explicit judgments about concept properties and category membership based on featural dimensions and categories that are predetermined by the experimenter (e.g., McRae et al., 2005; Crutch et al., 2013; Devereux et al., 2014; Villani et al., 2019; Lynott et al., 2020; Banks and Connell, 2023). However, such experimental paradigms might not be well-suited to mapping the representational space of abstract concepts because the featural dimensions of the space and the categories within it may not be known *a priori* or might be hard to articulate. In addition, these experiments ask people to make explicit verbal judgments about concepts that accurately reflect their mental representations. Prior findings suggest that these requirements might introduce unnecessary bias into the experiment. For example, when asked to freely list the semantic features of concepts, people generate only a small subset of relevant features (Plaut and Shallice, 1993; McRae et al., 2005). On the other hand, experiments that ask participants to rate concepts along featural dimensions might not be adequate for mapping the representational space of abstract concepts because the dimensions will be limited in number and thus biased by the experimenter’s choice of which dimensions to include in the experiment. The goal of the current study is to use an analytic approach that does not rely on *a priori* featural dimensions or explicit judgments from participants to identify behaviorally relevant categories in a set of abstract words. To do so, required us to choose a relatively small set of 50 abstract nouns so we could exhaustively sample the representational space of the concepts with tens of thousands of similarity judgments using an odd-one-out triad task (Experiment 1) and subsequently collect over a thousand trials during an automatic semantic priming in each participant (Experiment 2).

In Experiment 1, we used an odd-one-out similarity task, in which participants chose which of three abstract nouns was least like the other two across many trials, to determine the representational space of 50 abstract nouns (Hebart et al., 2020). Using this approach provides a non-verbal and implicit judgment about which dimensions are most relevant to the similarity between pairs of words across varying contexts provided by the third word in each trial, and thus makes it possible to capture data-driven dimensions that might

otherwise be missed. We then used the InfoMap clustering algorithm to define categories within the representational space (Rosvall and Bergstrom, 2008, 2011). In Experiment 2, we used automatic semantic priming to confirm that the categories found in Experiment 1 reflect intrinsic semantic categories rather than deliberative *ad hoc* categories (Barsalou, 1983) and to test whether reshuffling the words into categories based on explicit ratings along predetermined dimensions would also produce priming (Neely, 1977, 1991; Sperber et al., 1979; Lucas, 2000; McNamara, 2005).

Experiment 1

In the first experiment, we collected two sets of similarity judgments, each comprising tens of thousands of implicit judgments of the similarity between 50 abstract nouns on Amazon’s Mechanical Turk (M Turk). We then constructed two independent similarity matrices from those data and searched for clusters that replicated across them. The goal was to identify stable categories among the set of abstract words.

Methods

Participants

Eleven hundred and four individuals (mean (SD) age, 39.8 (11.9); 436 females) participated in the experiment on M Turk. Participants were required to be high school graduates, native English speakers, and live within the United States. These participants gave informed consent under the National Institutes of Health (NIH) Institutional Review Board-approved protocol number 000589 and were compensated financially for their time.

Experimental design

We created a set of 50 abstract nouns by randomly choosing common English words that were at least one standard deviation below the mean concreteness rating in the Abstract Conceptual Feature (ACF) database (Crutch et al., 2013; Troche et al., 2017). There were 169 words that were at least one standard deviation below the mean concreteness rating. Therefore, our method of randomly selecting words was to list and number those words and then randomly draw 50 numbers from the range of 1–169 using the *randperm* function in Matlab. We then posted an odd-one-out triad task on M Turk to obtain similarity judgements for these 50 abstract nouns. On each trial of the experiment, three words were presented in random order in a horizontal row under the instruction to “click on the odd-one out.” The participants used a computer mouse to click on the chosen word, which was followed by a blank screen for 500 milliseconds (ms), then the next trial. Participants were given as much time as needed to make their choice. The set of all possible triads (19,600 trials) was randomly ordered and divided into 980 unique blocks, each with 20 triad trials. To ensure that participants were attentive during each block, we included four catch trials in which participants saw the words “PLUS,” “MINUS,” and “EQUAL” randomly ordered in a horizontal line under the instruction to “click on EQUAL.”

Thus, there were 24 trials in each block. Individuals were allowed to participate in a maximum of five blocks. Before starting the experiment, participants were presented with instructions and an example triad that included words that were not used in the experiment. After collecting the full set of triads, we ran the same experiment again to obtain a second set of similarity judgements from a separate group of participants. Before posting the second set of triads, we randomized the order of triads before dividing them into blocks and excluded individuals that participated in the first posting. We excluded blocks in which the participant either (1) failed to supply the correct answer on a catch trial; (2) had an average trial response time of less than 500 ms; (3) or selected the same response on every trial. In the event of a rejected block, the individual was excluded from further participation in the experiment and the block was posted again. A total of 276 subjects were eliminated based on the criteria described above. In total, we used the data from 828 participants to collect two full sets of 19,600 trials each.

Data analysis

Based on the results of the odd-one-out triad task, we created two separate 50×50 similarity matrices (one for each full set of triads) by calculating the proportion of trials in which a given word pair was kept together across all triads containing that word pair, then calculating Euclidean distances between the words. The independent similarity matrices were strongly correlated ($r=0.91$). We then clustered the abstract nouns using the similarity matrices in the following way. First, the real-valued similarity matrices were thresholded into binary (0 or 1) undirected matrices at a range of threshold values (top %: 50, 60, 70, 80, 85, 90, 95, 99, 99.5). The thresholded matrices were then clustered using the InfoMap algorithm to form optimal two-level partitions (i.e., the optimal solution found on 100 searches). Across the range of thresholds, we found optimal agreement of number of clusters found in the two matrices and the percentage of words clustered across the matrices at the 90% threshold. We then constrained the clusters that replicated across the two matrices to contain at least four words to ensure that each cluster was stable and potentially interpretable. Since the two similarity matrices were strongly correlated, we averaged them together before performing hierarchical clustering for greater precision and stability.

Results

The clustering procedure separated 39 abstract nouns into five replicable clusters that each included at least four words, while the remaining 11 words were left unsorted based on the conservative inclusion criteria described above (Figure 1A). Our procedure clustered the abstract nouns into categories that seem to capture meaningful dimensions of the stimuli. To generate category labels, we asked OpenAI's text-davinci-003 GPT3 model (Brown et al., 2020) to list properties that are common across the words from each category [The exact prompt was "List five common properties of the following words: (list of words from the category)]." The top three labels for each category (and the cluster unsorted words) are listed at the bottom of Figure 1A. Based on these labels, the categories can be described as relating to positive emotions, negative emotions,

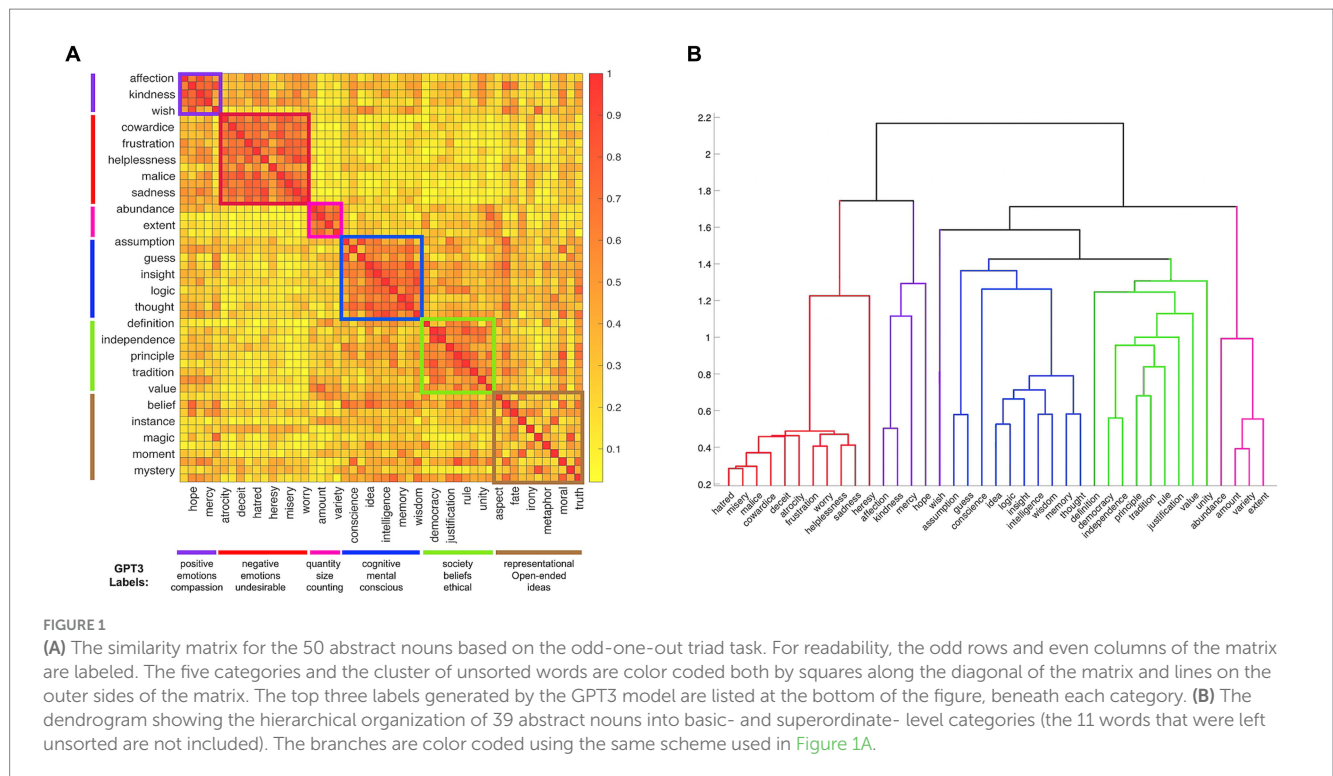
quantities, cognitive terms, and social terms, respectively. By contrast, the cluster of unsorted words was labeled with vague terms, such as "representational" and "open-ended."

Next, we asked whether the 39 abstract nouns that were clustered into five categories were further nested in a hierarchy comparable to the well-established organization of concrete nouns into subordinate, basic, and superordinate levels, Rosch (1973). A hierarchical clustering analysis based on the Euclidean distances between the categorized words indeed resulted in a three-level dendrogram with the individual words on the bottom; the words grouped into five categories at the middle level (positive and negative emotions, respectively, cognitive, social, and quantitative categories); and three groupings at the upper level. At the upper level of the hierarchy, one branch was affect-related, grouping the positive and negative emotions together, another branch grouped the cognition-related category with the social category, and the third branch was comprised the quantity-related category only (Figure 1B). Thus, like the category structure of concrete nouns, the category structure in our set of abstract nouns is hierarchical.

While our method uncovers a category structure within the space of abstract nouns that was replicated across two data sets, we wanted to compare it to more commonly used methods of investigating the representational space of abstract nouns. To do so, we re-clustered the data using k-means to cluster the 50 abstract nouns based on Euclidean distances derived from explicit feature ratings from the ACF database. We found that the $k=5$ cluster solution explained ~80% of the variance in these ratings. However, the clusters do not seem to have a reasonable organizing principle (Supplementary Figure 1). When we asked the GPT3 model to list properties that are common across the words from each of the five clusters, the labels were vague, often referring to the fact that each cluster contains abstract nouns from the English language (Supplementary Figure 1). Taken together, our method of collecting non-verbal and implicit judgments about the dimensional space of abstract words and using a rigorous clustering method revealed a replicable and intuitive category structure within our set of abstract nouns. Furthermore, our method seems to be an improvement over commonly used methods of characterizing the category structure of concepts based on participants making explicit judgments about concept properties along featural dimensions that are predetermined by the experimenter.

Experiment 2

The results from Experiment 1 demonstrated that using our method of collecting similarity judgments with a triad task to identify categories in a set of abstract nouns provides more reasonable categories compared to commonly used methods that collect explicit judgments about *a priori* semantic features of words. However, since there was no time limit for choosing the odd-one-out of a triad of abstract words, it could be argued that our triad task elicited a controlled and strategic process, as opposed to being implicit and automatic in nature. As a result, the triad task may have produced arbitrary categories rather than the intrinsic semantic categories that we intended to uncover (Barsalou, 1983). To address this potential difficulty, we sought to demonstrate that the five categories identified in Experiment 1 could be activated automatically by using an automatic semantic priming paradigm.



The assumption underlying automatic semantic priming is that the presentation of a word automatically activates its semantic representation in long-term memory and this activation partially includes semantically related concepts due to overlap of shared features, thus increasing the accessibility of those related concepts. Hence, when individuals are required to make a semantic judgment about a word (the probe), they are faster to respond when the probe word is immediately preceded by a semantically related word (the semantic prime) relative to when it is preceded by a semantically unrelated word. Critically, it has been shown that if the words are presented with a very brief time between them (i.e., a stimulus onset of asynchrony – SOA – of 250 ms), then priming reflects automatic activation of semantically related concepts from the same category (e.g., “bulldog” and “poodle”), instead of controlled and conscious associations between related concepts from different categories (e.g., “bulldog” and “leash”) that can occur with longer SOAs (Neely, 1977; Lucas, 2000). Thus, we used a 250 ms SOA to test the intrinsic nature of the categories found in Experiment 1. If this was indeed the case, then participants should be significantly faster to respond to words when they are preceded by words from the same category relative to when they are preceded by words from a different category.

Methods

Participants

Twenty-five individuals (mean (SD) age, 27.5 (8.9); 15 females) participated in the experiment in-person at the NIH. These participants gave informed consent under a National Institutes of Health Institutional Review Board-approved protocol (93-M-0170, clinical trial #NCT00001360). An additional 108 individuals (mean

(SD) age, 43.1 (9.7); 43 females) participated on M Turk. These participants gave informed consent under the National Institutes of Health Institutional Review Board-approved protocol number 000589. All participants were high school graduates, native English speakers, lived within the United States, and were compensated financially for their time.

Experimental design

Prior to generating the word pairs used in the experiment, we chose to exclude nine of the categorized abstract words so that we could present enough trials for evaluating priming, without overtaxing the participants. Thus, we selected 30 words from the categories identified in Experiment 1 to use as stimuli. At least four words were selected from each of the five categories (Figure 2). No other criteria were used to decide which words were excluded. We also selected 30 common concrete nouns from five categories (fruit, four-legged animals, dwellings, tools, and birds) included in the Connecticut Category Norms database (Battig and Montague, 1969) as a control condition (Supplementary Figure 2). We chose these five categories because they should be common to English speakers and they each had enough words that were frequently listed by respondents in the Connecticut Category Norms database so that we could exclude words that were ambiguous or polysemous – e.g., we excluded the word “orange” from the fruit category because it is also commonly used to refer to a color. We compared the magnitude of the automatic semantic priming effects elicited by these concrete words with the priming effects found across the abstract categories.

Semantic priming was tested using a 250 ms SOA. During each trial, a prime word was presented for 100 ms, followed by a 50 ms

affection	atrociti	abundance	conscience	definition
hope	cowardice	amount	idea	democracy
kindness	deceit	extent	insight	independence
mercy	frustration	variety	intelligence	rule
wish	hatred		memory	tradition
	helplessness		wisdom	unity
	malice			
	sadness			
	worry			

FIGURE 2

The 30 abstract words used in the automatic semantic priming experiment, separated (and color coded) by the categories identified in Experiment 1.

mask, and 100 ms blank screen, then a probe word for 250 ms, and finally a blank screen for 1,000 ms. The participants were asked to respond using the keyboard whether the probe word was abstract or concrete, by pressing either the “f” or “j” key, respectively. The mapping of response to key press was alternated across participants. Participants were able to respond as soon as the probe word appeared on the screen. Reaction times (RTs) were recorded as the time between the onset of the probe word and the button response on each trial. Before starting the experiment, participants completed 10 practice trials. During the experiment, each participant was presented with 1,312 trials so that every possible prime-probe word pair from within each category was shown twice along with an equal number of between-category word pairs. Participants also saw an equal number of trials in which the prime was an abstract noun while the probe was concrete and vice versa (mixed-word-type trials). Thus, there were an equal number of trials in which the correct response was ‘abstract’ or ‘concrete’ across the experiment. The trials were randomly shuffled and separated into eight runs for each participant. Each run lasted 6 min and 22 s.

Data analysis

The data were processed using Matlab. For each participant, trials with incorrect responses or RTs that were two and a half standard deviations above or below the mean were excluded from further analyses. Participants were excluded from the analysis if their accuracy was less than 80% across all trials or the average RT across the experiment was two and a half standard deviations above or below the group mean. Across participants, accuracy was very high (94.7%) and there was not a significant difference in accuracy across within- and between-category trials for abstract or concrete words (both t 's < 1.57, both p 's > 0.10). None of the 25 participants were excluded from the analysis. The RTs for all within- and between- category trials were separately averaged for the abstract and concrete words. The mixed-word-type trials were not analyzed further. Repeated-measures ANOVAs and planned paired-sample t -tests were used to compare the within- and between- category RTs for abstract and concrete words across participants.

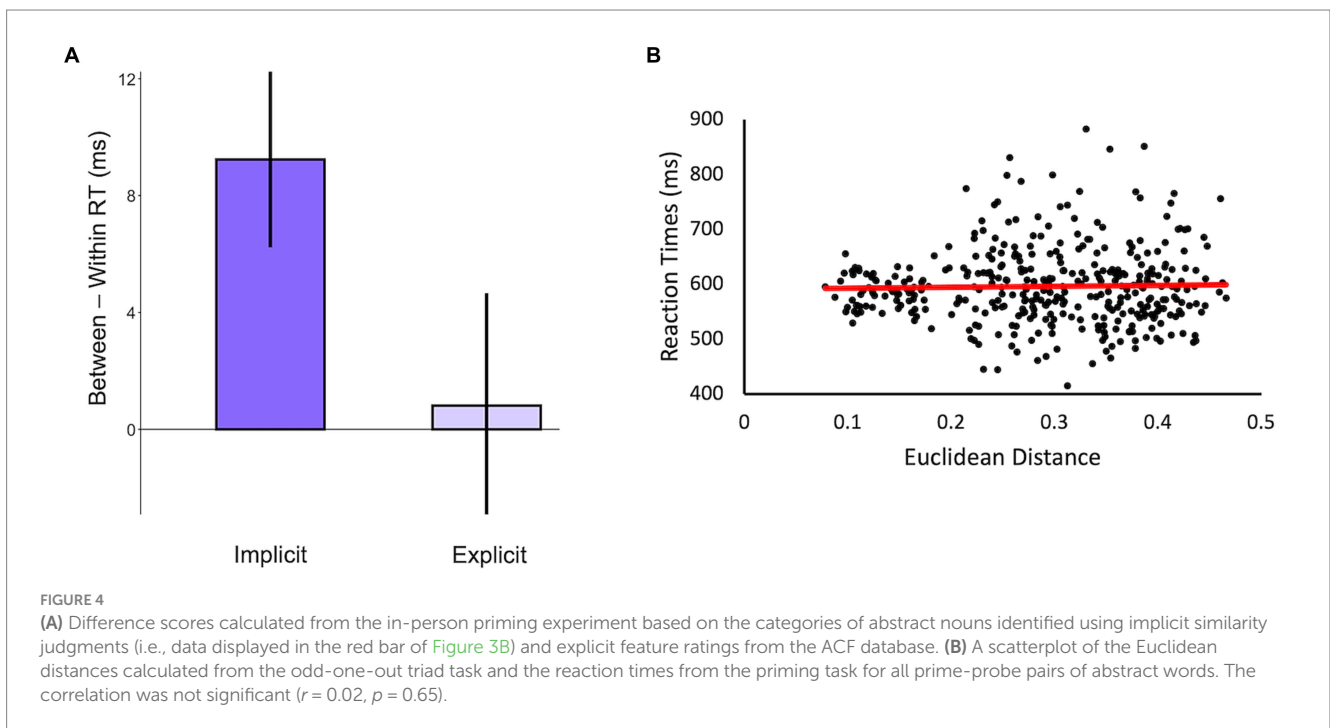
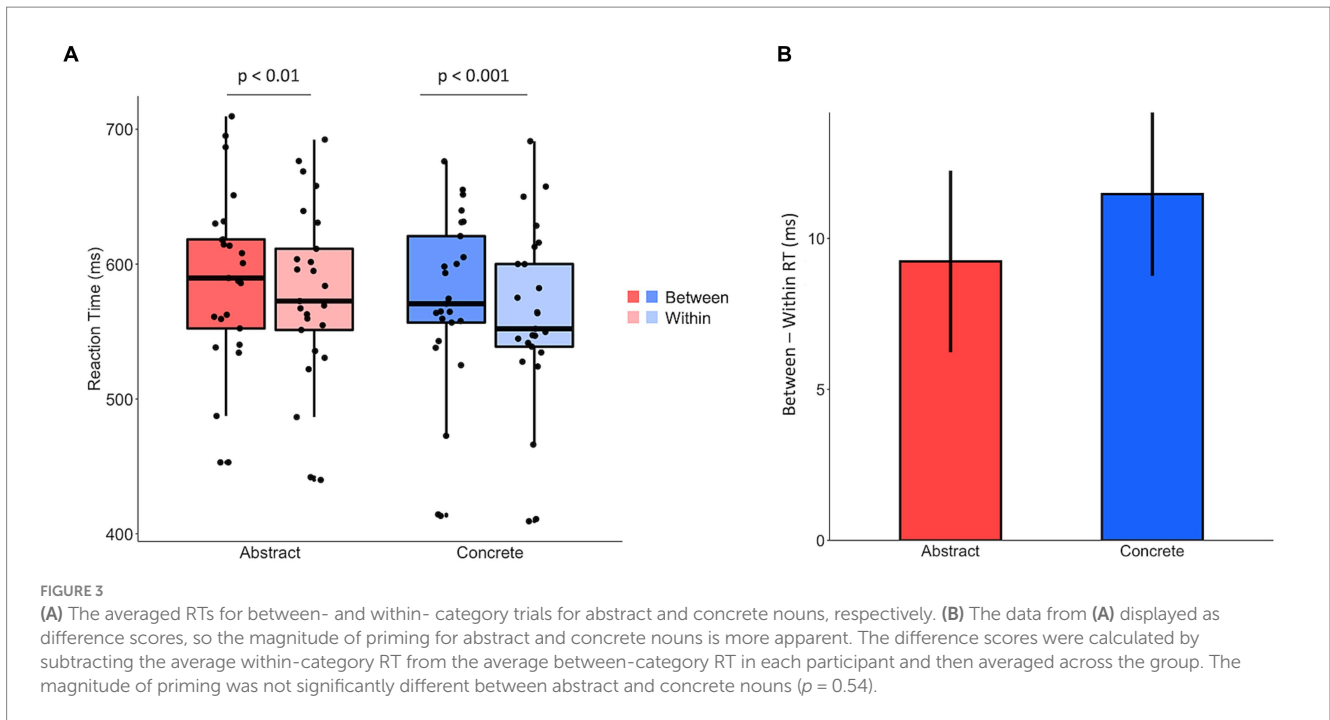
M Turk replication

One-hundred and eight people were recruited from the M Turk Master Subject pool to complete the same automatic semantic priming task as described above. Reaction time data were collected and analyzed using the same procedure as the in-person semantic priming task described above. After data processing, two participants were excluded from the analysis: one because the average RT was above two and a half standard deviations from the group mean and the other for chance-level accuracy.

Results

Paired-sample t -tests revealed significant priming effects (i.e., greater between- than within- category RTs) for both abstract ($t_{(24)} = 3.07$, $p < 0.01$) and concrete ($t_{(24)} = 4.23$, $p < 0.001$) words (Figure 3A). Furthermore, a 2 (Abstract, Concrete) \times 2 (Within, Between) repeated-measures ANOVA revealed a main effect of word type ($F_{(1,24)} = 11.64$, $p < 0.01$), but not a significant interaction ($F_{(1,24)} = 0.385$, $p = 0.54$). These results demonstrate that while participants were faster to respond to concrete words, the magnitude of priming was not significantly different for abstract and concrete nouns (Figure 3B). To ensure that the priming effect for the abstract categories was robust, we replicated the result in 106 participants on M Turk. Paired-sample t -tests again found significant priming effects for both abstract ($t_{(105)} = 2.07$, $p < 0.05$) and concrete ($t_{(105)} = 7.33$, $p < 0.001$) words.

Next, we tested whether the significant priming effect for abstract words was specific to our method of identifying categories or if using explicit feature ratings would also suffice. To do so, we reshuffled the category membership of the words using the clusters derived from explicit feature ratings in the ACF database (Supplementary Figure 1) and calculated the magnitude of priming assuming these categories. We then compared the magnitude of priming to the priming effect found using the categories that we identified using similarity judgments about the abstract words. As predicted, we did not find a significant priming effect for abstract nouns when the categories were based on explicit feature ratings ($t_{(24)} = 0.21$, $p = 0.84$) and a



repeated-measures ANOVA revealed an interaction between the priming magnitudes, such that there is less priming when categories are based on explicit feature ratings compared to when categories are based on implicit judgments of similarity ($F_{(1,24)}=3.74, p=0.06$ – Figure 4A).

Finally, we addressed the concern that abstract nouns do not fit naturally into categories but are instead organized within a continuous semantic space. In this case, the categories identified using our method might simply be an artifact of using a clustering algorithm on the similarity matrix of abstract nouns. If so, then

we should expect to find that RTs for pairs of abstract nouns are correlated with the Euclidean distances between the words, regardless of whether they are from the same category. As predicted, we did not find a significant correlation between RTs and Euclidean distances across prime-probe pairs of abstract nouns ($r=0.02, p=0.65$; Figure 4B). This, combined with the replicated pattern of clustering across independent datasets and significant priming for within-category abstract words indicates that the categories identified in this study reflect intrinsic categories within the semantic space of abstract nouns.

Discussion

We collected non-verbal and implicit judgments about which dimensions are relevant to the similarity between pairs of 50 abstract nouns and then used a rigorous clustering algorithm to identify categories within the representational space of the words. We then used automatic semantic priming to ensure that we identified intrinsic categories, rather than arbitrary categories resulting from a strategic process. We found five intuitive and replicable semantic categories within the representational space of the 50 abstract nouns and found significant automatic semantic priming for abstract nouns and a control set of concrete nouns. To the extent that our results generalize to the huge corpus of abstract words, our results suggest that the semantic space of abstract words has a category-like structure that is similar to the taxonomic structure of concrete words.

Importantly, we also found that our method of mapping the representational space of abstract words provides a significant advantage over more commonly used methods that are based on collecting explicit judgments about abstract words. In Experiment 1, we used chatGPT to show that the categories from the triad task are more easily labeled than the categories defined using dimensional weights from an explicit rating experiment (Troche et al., 2017). In Experiment 2, we showed that the significant automatic semantic priming effect that was replicated across two independent datasets is eliminated when the words are re-categorized based on the dimensions from the explicit ratings. The practice of collecting explicit judgments about (abstract and concrete) concepts – whether by asking people to list relevant features of concepts or to rate the concepts along predetermined dimensions – is ubiquitous in cognitive science (McRae et al., 2005; Crutch et al., 2013; Devereux et al., 2014; Villani et al., 2019; Lynott et al., 2020; Banks and Connell, 2023). However, this method is not ideal for characterizing the representational space of concepts because people may not always list all of the relevant features of the concepts (McRae et al., 2005; Löhr, 2023) or the experimenter may choose suboptimal feature dimensions to include in the experiment, thus leaving the representational space incomplete – this is true for concrete concepts and it is doubly true for abstract concepts since they lack easily referenced perceivable features (Borghi et al., 2017). By contrast, the odd-one-out similarity task used in Experiment 1 collects judgments of similarity between word pairs in the presence of a third word in each trial. The third word provides a context for the word pair and thus implicitly highlights the relevant dimension (s) of similarity between the word pair. For example, “sadness” and “hope” might be judged as similar when shown in a trial with “variety” because the participant chooses “variety” as the odd one out, but the words might be judged as dissimilar when shown with “worry”, since the participant might now choose “hope” as the odd one out. From these two trials, it seems that affect and valence are important dimensions underlying the semantic relation between “sadness” and “hope”. Across enough trials (and different contexts), we can map the multidimensional feature space of a set of concepts (Hebart et al., 2020; Muttenthaler et al., 2022). In the current study, we demonstrated the utility of this method by using it to identify behaviorally relevant categories in a set of abstract nouns.

To be clear, we are not claiming that results from experiments like the one used in Troche et al. (2017) are wrong, rather our results suggest that we can better derive validated categorical boundaries using the triad task. Far from thinking that collecting explicit ratings is a futile endeavor, we believe that results from such rating paradigms, and other methods used to understand word meaning, are complimentary to information gained from using our method of collecting implicit measures of

similarity. For example, it would be interesting to further compare the dimensions derived from different methods and then potentially optimally combine them to better map the representational space of words. Furthermore, other information that is not directly related to conceptual content (e.g., age of acquisition, modality of acquisition, reliance on others to learn a given word) would also likely be quite valuable in future pursuits of mapping the representational space of abstract (and concrete) concepts.

The main limitation to the current study is the relatively small set of words used in the experiment. However, our decision to use a set of 50 abstract nouns was a practical one. Since the odd-one-out task had not been used with abstract concepts, we chose a small, but adequate, number of words to ensure that we could sample the entire space of triads and thus get a reliable estimate of the representational space of the abstract concepts. Indeed, we were able to collect two complete sets of triads for the 50 words so we could replicate the results across independent datasets (19,600 trials per set). Our method of replicating the category structure within this subset of abstract concepts across large groups of participants and different experimental paradigms produced results that should generalize to the full domain of abstract concepts. However, the methods in Experiment 1 were intentionally conservative to ensure that the categories were stable, thus leaving 11 words unsorted. The 11 words were unsorted here because of our decisions to include a size constraint for the categories and to require the categories to replicate across independent datasets, not necessarily because they are uncategorizable concepts – in fact, this is an empirical question that can be answered by extending our method to a larger set of abstract concepts. Thus, the results of our study demonstrate that it is a worthwhile endeavor to collect similarity judgments using the odd-one-out triad task on a larger set of abstract (and concrete) words to obtain a more complete map of the representational space that underlies our conceptual knowledge. Even prior to generalizing the results of this study to a larger set of concepts, the categories identified here can be used to select validated stimuli in future experiments.

In conclusion, we identified behaviorally valid categories within the representational space of 50 abstract nouns. To do so, we used a method of collecting implicit measures of similarity between the concepts to initially define the category structure and showed that this method provides more stable and interpretable results than commonly used methods that are based on collecting explicit judgments about abstract concepts. We then used automatic semantic priming as a strong test of the intrinsic nature of the categories. Taken together, our results demonstrate that abstract concepts can be characterized beyond their negative relation to concrete concepts by investigating their relation to one another using the appropriate methods.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author. The raw data used in this study are available on the Open Science Foundation webpage. doi: [10.17605/OSF.IO/MJUKT](https://doi.org/10.17605/OSF.IO/MJUKT).

Ethics statement

The studies involving humans were approved by NIH Institutional Review Board. The studies were conducted in accordance with the

local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AP: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft. JS: Data curation, Formal analysis, Investigation, Writing – original draft. JD: Data curation, Formal analysis, Investigation, Writing – original draft. SG: Conceptualization, Methodology, Supervision, Writing – review & editing. AM: Conceptualization, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the NIMH Intramural Research Program. (#ZIA MH002588, clinical trials number NCT00001360).

Acknowledgments

We thank Martin Hebart for technical assistance.

References

- Allen, R., and Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *J. Mem. Lang.* 55, 64–88. doi: 10.1016/j.jml.2006.02.002
- Banks, B., and Connell, L. (2023). Multi-dimensional sensorimotor grounding of concrete and abstract categories. *Philos. Trans. R. Soc. B Biol. Sci.* 378:20210366. doi: 10.1098/rstb.2021.0366
- Barsalou, L. W. (1983). Ad hoc categories. *Mem. Cogn.* 11, 211–227. doi: 10.3758/BF03196968
- Battig, W. F., and Montague, W. E. (1969). Category norms of verbal items in 56 categories a replication and extension of the Connecticut category norms. *J. Exp. Psychol.* 80, 1–46. doi: 10.1037/h0027577
- Bleasdale, F. A. (1987). Concreteness-dependent associative priming: separate lexical organization for concrete and abstract words. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 582–594. doi: 10.1037/0278-7393.13.4.582
- Borghini, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., and Tummolini, L. (2017). The challenge of abstract concepts. *Psychol. Bull.* 143, 263–292. doi: 10.1037/bul0000089
- Breedin, S. D., Saffran, E. M., and Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cogn. Neuropsychol.* 11, 617–660. doi: 10.1080/02643299408251987
- Brown, T., et al. (2020). Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* 33, 1877–1901.
- Cree, G. S., and 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). *J. Exp. Psychol. Gen.* 132, 163–201. doi: 10.1037/0096-3445.132.2.163
- Crutch, S. J., Troche, J., Reilly, J., and Ridgway, G. R. (2013). Abstract conceptual feature ratings: the role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Front. Hum. Neurosci.* 7:186. doi: 10.3389/fnhum.2013.00186
- Desai, R. H., Reilly, M., and van Dam, W. (2018). The multifaceted abstract brain. *Philos. Trans. R. Soc. B Biol. Sci.* 373:122. doi: 10.1098/rstb.2017.0122
- Devereux, B. J., Tyler, L. K., Geertzen, J., and Randall, B. (2014). The Centre for Speech, language and the brain (CSLB) concept property norms. *Behav. Res. Methods* 46, 1119–1127. doi: 10.3758/s13428-013-0420-4
- Franklin, S., Howard, D., and Patterson, K. (1995). Abstract word anomia. *Cogn. Neuropsychol.* 12, 549–566. doi: 10.1080/02643299508252007
- Gilhooly, K. J., and Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behav. Res. Methods Instrum.* 12, 395–427. doi: 10.3758/BF03201693
- Hebart, M. N., Zheng, C. Y., Pereira, F., and Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nat. Hum. Behav.* 4, 1173–1185. doi: 10.1038/s41562-020-00951-3
- Löhr, G. (2023). Does the mind care about whether a word is abstract or concrete? Why concreteness is probably not a natural kind. *Mind Lang.* 12473. doi: 10.1111/mila.12473
- Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychon. Bull. Rev.* 7, 618–630. doi: 10.3758/BF03212999
- Lynott, D., Connell, L., Brysbaert, M., Brand, J., and Carney, J. (2020). The Lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behav. Res. Methods* 52, 1271–1291. doi: 10.3758/s13428-019-01316-z
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. 1st Edn. New York: Psychology Press.
- McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559. doi: 10.3758/BF03192726
- Muttenthaler, L., Zheng, C. Y., McClure, P., Vandermeulen, R. A., Hebart, M. N., and Pereira, F. (2022). VICE: Variational Interpretable Concept Embeddings. *Adv. Neural Inf. Process. Syst. Adv. Neural Inf. Process. Syst.* 35, 33661–33675.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol. Gen.* 106, 226–254. doi: 10.1037/0096-3445.106.3.226
- Neely, J. H. (1991). “Semantic priming effects in visual word recognition: a selective review of current findings and theories” in *Basic processes in reading: Visual word recognition*. eds. D. Besner and G. W. Humphreys (Hillsdale, NJ: Erlbaum)
- Paivio, A. (1991). Dual coding theory: retrospect and current status. *Can. J. Psychol./Rev. Can. Psychol.* 45, 255–287. doi: 10.1037/h0084295
- Plaut, D. C., and Shallice, T. (1993). Deep dyslexia: a case study of connectionist neuropsychology. *Cogn. Neuropsychol.* 10, 377–500. doi: 10.1080/02643299308253469
- Riddoch, M. J., Humphreys, G. W., Coltheart, M., and Funnell, E. (1988). Semantic systems or system? Neuropsychological evidence re-examined. *Cogn. Neuropsychol.* 5, 3–25. doi: 10.1080/02643298808252925

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1278744/full#supplementary-material>

- Rosch, E. H. (1973). Natural categories. *Cogn. Psychol.* 4, 328–350. doi: 10.1016/0010-0285(73)90017-0
- Rosch, E. (1975). Cognitive representations of semantic categories. *J. Exp. Psychol. Gen.* 104, 192–233. doi: 10.1037/0096-3445.104.3.192
- Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U. S. A.* 105, 1118–1123. doi: 10.1073/pnas.0706851105
- Rosvall, M., and Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical Organization in Large Integrated Systems Rapallo F, ed. *PLoS One* 6:e18209. doi: 10.1371/journal.pone.0018209
- Schwanenflugel, P. J., Harnishfeger, K. K., and Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *J. Mem. Lang.* 27, 499–520. doi: 10.1016/0749-596X(88)90022-8
- Schwanenflugel, P. J., and Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 82–102. doi: 10.1037/0278-7393.9.1.82
- Sperber, R. D., McCauley, C., Ragain, R. D., and Weil, C. M. (1979). Semantic priming effects on picture and word processing. *Mem. Cogn.* 7, 339–345. doi: 10.3758/BF03196937
- Troche, J., Crutch, S. J., and Reilly, J. (2017). Defining a conceptual topography of word concreteness: clustering properties of emotion, sensation, and magnitude among 750 English words. *Front. Psychol.* 8:1787. doi: 10.3389/fpsyg.2017.01787
- Villani, C., Lugh, L., Liuzza, M. T., and Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Lang. Cogn.* 11, 403–430. doi: 10.1017/langcog.2019.23
- Walker, I., and Hulme, C. (1999). Concrete words are easier to recall than abstract words: evidence for a semantic contribution to short-term serial recall. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 1256–1271. doi: 10.1037/0278-7393.25.5.1256
- Warrington, E. K. (1975). The selective impairment of semantic memory. *Q. J. Exp. Psychol.* 27, 635–657. doi: 10.1080/14640747508400525