JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles | Behavioral/Cognitive

Neural representations of concreteness and concrete concepts are specific to the individual

https://doi.org/10.1523/JNEUROSCI.0288-24.2024

Received: 12 February 2024
Revised: 29 August 2024
Accepted: 9 September 2024

Neural representations of concreteness and concrete concepts are specific to the individual

Thomas L. Botch[1*] & Emily S. Finn[1*]

[1] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH, 03755

[*] Correspondence:

Thomas L. Botch, *tlb.gr@dartmouth.edu*

Emily S. Finn, *emily.s.finn@dartmouth.edu*

Abstract

Different people listening to the same story may converge upon a largely shared interpretation while still developing idiosyncratic experiences atop that shared foundation. What linguistic properties support this individualized experience of natural language? Here, we investigate how the "concrete-abstract" axis — i.e., the extent to which a word is grounded in sensory experience — relates to within- and across-subject variability in the neural representations of language. Leveraging a dataset of human participants of both sexes who each listened to four auditory stories while undergoing functional MRI, we demonstrate that neural representations of "concreteness" are both reliable across stories and relatively unique to individuals, while neural representations of "abstractness" are variable both within individuals and across the population. Using natural language processing tools, we show that concrete words exhibit similar neural representations despite spanning larger distances within a high-dimensional semantic space, which potentially reflects an underlying representational signature of sensory experience — namely, imageability — shared by concrete words but absent from abstract words. Our findings situate the concrete-abstract axis as a core dimension that supports both shared and individualized representations of natural language.

Significance Statement

The meaning of spoken language is often ambiguous. As a result, people may form different interpretations despite being presented with the same information. What properties of language does the brain leverage to form this diverse, individual experience? Analyses of functional MRI data demonstrated that "concreteness", the extent to which language is related to sensory experience, evoked reliable neural patterns that were unique to individual subjects and allowed us to identify individuals solely based on their neural data. Application of machine learning methods showed that sets of concrete concepts, but not abstract concepts, show stable neural patterns, potentially due to a sensory signature: imageability. Overall, this study characterizes concreteness as a central property supporting the individualized experience of real-world language.

Introduction

The success of language as a means of communication relies on a shared understanding of the meanings of words as links to mental concepts (Elman, 2004; Stolk et al., 2016; Thompson et al., 2020). While people generally converge in how they understand and represent language (Fedorenko & Thompson-Schill, 2014; Malik-Moraleda et al., 2022), the conceptual associations evoked by a given word can also be highly individualized and informed by experience (Elman, 2009; Yee & Thompson-Schill, 2016). What linguistic properties scaffold common conceptual knowledge while also providing the foundation for idiosyncratic representations?

A large body of empirical and theoretical work has suggested that human knowledge is organized along an axis that moves from concrete, sensory-based representations to abstract, language-derived representations (Bedny & Caramazza, 2011; Bi, 2021; Borghi et al., 2017; Paivio, 1991). Within this framework, "concrete" words are experienced directly through senses or actions (e.g., dog, table) while "abstract" words have meanings dependent on language (e.g., idea, plan).

Together, concreteness and abstractness represent ends of a continuum of "sensory grounding", where a given word can be placed along this axis based on the degree to which it can be experienced directly through one's senses. Accordingly, each word is assumed to share this property with other words at a similar position along the axis, irrespective of their meanings. Theories of "grounded cognition" (Barsalou, 2008; Binder & Desai, 2011) propose that concrete words benefit from being jointly represented across both sensory and linguistic domains and, as a result, exhibit more stable representations than abstract words. Recent findings from human neuroimaging provide support for these theories, demonstrating close topographical and functional correspondence between representations of sensory and linguistic information (Deniz et al., 2019; Huth et al., 2016; Popham et al., 2021). In turn, the concreteness of words benefits behavior: concrete words are processed faster (Kroll & Merves, 1986; Paivio & Begg, 1971; Roxbury et al., 2014; Schwanenflugel et al., 1988), are more imageable (Altarriba et al., 1999; Tuckute et al., 2018), and are more easily recalled than abstract words (Aka et al., 2021; Gorman, 1961; M. Hamilton & Rajaram, 2001; Romani et al., 2008; Walker & Hulme, 1999).

While studies often highlight population-level commonalities in how people process and represent concrete versus abstract words, researchers have also identified differences in how individuals organize and represent concrete versus abstract language (X. Wang & Bi, Yanchao, 2021). Specifically, concrete words are more similar both across (X. Wang & Bi, Yanchao, 2021) and within subjects (Musz & Thompson-Schill, 2015) in both their conceptual organization (as measured behaviorally with a semantic distance task) and neural representations. However, the extent to which representations of the concrete-abstract axis itself, rather than individual words along that axis, are stable across experiences and unique to each person remains unclear. On one hand, representations of individual concrete words may be more stable due to each word's unique sensory grounding that stabilizes its own representation and distinguishes it from other words. On the other hand, the property of concreteness may provide a shared structure that supports the representation of each individual word, elevating the similarity among all concrete words as a class despite differences in sensory grounding and word meaning. Together, this complicates the interpretation of previous findings: across subjects, the low similarity of abstract word representations may result not only from variability in individual word representations, but also variability in representing the property of "abstractness" more generally (Figure 1C).

One possibility is that representations of abstractness might be highly individualized—in other words, both unique to the individual and shared across distinct abstract words within that individual. Such individual-specific representations would be evidenced by high within-subject similarity across exposures to different abstract words, despite low across-subject similarity. Another possibility is that low similarity results from unstable representations of abstractness. In this case, representations would show low similarity both within and across subjects that could result from high variability in abstractness across contexts. Yet, without evaluating the reliability of representations *within* subjects and *across* words, the low similarity of abstract word representations *across* subjects is difficult to interpret.

Here, we aimed to understand how the concrete-abstract axis provides a foundation for individual differences in the neural representation of language. We investigated this question within a large dataset of subjects who listened to four naturalistic auditory stories during functional magnetic resonance imaging (fMRI). Unlike many previous investigations that used isolated single-word or otherwise simplified paradigms (Binder et al., 2005; Fernandino et al., 2022; Friederici et al., 2000; Musz & Thompson-Schill, 2015; Roxbury et al., 2014; Vignali et al., 2023; X. Wang & Bi, Yanchao, 2021; West & Holcomb, 2000), these data allowed us to characterize neural representations of the concrete-abstract axis within contextualized speech, as language is used in everyday life (L. S. Hamilton & Huth, 2020). We tested not only the extent to which neural representations of concreteness and abstractness are consistent across subjects, but also the degree to which these representations are reliable within and unique to a given subject across stories. Then, by leveraging tools from natural language processing, we relate our findings on concreteness and abstractness to prior work on word meanings by taking sets of similar concepts as a proxy for repeated words across stories. Specifically, we examined how the organization of words within a high-dimensional semantic space relates to differential reliability of how concreteness versus abstractness are represented in the human brain.

Methods

Participants

We used a subset of data from the publicly available *Narratives* dataset (Nastase et al., 2021). Specifically, we used data from 45 subjects (N=33 female; mean age 23.3 +/- 7.4 years) who each listened to four auditory stories ("Running from the Bronx", 8:56 min; "Pie Man (PNI)", 6:40 min; "I Knew You Were Black", 13:20 min; "The Man Who Forgot Ray Bradbury", 13:57 min) during fMRI scans at the Princeton Neuroscience Institute (Figure 1A). All stories were collected within the same testing session and each story was collected within a separate run. Across participants, the order of stories was pseudo-randomized such that "Bronx" and "Pie Man (PNI)" were always presented in the first half of the session while "Black" and "Forgot" were presented in the second half of the session. The order of the stories presented within each half of the session was then randomized, resulting in four possible presentation orders across participants. All participants completed written informed consent, were screened for MRI safety and reported fluency in English, having normal hearing, and no history of neurological disorders. The study was approved by the Princeton University Institutional Review Board.

MRI data acquisition and preprocessing

Functional and anatomical images were collected on a 3T Siemens Magnetom Prisma with a 64-channel head coil. Whole-brain images were acquired (48 slices per volume, 2.5mm isotropic resolution) in an interleaved fashion using a gradient-echo EPI (repetition time (TR) = 1.5s, echo time (TE) = 31ms, flip angle (FA) = 67°) with a multiband acceleration factor of 3 and no in-plane acceleration. A total of 1717 volumes were collected for each participant across four separate scan runs, where a single story was presented within each run.

We used preprocessed data provided by Nastase et al., 2021. In brief, data were preprocessed using *fmriprep* (Esteban et al., 2019) including co-registration, slice-time correction, and non-linear alignment to the MNI152 template brain. Time-series were detrended with regressors for motion, white matter, cerebrospinal fluid and smoothed with a 6mm FWHM gaussian kernel. For more information about data acquisition and preprocessing, please refer to Nastase et al., 2021.

As an additional preprocessing step, we performed functional alignment on these data using a shared response model (Chen et al., 2015) as implemented in *BrainIAK* (Kumar et al., 2021). Previous work has demonstrated better functional alignment by fitting a SRM within each parcel (Bazeille et al., 2021). Accordingly, we restricted our analyses to the neocortex and used the 200-parcel Schaefer parcellation (Schaefer et al., 2018) and removed any parcel without at least 75% coverage across all participants and stories (total parcels removed: 9/200, or 4.5%). Within each remaining parcel, we then fit a model to capture reliable responses to all stories across participants in a lower dimensional feature space (number of features = 50). We then inverted the parcel-wise models to reconstruct the individual voxel-wise time courses for each participant and each story (Yates et al., 2021). This procedure served as an additional denoising step to improve the consistency of stimulus-driven spatiotemporal patterns across participants. All analyses were conducted in volume space and projected to surface space (*fsaverage*) using *nilearn* (Abraham et al., 2014) for visualization purposes only.

Stimulus preprocessing

Each story was originally transcribed and aligned to the audio file using the Gentle forced-alignment algorithm by the authors of Nastase et al., 2021. We applied additional preprocessing to the transcripts using the Natural Language Toolkit (Bird et al., 2009). First, we obtained parts-of-speech and word lemmas — the base form of a word (e.g., “go” is the lemma for “going”, “gone” and “went”) — for each word, and excluded stop-words (uninformative, common words) such as “the”, “a”, and “is”.

To address our hypotheses, we leveraged an existing corpus of human ratings of word concreteness (Brysbaert et al., 2014). In this study, online participants rated a total of 40,000 English word lemmas on a 5-point Likert scale from abstract (lower) to concrete (higher). Each word was rated by at least 25 participants. Participants were instructed to consider a word as more concrete if it refers to something that exists in reality and can be experienced directly through senses or actions, and, in contrast, to consider a word as more abstract if its meaning depends on language and cannot be experienced directly through senses or actions. Henceforth, we use “concrete-abstract axis” to refer to this general linguistic dimension, and “concreteness” as a word’s specific position on this axis.

For each word in each story, we assigned a value of concreteness using the average human rating for that word’s lemma if it was present in the concreteness corpus (Figure 1B). In addition to our critical predictor (concreteness), we included three other linguistic properties as controls: frequency (Brysbaert et al., 2019; Brysbaert & New, 2009), a measure of how often a word occurs in language, and two affective properties, valence and arousal (Warriner et al., 2013). Word frequency was derived objectively by calculating the number of occurrences of a word per million words (51 million total words), while valence and arousal were derived from human ratings analogous to the concreteness ratings described above. Previous research investigating word frequency effects have demonstrated that less frequent words drive stronger neural responses within the language network (Fiebach et al., 2002; Schuster et al., 2016). A separate set of studies investigating affect have demonstrated that valence and arousal contribute to representations of language within areas related to emotion processing and memory (Brooks et al., 2016; Kensinger & Schacter, 2006). While the selected control properties are not a definitive list, including them as "competition" allows us to make inferences that are more specific to the concrete-abstract axis. Our analysis was then constrained to the set of words with a value for any of the four properties (i.e., the union), resulting in 97.7% of content words sampled on average across stories (2449 words of the possible 2500 content words). We were able to model the majority of these content words within each linguistic predictor (concreteness: 96.4%, frequency: 97.7%, valence: 83%, arousal: 83%). Importantly, collinearity between the critical regressor, concreteness, and other linguistic properties varied, showing a moderate relationship with word frequency and weak relationships with all other properties (average Pearson's r across stories: arousal = -0.10; frequency = -0.30; valence = -0.05).

fMRI Analysis

Modeling representations of word properties

For each story and participant, we used a general linear model (GLM) to estimate BOLD responses for each linguistic property (concreteness, frequency, valence, arousal), plus a low-level auditory feature regressor (loudness: the root mean square of the auditory waveform). We collectively refer to these linguistic and sensory properties as “word properties”.

To construct a continuous, amplitude-modulated regressor, each word property was assigned a value at each timepoint of the story timeseries based on the word(s) spoken at that timepoint. We then modeled BOLD signal as a function of these regressors using AFNI (Cox, 1996). The model yields a map of beta values that correspond to responses to each property, where higher and lower values indicate higher and lower values of a given linguistic property (e.g., higher = more concrete, lower = more abstract). As all word properties were included in the same model, the resulting beta values represent the BOLD response to a given property while controlling for all other properties.

Using the outputs from these models, we first examined group-level univariate responses to each word property using a linear-mixed effects model. At each voxel, the model predicts BOLD activity from the fixed effects of each property plus the random effects of subject and story. The model therefore yields a map of beta values that describes consistent neural responses to each property across stories and subjects. All voxel-wise results are shown following correction for multiple comparisons ($q_{FDR} < 0.05$).

Evaluating the reliability of representations of the concrete-abstract axis and other word properties

To understand whether word properties elicit reliable representations during story listening (Figure 1C), we examined the within- and across-subject multivariate pattern similarity of evoked responses for each property across stories. We first divided the cortex into 200 parcels using the Schaefer parcellation (Schaefer et al., 2018). Then, within each parcel, we correlated the multivoxel pattern of beta values between all pairs of participants, repeating this process for each unique pair of stories (six total pairs). Lastly, we averaged across all story-pair matrices to obtain a subject similarity matrix for each parcel (denoted as *M* within the following equations). We repeated this procedure for each property to understand the similarity of neural representations across stories both within- and across-subjects. See Figure 1D for a schematic of this analysis.

We evaluated two multivariate signatures of these neural representations (Figure 1D). Our first method, reliability, assesses the similarity of a subject's representations to themselves across stories compared to the similarity of their representations to those of other subjects. Specifically, reliability is calculated as the difference between the similarity of a subject to themselves (within-subject similarity) and the average pairwise similarity of a subject to all other subjects (across-subject similarity).

$$reliability = \frac{1}{N}\sum_{i=1}^{N} M_{i,i} - (\frac{1}{N}\sum_{j=1, i\neq j}^{N} M_{i,j})$$

Our second method, identifiability, measures how unique representations are to each subject. A subject is said to be identifiable based on their representations when, across stories, similarity of a subject to themselves is higher than similarity to all other participants of the group. For each

$$identifiability = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}1 & argmax(M_{i,1:N}) = M_{i,i} \\ 0 & otherwise\end{cases}$$

parcel, we calculate identifiability as fingerprinting accuracy: the average number of participants identifiable based on their neural representations (Finn et al., 2015).

For both reliability and identifiability analyses, statistical significance was evaluated via permutation testing. Specifically, for each parcel, we permuted the rows of the subject similarity matrix and recalculated reliability and identifiability values. This process was repeated 10,000 times and observed values were tested against this null distribution. Resulting *p*-values for each signature were corrected for multiple-comparisons across 200 parcels using the Benjamini-Hochberg method ($q_{FDR} < 0.05$). To evaluate reliability and identifiability at a whole-brain level, for each signature, we used a linear-mixed effects model to predict reliability/identifiability from the fixed-effect of word property while controlling for the random effect of parcel in both models and a random effect of subject within the reliability model. We tested for significant differences between word properties by conducting pairwise statistical tests between model fits to each property.

To understand what was driving observed reliability — i.e., high within-subject consistency, low across-subject similarity, or both — we compared within-subject similarity to across-subject similarity. Specifically, we calculated across-subject similarity in two ways: 1) in the *same* stories and 2) across *different* stories. For each word property, we used one-sample tests to assess significance of similarity of representations for each form of similarity. Then, we used a linear-mixed effects model to evaluate whether within-subject similarity was higher than both forms of across-subject similarity. All tests were two-tailed, tested at alpha $p < 0.05$, and corrected for multiple-comparisons using FDR correction.

Disentangling the reliability of representations of concreteness versus abstractness

We next aimed to understand whether concreteness and abstractness differentially contribute to the reliability of neural representations of the concrete-abstract axis. To this end, within each story, we limited our analysis to nouns (as verbs were more prevalent at the abstract end) and dichotomized the concrete-abstract axis by selecting the top 30% of concrete and top 30% of abstract words (Figure 4A). Specifically, we asked if and where representations of concreteness are more reliable than representations of abstractness or vice versa.

We used a GLM to estimate separate BOLD response patterns for concreteness and abstractness (using regressors defined based on the top 30% of words at each end). Within this model, we specified concreteness and abstractness as event regressors, discarding the amplitude component and treating all words of a given property as contributing equally to the model of BOLD response. The regressors for concreteness and abstractness each contained a total of 187 words aggregated across stories, resulting in a total of 374 words modeled across stories (*black*: 94 words; *bronx*: 92 words; *piemanpni*: 68 words; *forgot*: 120 words). We also included two amplitude-modulated regressors, word frequency and loudness, to control for differences in low-level linguistic and sensory features. We then repeated our analysis of reliability and identifiability (described above) on the beta maps of concreteness and abstractness separately.

For each parcel, we contrasted the reliability of concreteness and abstractness within each subject by applying Fisher's z-transformation and taking the difference between the reliability scores (concrete minus abstract), limiting our analysis to parcels that showed significant reliability for *either* concreteness or abstractness. Then, within each parcel, we conducted paired t-tests to identify parcels that significantly differed in their reliability of concreteness and abstractness representations. All tests were two-tailed, tested at alpha $p < 0.05$, and corrected for multiple-comparisons using FDR correction.

Evaluating the stability of representations of concrete versus abstract concept clusters

In light of the finding that representations of concreteness are more reliable than those of abstractness (cf. Figure 4B), we asked whether this higher reliability is driven by closer and more stable semantic relationships between words at the concrete end of the spectrum. To define semantic relationships between words, we used a natural language processing model (GloVe; (Pennington et al., 2014) to embed each word in both the top 30% concrete and top 30% abstract word sets, aggregated across stories, within a high-dimensional semantic space (Figure 5A). We then applied spectral clustering (Shi & Malik, 2000) over the concrete and abstract word embeddings to obtain clusters for each end of the spectrum (k=3 each for the concrete and abstract ends, so six total) composed of semantically similar words, which we refer to as "concept clusters". While we selected k=3 clusters because this value of *k* yielded the most balanced number of words in each cluster, similar results were obtained at both k=2 and k=4 clusters. These clusters grouped concrete and abstract words into sets of related concepts — such as a food-related concrete cluster containing the words "bread" and "cheese" — that were visually distinct when projected into a 2-dimensional space using UMAP (Figure 5B(McInnes et al., 2020). Importantly, words within each concept cluster could come from within the same story or from different stories.

In addition to visualizing the qualitative organization of concept clusters, we also formally tested the semantic similarity of words in the same or in different clusters, within and between ends of the concrete-abstract axis. Importantly, because the clustering itself was done on semantic distances, we expect that distances will be lower between words in the same versus different clusters, but this analysis also lets us quantify if and how semantic spread *across* clusters is greater at one end of the concrete-abstract axis than the other. Specifically, we calculated the cosine similarity between all pairs of words embedded within the semantic space. We then grouped these pairwise similarity values into the following categories: a) pairs of words within the same cluster, b) pairs of words in different clusters at the same end of the concrete-abstract axis (i.e., either concrete or abstract), and c) pairs of words at different ends of the concrete-abstract axis, which were (by definition) in different clusters. To compare these groups of similarity values, we used a linear-mixed effects model to evaluate how end of the property spectrum (concrete vs. abstract), cluster membership (within vs. between), and the interaction between these two features relate to the semantic similarity of cluster words while controlling for the random effect of word. To help interpret any resulting differences, we also conducted follow-up pairwise statistical tests. All tests were two-tailed, tested at alpha $p < 0.05$, and corrected for multiple-comparisons using FDR correction.

Next, we used a GLM to estimate BOLD responses to words within each concept cluster and evaluated both within- and across-subject similarity of these neural concept-cluster representations across stories. Similar to our analysis of semantic space, we calculated a) the similarity of neural representations of the same cluster across stories, b) the similarity of neural representations of different clusters at the same end of the spectrum (e.g., concrete clusters to other concrete clusters), and c) the similarity of neural representations between concrete clusters and abstract clusters. Crucially, all analyses of cluster similarity, both within- and across-subjects, are calculated as the similarity of clusters *across stories*; this allowed us to evaluate the stability and uniqueness of concept-cluster representations across distinct presentations and contexts.

Using two separate linear-mixed effects models, we examined how end of the property spectrum (concrete vs. abstract), cluster membership (within vs. between), and specific cluster relationship (e.g., within-concrete, between-concrete, etc.) differentially contribute to whole-brain similarity of neural representations while controlling for random effects of subject and parcel. Our first model predicts similarity from the fixed-effects of end of the property spectrum and cluster membership, and evaluates their main effects as well as their interaction. Then, in a separate model, we predict similarity from the fixed-effect of specific cluster relationship, specifying each cluster relationship as a separate level of the fixed effect. Using this second model, we tested for significant differences between cluster relationships by conducting pairwise statistical tests. All tests were two-tailed, tested at alpha $p < 0.05$, and corrected for multiple-comparisons using FDR correction.

Results

We aimed to understand how neural representations of the concrete-abstract axis vary within individuals and across the population during naturalistic story listening. Using a dataset of subjects (N=45) that listened to four stories each, we replicated previous findings that univariate neural responses to the concrete-abstract axis show group-level consistency. Complementing this consistency, we also found idiosyncratic multivariate representations of this axis that were unique to individuals and stable across stories, allowing us to identify subjects with a high degree of accuracy. Furthermore, by placing words within a high-dimensional semantic space, we demonstrated that neural representations of concrete words are particularly stable and stereotyped, and that this consistency primarily drives the reliability of the concrete-abstract axis, while representations of abstract words are more variable both within and across subjects.

Consistent group-level activations to the concrete-abstract axis

We first sought to replicate prior work demonstrating group-level consistency of univariate activity to the concrete-abstract axis. For each subject and story, we modeled brain activity as a function of the time-varying concreteness level of its content (as given by word-level norms provided by a separate set of human raters). Our model also included time-varying regressors for other linguistic properties — namely, frequency, valence, and arousal — plus loudness, a low-level sensory control.

All properties, both sensory and linguistic, demonstrated univariate neural responses that were consistent across both subjects and stories (Figure 2; $q_{FDR} < 0.05$). For example, as expected, loudness evoked responses in bilateral primary auditory cortex. Critically, the concrete-abstract axis evoked neural responses across a wide swath of cortex: more concrete words drove higher responses in regions including bilateral angular gyrus, bilateral parahippocampal cortex, and bilateral inferior frontal gyrus, while more abstract words drove responses in regions such as bilateral superior temporal gyrus and bilateral anterior temporal lobe. These results align with previous research that has reported similar cortical regions engaged in processing concrete and abstract concepts (Montefinese, 2019; J. Wang et al., 2010). Importantly, all linguistic properties exhibited responses that agree with prior research: frequency modulation in the left inferior frontal gyrus (Schuster et al., 2016), valence in the right temporoparietal junction (Tamir et al., 2016), and arousal in posterior cingulate (Maddock & Buonocore, 1997) and ventromedial prefrontal cortex (Kensinger & Schacter, 2006).

Representations of the concrete-abstract axis are reliable within individuals

Having shown that the concrete-abstract axis drives consistent univariate activity at the group level, we next investigated the stability of multivariate representations of this axis, as well other word properties, across stories. Representations were operationalized as multivoxel patterns of activity within each cortical parcel evoked by a given property in a given story. Specifically, we compared representations both within and across individuals, allowing us to understand the extent to which representations of these common linguistic dimensions are shared versus individualized.

We found that representations of all word properties except valence exhibited individual reliability across stories in at least some brain regions (Figure 3A; n = 10,000 permutations, all $q_{FDR} < 0.05$), where reliability was defined as the difference between within-subject and average across-subject similarity. Importantly, while the low-level sensory property of loudness showed the highest average reliability across parcels ($r = 0.11$), the concrete-abstract axis showed the second highest average reliability ($r = 0.09$) and was significantly more reliable than all other linguistic (i.e., non-sensory) properties (frequency: $r = 0.04$, $\beta = 0.01$, $t(42967) = 8.71$; valence: $r = -0.002$, $\beta = 0.05$, $t(42967) = 41.83$; arousal: $r = 0.02$, $\beta = 0.03$, $t(42967) = 23.74$; all ps < 0.001).

We next disentangled the separate contributions of within- and across-subject similarity in driving reliability of individual representations. In theory, high individual reliability of representations across stories could result from 1) highly *similar* representations within subjects, 2) highly *dissimilar* representations across subjects, or 3) a combination of the two. Accordingly, for each word property, we calculated the within- and across-subject similarity of representations. Specifically, we calculated the similarity of across-subject representations both within the *same* stories and across *different* stories. We compared the similarity of within-subject representations to both forms of across-subject similarity. Importantly, this comparison ensured that any observed differences in reliability stemmed from individualized representations (within-subject similarity) above and beyond characteristics of the presented stories.

For all word properties with significant reliability (i.e., all except valence), participants' representations were significantly similar to themselves across different stories (Figure 3B; one-sample t-tests, all *p*s < 0.001). Critically, participants' were significantly more similar to themselves than to other participants, even when across-subject representations were compared within the same story (LME range of β values = -0.01 – 0.03, all *p*s < 0.001).

We then examined whether there was a relationship between within- and across-subject similarity of word property representations. By correlating within- and across-subject similarity values across parcels, we found that brain areas with word property representations that were more similar within subjects also showed higher similarity in representations across subjects (loudness (r = 0.874), concrete-abstract (r = 0.784), frequency (r = 0.797), valence (r = 0.428), arousal (r = 0.599); all *p*s < 0.001). This finding recapitulates a seemingly paradoxical phenomenon previously shown in functional connectivity fingerprinting: brain states that make individuals more similar to others also make them more similar to themselves (Finn et al., 2017).

Individuals are identifiable from their representations of the concrete-abstract axis

The previous analyses revealed that individuals' representations of the concrete-abstract axis are stable across stories, but how *unique* are these representations? High reliability does not necessarily imply uniqueness: low average across-subject similarity could be due to high *variability* in across-subject similarity. In other words, certain pairs of subjects may have highly similar representations of the concrete-abstract axis, despite most of the group exhibiting low similarity. To test the extent to which word property representations are unique to each individual, we evaluated our ability to identify subjects from their representations of each word property.

Across cortical parcels, we were able to identify subjects from representations of both sensory response (loudness) and all four linguistic properties across much of the brain (Figure 3C; null = 10,000 permutations, all $q_{FDR} < 0.05$). Of note, the average identification rates across cortical parcels were low in an absolute sense but still significantly above chance (chance = 2.22%; Figure 3D). Overall, representations of loudness provided the best ability to identify subjects (22.1%), demonstrating significantly higher identification rates, on average, than the concrete-abstract axis (16.5%; $\beta = 10.41$, $t(948) = 14.77$, $p < 0.001$). However, representations of the concrete-abstract axis enabled significantly higher identification accuracy than representations of other linguistic properties (frequency: 8.8%, β = 2.9, $t(948) = 4.11$; valence: 4.4%, β = 7.24, $t(948) = 10.27$; arousal: 6.6%, β = 5.08, $t(948) = 7.2$; all ps < 0.001).

We then applied a winner-takes-all approach to identifiability maps to understand the cortical parcels where concrete-abstract axis representations showed the highest accuracy out of all word properties. We found that the concrete-abstract axis enabled the highest identification of subjects—even higher than loudness—within regions including left anterior temporal lobe, left inferior frontal gyrus, and bilateral retrosplenial cortex (RSC). These results dovetail with previous studies that have shown that areas within the left-lateralized language network and multimodal cortex are important in representing concrete and abstract concepts (Binder et al., 2005; Roxbury et al., 2014; J. Wang et al., 2010; Zhang et al., 2020).

Representations of concreteness are more reliable than representations of abstractness and drive individual identifiability

Thus far, we have shown that representations of the concrete-abstract axis are reliable within and unique to individual subjects across experiences. Yet it remains unclear whether both ends of this continuum – concreteness and abstractness – contribute equally this reliability and uniqueness.

On one hand, representations of concreteness may be more reliable than those of abstractness due to greater associations with sensory experience. On the other hand, representations of abstractness may be more idiosyncratic, as uniquely language-based representations could depend more heavily on individual experience to create meaning. While prior work suggests that representations of abstract words exhibit lower similarity across individuals than concrete words, disentangling the source of this difference requires 1) evaluating the stability of concreteness and abstractness as classes, and 2) assessing similarity within the same individual across experiences.

To understand the differential contributions of concreteness and abstractness in driving reliability, we dichotomized the continuous concrete-abstract axis and estimated reliability separately for each end of the spectrum. Specifically, we first limited our analysis to nouns to avoid confounds associated with different parts of speech, as verbs are more prevalent at the abstract end of the axis. We then separated the top 30% of words at each end of the concrete-abstract axis into two classes representing “concreteness” and “abstractness”. Lastly, we used a GLM to estimate separate BOLD response patterns for “concreteness” and “abstractness”.

We observed that representations of concreteness and abstractness each demonstrated significant reliability across stories in several brain regions (Figure 4B; null = 10,000 permutations, both $q_{FDR} < 0.05$). By contrasting the reliability maps, we found that many cortical parcels (36%, or 72/200) exhibited more reliable responses to concreteness than abstractness. On the other hand, no parcels showed greater reliability for representations of abstractness over concreteness. We then repeated our identifiability analysis (see Methods) to understand whether these representations of concreteness and abstractness were unique enough to discriminate individual subjects from one another. Across the majority of parcels, we were able to identify individuals based on their representations of both concreteness and abstractness significantly above chance (Figure 4C; null = 10,000 permutations, both $q_{FDR} < 0.05$). However, at a whole-brain level, representations of concreteness showed a significantly higher rate of identification compared to representations of abstractness (Figure 4D; concreteness: 14%; abstractness: 6.4%; $\beta = 3.83$, $t(190) = 12.79$; $p < 0.001$). Together, these findings suggest that representations of concreteness primarily drive reliable responses of the concrete-abstract axis and are more individualized than representations of abstractness, extending previous, population-level findings to individual patterns of neural responses (Binder et al., 2005; Roxbury et al., 2014; Tong et al., 2022; X. Wang & Bi, Yanchao, 2021; West & Holcomb, 2000).

Concrete concepts share an underlying representational signature that drives reliability of representations across experiences

Why might neural representations of the concrete end of the spectrum be more reliable than representations of the abstract end? One potential explanation is that concrete words share the property of imageability, which carries its own representational signature that undergirds the representations of individual concrete words despite their differences in meaning. This representational signature could serve to stabilize the representations of individual concrete words across different contexts and in relation to other concrete words. While the naturalistic nature of these stimuli means that we did not necessarily have repeated presentation of the *same* word(s) across stories, we can use natural language processing (NLP) techniques to group words into clusters of semantically related words and use these clusters to help understand why representations of concreteness are more reliable than those of abstractness, even when generalizing over individual words and concepts.

Numerous recent studies have demonstrated parallels in language representation between humans and NLP models (Caucheteux & King, 2022; Goldstein et al., 2022; Huth et al., 2016; Schrimpf et al., 2021; Tuckute et al., 2024). Here, we used a word-embedding NLP model (GloVe; (Pennington et al., 2014) to understand how the semantic relationships among concrete and abstract words relate to the reliability of representations of the concrete-abstract axis. Specifically, we embedded concrete and abstract words within a high-dimensional semantic space and clustered words based on their semantic similarity. We then analyzed the similarity of these "concept clusters" in semantic space and, analogously, the similarity of neural responses to each cluster across stories using linear mixed-effects models (see Methods).

The semantic-embedding analysis confirmed that words within the same concept cluster were more similar to each other than to words in different clusters (Figure 5C; , β = 0.03, $t(610) = 14.71$, $p < 0.001$), a pattern of results consistent across both concrete and abstract clusters (pairwise comparisons; concrete: $t(306) = 10.76$; abstract: $t(306) = 10.03$; both *p*s < 0.001). This was expected given that the clustering was performed on semantic distances, but still served as a useful check on the appropriateness of the cluster solution. But we also observed a somewhat puzzling result: within semantic space, abstract clusters were generally more similar to one another than concrete clusters were to one another (β = 0.03, $t(610) = 5.87$, $p < 0.001$). This finding was particularly surprising given the results from the previous analysis (cf. Figure 4B) that showed that neural representations of concreteness are more reliable than representations of abstractness. Why might the concrete end of the spectrum, which encompasses *more* variability in (i.e., spans more of) semantic space, show *less* variability in its neural representations?

We next turned to analyze within-subject neural representations of concrete and abstract concept clusters. Echoing the results in semantic space, representations of words within the same cluster were more similar across stories than representations of words in different clusters (Figure 5D; $\beta = 0.007$, $t(34373) = 20.04$, $p < 0.001$), and this was true for both the concrete and abstract ends of the spectrum (concrete $z = 4.36$, abstract $z = 23.99$, both ps < 0.001). In contrast to the similarity of clusters in semantic space (Figure 5C), neural representations of concrete clusters exhibited greater similarity than abstract clusters regardless of semantic distance (same or different clusters; $\beta = 0.01$, $t(34373) = 29.45$, $p < 0.001$; Figure 5D).

Critically, there was also an interaction such that the similarity advantage for same- over different-cluster representation was smaller for concrete clusters than for abstract clusters ($\beta = -0.005$, $t(34373) = -13.88$, $p < 0.001$). Strikingly, neural representations of *different* concrete clusters were more similar within subjects across stories than neural representations of the *same* abstract cluster (Figure 5D; mean difference = 0.007, $z = 7.12$, $p < 0.001$). Furthermore, this pattern of results persisted when analyzing similarity *across* subjects (within > across: $\beta = 0.002$, $t(34373) = 24.11$; concrete > abstract: $\beta = 0.001$, $t(34373) = 17.07$; interaction: $\beta = -0.001$, $t(34373) = -13.27$; all ps < 0.001; data not shown), suggesting that a consistent principle drives how concreteness is represented across similar words, within individuals and across the population.

Considered together, neural representations of semantically similar concrete words were more alike than those of semantically similar abstract words, despite concrete words spanning greater distances within semantic space than abstract words. These divergent results between the NLP model and neural data suggest that concrete words share a representational signature beyond linguistic representations due to sensory associations that could stem from integrating visual information into the neural representations.

Discussion

Word meanings vary across both people and contexts, often informed by conceptual associations specific to the individual as well as different situations in which the word is used. What linguistic properties provide a stable foundation for conceptual knowledge while simultaneously supporting unique, individual experience? Here, we found that the concrete-abstract axis provides a basis for both population stability and individual variability in the representation of natural language.

Many studies have demonstrated that while both concrete and abstract words evoke responses within the language network (Binder et al., 2005; Del Maschio et al., 2021; Friederici et al., 2000; Moseley & Pulvermüller, 2014), concrete words exhibit stronger and longer-lasting responses (Barber et al., 2013; Vignali et al., 2023; West & Holcomb, 2000) and also engage multimodal cortices, such as bilateral angular gyrus, posterior cingulate, and precuneus, more than abstract words (Binder et al., 2005; Roxbury et al., 2014; Tang et al., 2021; J. Wang et al., 2010; Zhang et al., 2020). In our study, we assessed whether reliability exists uniformly across the concrete-abstract axis, enabling us to understand if previously observed variability in abstract word representations can be explained by variability in representations of abstractness itself. We found reliable representations of the concrete-abstract axis within regions related to the language network and within multimodal cortex that were unique to individual subjects across diverse, naturalistic stories. Critically, representations of the concrete-abstract axis were more reliable than representations of other linguistic properties (i.e., frequency, valence, arousal), and this effect was driven primarily by the stable representations of the concrete end of the axis. Together, our results suggest that word representations are stabilized by consistent representations of concreteness more so than abstractness, potentially due to the engagement of multimodal areas known to integrate sensory and linguistic information.

Traditionally, neural representations of language have been probed by presenting participants with single words, sentences, and short paragraphs (Bookheimer, 2002; Hagoort, 2019). These studies have revealed neural territory specific to language (Fedorenko et al., 2011; Malik-Moraleda et al., 2022) that closely interacts with other networks involved in cognitive control and theory of mind (Fedorenko & Thompson-Schill, 2014; Paunov et al., 2019, 2022). In contrast to these carefully controlled experiments, everyday language is dynamic and contextualized, such that the meanings of words and sentences are informed by larger narrative structure (L. S. Hamilton & Huth, 2020; Willems et al., 2020). It is therefore crucial to evaluate the degree to which findings of carefully controlled studies extend to naturalistic language perception (Nastase et al., 2020). Within the present study, participants were presented with naturalistic auditory narratives representative of how language is used in day-to-day life. Importantly, we found that representations of abstractness, as well as clusters of related abstract words, were more variable both within and across subjects than representations of concrete words.

The finding of higher across-subject variability for abstractness aligns with another recent study that used a single-word paradigm to study abstract words (X. Wang & Bi, Yanchao, 2021); the authors of that study interpreted this heightened variability as reflecting individual differences in meaning of abstract words in particular. However, the appeal to individual differences implies a stability of representations *within* the same subject over time, which was not tested. Our study differs from this previous work in two ways: first, we examined neural representations to the concrete-abstract axis across words within distinct, naturalistic stories, and second, we evaluated the reliability of representations *within* subjects, *across* stories to understand if abstractness is idiosyncratically represented. We found that compared to representations of concreteness, representations of abstractness were more variable not only across subjects, but also within the same individual across distinct experiences. This suggests that variability in abstract words stems less from individual differences in meaning and more from a general instability of representations of abstractness.

Recent developments in natural language processing (NLP) models have provided researchers with tools to better investigate how the human brain organizes and processes natural language (Caucheteux & King, 2022; Goldstein et al., 2022; Huth et al., 2016; Schrimpf et al., 2021; Tuckute et al., 2024). These computational models not only capture semantic relationships between words, but also contain rich knowledge regarding how words relate within various contexts (Erk, 2012). Importantly, the contextual relationships between concrete words — that a fish and a whale may be semantically similar in terms of "wetness" but different in terms of "size" — closely correspond to human judgements of the same categories (Grand et al., 2022). Yet, within our study, we found that clusters of concrete words were less similar than clusters of abstract words within an NLP model but *more* similar in the human brain. This dissociation supports theories of grounded cognition that suggest representations of concreteness carry additional information beyond pure linguistic representation (Altarriba et al., 1999; Tuckute et al., 2018). Indeed, recent computational work has demonstrated that visual grounding is essential for linguistic representations to capture human ratings of the concrete-abstract axis (Zhang et al., 2021). While prior work has revealed subsets of abstract words that also exhibit sensory associations (Barsalou & Wiemer-Hastings, 2005; Ghio et al., 2013; Kiefer & Harpaintner, 2020), the lower similarity of abstract words even within a concept cluster suggests that the representational signature of sensory experience may be weaker or not present for abstract words. Together, these findings suggest that concrete words, but not abstract words, carry a shared signature of sensory grounding that stabilizes their neural representations both within and across subjects.

Though our work aligns with and extends past work on the concrete-abstract axis, it has some limitations. First, it is possible that we have underestimated the extent to which neural representations of the other properties (valence, arousal, frequency) are also idiosyncratic. In the current study, we leveraged pre-existing human ratings of these properties, but these behavioral ratings were collected by presenting participants with individual words out of context. Similarly, we leveraged an NLP model that does not incorporate contextual information into the word-level representations. Some of these other properties, especially valence and arousal, may be more context-dependent and require ratings specific to a given story or individual to understand the idiosyncrasies in neural representations. In addition, the moderate negative relationship between the concrete-abstract axis and word frequency in our dataset also leaves open the possibility that some effects attributed to concreteness may be shared with (inverse) frequency. Second, due to the diversity of content across the auditory narratives, we were limited in our ability to compare representations of the *same* words across stories. We addressed this by comparing the neural representations of clusters of similar words across stories, extending prior work on single words to the organization of broader concepts in semantic space. Future work could select stories that contain the same words but vary in narrative content to understand the stability of both specific words and semantic organization more generally across experiences.

In sum, our work establishes the concrete-abstract axis as a critical dimension for promoting both shared and individualized representations of language. In particular, these findings disentangle the sources of individual variability of concrete and abstract word representation and reveal a representational signature of sensory experience specific to concrete words that boosts their representational stability. Our results underscore the importance of considering within-subject variability when identifying underlying drivers of common versus idiosyncratic processing of natural language.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Aka, A., Phan, T. D., & Kahana, M. J. (2021). Predicting recall of words and lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(5), 765–784. https://doi.org/10.1037/xlm0000964

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

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

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating Abstract Concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition* (1st ed., pp. 129–163). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.007

Bazeille, T., DuPre, E., Richard, H., Poline, J.-B., & Thirion, B. (2021). An empirical evaluation of functional alignment using inter-subject decoding. *NeuroImage*, *245*, 118683. https://doi.org/10.1016/j.neuroimage.2021.118683

Bedny, M., & Caramazza, A. (2011). Perception, action, and word meanings in the human brain: The case from action verbs: Word meanings and sensory experience. *Annals of the New York Academy of Sciences*, *1224*(1), 81–95. https://doi.org/10.1111/j.1749-6632.2011.06013.x

Bi, Y. (2021). Dual coding of knowledge in the human brain. *Trends in Cognitive Sciences*, *25*(10), 883–895. https://doi.org/10.1016/j.tics.2021.07.006

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct Brain Systems for Processing Concrete and Abstract Concepts. *Journal of Cognitive Neuroscience*, *17*(6), 905–917. https://doi.org/10.1162/0898929054021102

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python* (1st ed). O'Reilly.

Bookheimer, S. (2002). Functional MRI of Language: New Approaches to Understanding the Cortical Organization of Semantic Processing. *Annual Review of Neuroscience*, *25*(1), 151–188. https://doi.org/10.1146/annurev.neuro.25.112701.142946

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, *143*(3), 263–292. https://doi.org/10.1037/bul0000089

Brooks, J. A., Shablack, H., Gendron, M., Satpute, A. B., Parrish, M. H., & Lindquist, K. A. (2016). The role of language in the experience and perception of emotion: A neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience*, nsw121. https://doi.org/10.1093/scan/nsw121

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

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

Caucheteux, C., & King, J.-R. (2022). Brains and algorithms partially converge in natural language processing. *Communications Biology*, *5*(1), 134. https://doi.org/10.1038/s42003-022-03036-1

Chen, P.-H. (Cameron), Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., & Ramadge, P. J. (2015). A Reduced-Dimension fMRI Shared Response Model. *Advances in Neural Information Processing Systems*, *28*. https://proceedings.neurips.cc/paper_files/paper/2015/hash/b3967a0e938dc2a6340e258630febd5a-Abstract.html

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Del Maschio, N., Fedeli, D., Garofalo, G., & Buccino, G. (2021). Evidence for the Concreteness of Abstract Language: A Meta-Analysis of Neuroimaging Studies. *Brain Sciences*, *12*(1), 32. https://doi.org/10.3390/brainsci12010032

Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *The Journal of Neuroscience*, *39*(39), 7722–7736. https://doi.org/10.1523/JNEUROSCI.0675-19.2019

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301–306. https://doi.org/10.1016/j.tics.2004.05.003

Elman, J. L. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

Erk, K. (2012). Vector Space Models of Word Meaning and Phrase Meaning: A Survey: Vector Space Models of Word and Phrase Meaning. *Language and Linguistics Compass*, *6*(10), 635–653. https://doi.org/10.1002/lnco.362

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428–16433. https://doi.org/10.1073/pnas.1112937108

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, *18*(3), 120–126. https://doi.org/10.1016/j.tics.2013.12.006

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences*, *119*(6), e2108091119. https://doi.org/10.1073/pnas.2108091119

Fiebach, C. J., Friederici, A. D., Müller, K., & Cramon, D. Y. V. (2002). fMRI Evidence for Dual Routes to the Mental Lexicon in Visual Word Recognition. *Journal of Cognitive Neuroscience*, *14*(1), 11–23. https://doi.org/10.1162/089892902317205285

Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*, *160*, 140–151. https://doi.org/10.1016/j.neuroimage.2017.03.064

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Friederici, A. D., Opitz, B., & Von Cramon, D. Y. (2000). Segregating Semantic and Syntactic Aspects of Processing in the Human Brain: An fMRI Investigation of Different Word Types. *Cerebral Cortex*, *10*(7), 698–705. https://doi.org/10.1093/cercor/10.7.698

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

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., … Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, *25*(3), 369–380. https://doi.org/10.1038/s41593-022-01026-4

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, *61*(1), 23–29. https://doi.org/10.1037/h0040561

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-022-01316-8

Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, *366*(6461), 55–58. https://doi.org/10.1126/science.aax0289

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *35*(5), 573–582. https://doi.org/10.1080/23273798.2018.1499946

Hamilton, M., & Rajaram, S. (2001). The Concreteness Effect in Implicit and Explicit Memory Tests. *Journal of Memory and Language*, *44*(1), 96–117. https://doi.org/10.1006/jmla.2000.2749

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(2), 110–126. https://doi.org/10.3758/CABN.6.2.110

Kiefer, M., & Harpaintner, M. (2020). Varieties of abstract concepts and their grounding in perception or action. *Open Psychology*, *2*(1), 119–137. https://doi.org/10.1515/psych-2020-0104

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

Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., Chen, P.-H. C., Ellis, C. T., Henselman-Petrusek, G., Huberdeau, D., Hutchinson, J. B., Li, Y. P., Lu, Q., Manning, J. R., Mennen, A. C., Nastase, S. A., Richard, H., Schapiro, A. C., Schuck, N. W., … Norman, K. A. (2021). BrainIAK: The Brain Imaging Analysis Kit. *Aperture Neuro*, *1*(4), 10.52294/31bb5b68-2184-411b-8c00-a1dacb61e1da. https://doi.org/10.52294/31bb5b68-2184-411b-8c00-a1dacb61e1da

Maddock, R. J., & Buonocore, M. H. (1997). Activation of left posterior cingulate gyrus by the auditory presentation of threat-related words: An fMRI study. *Psychiatry Research: Neuroimaging*, *75*(1), 1–14. https://doi.org/10.1016/S0925-4927(97)00018-8

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience*. https://doi.org/10.1038/s41593-022-01114-5

McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (arXiv:1802.03426). arXiv. http://arxiv.org/abs/1802.03426

Montefinese, M. (2019). Semantic representation of abstract and concrete words: A minireview of neural evidence. *Journal of Neurophysiology*, *121*(5), 1585–1587. https://doi.org/10.1152/jn.00065.2019

Moseley, R. L., & Pulvermüller, F. (2014). Nouns, verbs, objects, actions, and abstractions: Local fMRI activity indexes semantics, not lexical categories. *Brain and Language*, *132*, 28–42. https://doi.org/10.1016/j.bandl.2014.03.001

Musz, E., & Thompson-Schill, S. L. (2015). Semantic variability predicts neural variability of object concepts. *Neuropsychologia*, *76*, 41–51. https://doi.org/10.1016/j.neuropsychologia.2014.11.029

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

Nastase, S. A., Liu, Y.-F., Hillman, H., Zadbood, A., Hasenfratz, L., Keshavarzian, N., Chen, J., Honey, C. J., Yeshurun, Y., Regev, M., Nguyen, M., Chang, C. H. C., Baldassano, C., Lositsky, O., Simony, E., Chow, M. A., Leong, Y. C., Brooks, P. P., Micciche, E., … Hasson, U. (2021). The “Narratives” fMRI dataset for evaluating models of naturalistic language comprehension. *Scientific Data*, *8*(1), 250. https://doi.org/10.1038/s41597-021-01033-3

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

Paivio, A., & Begg, I. (1971). Imagery and comprehension latencies as a function of sentence concreteness and structure. *Perception & Psychophysics*, *10*(6), 408–412. https://doi.org/10.3758/BF03210323

Paunov, A. M., Blank, I. A., & Fedorenko, E. (2019). Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *Journal of Neurophysiology*, *121*(4), 1244–1265. https://doi.org/10.1152/jn.00619.2018

Paunov, A. M., Blank, I. A., Jouravlev, O., Mineroff, Z., Gallée, J., & Fedorenko, E. (2022). Differential Tracking of Linguistic vs. Mental State Content in Naturalistic Stimuli by Language and Theory of Mind (ToM) Brain Networks. *Neurobiology of Language*, 1–29. https://doi.org/10.1162/nol_a_00071

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nature Neuroscience*, *24*(11), 1628–1636. https://doi.org/10.1038/s41593-021-00921-6

Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness Effects in Different Tasks: Implications for Models of Short-Term Memory. *Quarterly Journal of Experimental Psychology*, *61*(2), 292–323. https://doi.org/10.1080/17470210601147747

Roxbury, T., McMahon, K., & Copland, D. A. (2014). An fMRI study of concreteness effects in spoken word recognition. *Behavioral and Brain Functions*, *10*(1), 34. https://doi.org/10.1186/1744-9081-10-34

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex (New York, NY)*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, *118*(45), e2105646118. https://doi.org/10.1073/pnas.2105646118

Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., & Richlan, F. (2016). Words in Context: The Effects of Length, Frequency, and Predictability on Brain Responses During Natural Reading. *Cerebral Cortex*, *26*(10), 3889.2-3904. https://doi.org/10.1093/cercor/bhw184

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

Shi, J., & Malik, J. (2000). Normalized Cuts and Image Segmentation. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *22*(8).

Stolk, A., Verhagen, L., & Toni, I. (2016). Conceptual Alignment: How Brains Achieve Mutual Understanding. *Trends in Cognitive Sciences*, *20*(3), 180–191. https://doi.org/10.1016/j.tics.2015.11.007

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences*, *113*(1), 194–199. https://doi.org/10.1073/pnas.1511905112

Tang, J., LeBel, A., & Huth, A. G. (2021). *Cortical Representations of Concrete and Abstract Concepts in Language Combine Visual and Linguistic Representations* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.05.19.444701

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*(10), 1029–1038. https://doi.org/10.1038/s41562-020-0924-8

Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A Distributed Network for Multimodal Experiential Representation of Concepts. *The Journal of Neuroscience*, *42*(37), 7121–7130. https://doi.org/10.1523/JNEUROSCI.1243-21.2022

Tuckute, G., Mahowald, K., Isola, P., Oliva, A., Gibson, E., & Fedorenko, E. (2018). *Intrinsically memorable words have unique associations with their meanings* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/p6kv9

Tuckute, G., Sathe, A., Srikant, S., Taliaferro, M., Wang, M., Schrimpf, M., Kay, K., & Fedorenko, E. (2023). *Driving and suppressing the human language network using large language models* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.04.16.537080

Tuckute, G., Sathe, A., Srikant, S., Taliaferro, M., Wang, M., Schrimpf, M., Kay, K., & Fedorenko, E. (2024). Driving and suppressing the human language network using large language models. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01783-7

Vignali, L., Xu, Y., Turini, J., Collignon, O., Crepaldi, D., & Bottini, R. (2023). Spatiotemporal dynamics of abstract and concrete semantic representations. *Brain and Language*, *243*, 105298. https://doi.org/10.1016/j.bandl.2023.105298

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256–1271. https://doi.org/10.1037/0278-7393.25.5.1256

Wang, J., Conder, J. A., Blitzer, D. N., & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping*, *31*(10), 1459–1468. https://doi.org/10.1002/hbm.20950

Wang, X. & Bi, Yanchao. (2021). Idiosyncratic Tower of Babel: Individual Differences in Word-Meaning Representation Increase as Word Abstractness Increases. *Psychological Science*, *32*(10), 1617–1635. https://doi.org/10.1177/09567976211003877

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

West, W. C., & Holcomb, P. J. (2000). Imaginal, Semantic, and Surface-Level Processing of Concrete and Abstract Words: An Electrophysiological Investigation. *Journal of Cognitive Neuroscience*, *12*(6), 1024–1037. https://doi.org/10.1162/08989290051137558

Willems, R. M., Nastase, S. A., & Milivojevic, B. (2020). Narratives for Neuroscience. *Trends in Neurosciences*, *43*(5), 271–273. https://doi.org/10.1016/j.tins.2020.03.003

Yates, T. S., Ellis, C. T., & Turk-Browne, N. B. (2021). Emergence and organization of adult brain function throughout child development. *NeuroImage*, *226*, 117606. https://doi.org/10.1016/j.neuroimage.2020.117606

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Zhang, Y., Choi, M., Han, K., & Liu, Z. (2021). *Explainable Semantic Space by Grounding Language to Vision with Cross-Modal Contrastive Learning* (arXiv:2111.07180). arXiv. http://arxiv.org/abs/2111.07180

Zhang, Y., Han, K., Worth, R., & Liu, Z. (2020). Connecting concepts in the brain by mapping cortical representations of semantic relations. *Nature Communications*, *11*(1), 1877. https://doi.org/10.1038/s41467-020-15804-w

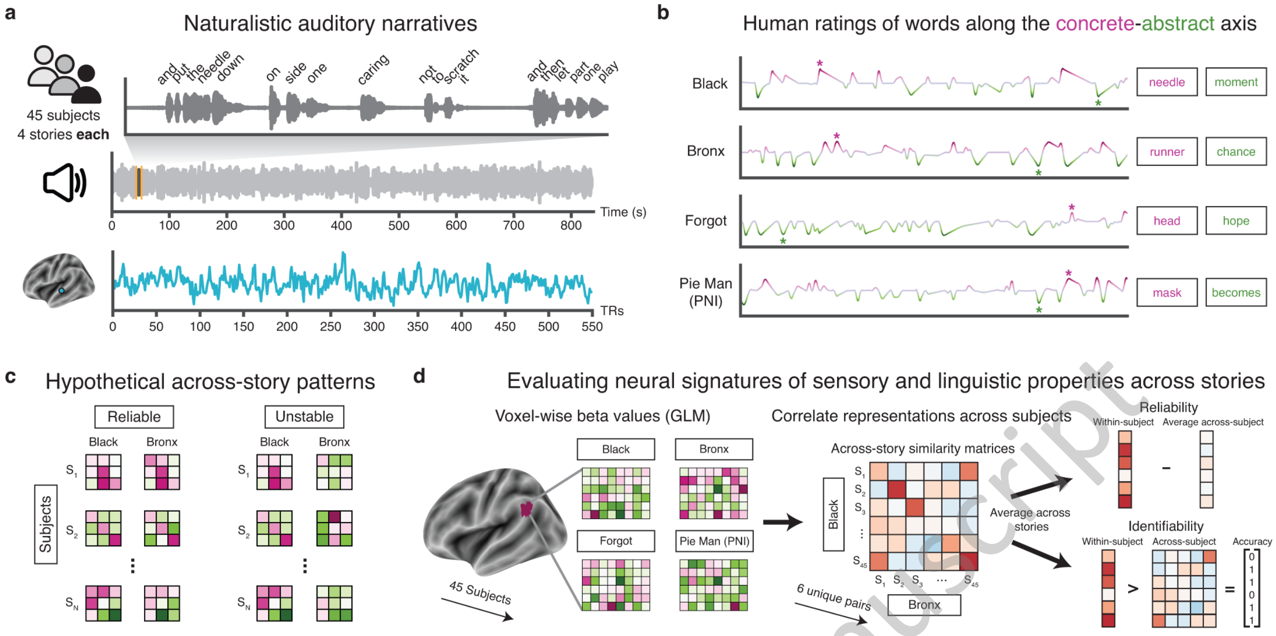
Figure 1. Experimental methods. (a) 45 subjects listened to four auditory stories during fMRI scanning (Nastase et al., 2021). **(b)** Human ratings were used to assign a continuous value of concreteness (i.e., position along the concrete-abstract axis) for as many words as possible within each story. This process was repeated with other linguistic properties including frequency, valence, and arousal (not shown). **(c)** Any apparent variation across subjects in neural representations of word properties could stem from two possible underlying patterns: neural representations could be reliably idiosyncratic within subjects, evidenced by high similarity of representations within the same subject across distinct experiences (here, stories), or these representations could be unstable both within and across subjects, evidenced by variability within the same subject across stories. **(d)** Example procedure for calculating reliability and identifiability for one word property. For each story, voxel-wise beta values were estimated within a generalized linear model. Then, within each of 200 parcels (Schaefer parcellation), beta values were correlated between all subjects for each pair of stories (6 unique pairs). These story similarity matrices were then averaged and used to estimate two indices of stable, individualized neural representations: 1) reliability, defined as the difference between within-subject and average across-subject similarity, and 2) identifiability, defined as the fingerprinting accuracy of discriminating one subject from all other subjects based on their neural representations. This process was repeated for each word property.

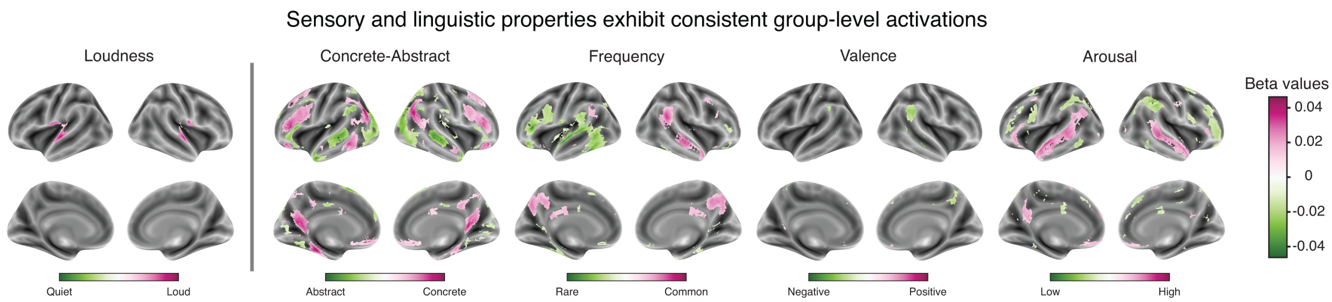
Figure 2. Group-level univariate activation to sensory and linguistic properties. Across stories and subjects, multiple regions exhibited significant activation to the intensity of sound and word-level linguistic properties including the concrete-abstract axis, frequency, valence, and arousal. Results shown are from a single linear mixed-effects model containing fixed effects for all properties plus random effects for story and subject. Results are displayed at a voxel-wise threshold of $q_{FDR} < 0.05$.

Figure 3. Within- and across-subject reliability of neural representations of word properties. We compared representations of word properties across four naturalistic stories both within and across subjects. **(a)** Across stories, all properties except valence exhibited high within-subject reliability across much of cortex ($q_{FDR} < 0.05$, null = 10,000 permutations). While a simple sensory property, loudness, exhibited the highest reliability, representations of the concrete-abstract axis were more reliable than other linguistic properties (frequency, valence, arousal). **(b)** At the whole-brain level, across all properties, within-subject across-story similarity was consistently higher than across-subject similarity, even when comparing representations across subjects within the same stories. Each data point represents average similarity value in one parcel of the Schaefer parcellation (200 total). **(c)** Representations of all properties enabled accurate identification of subjects across much of cortex. All plots are thresholded at chance (2.22%). **(d)** Out of tested linguistic properties, subjects were most identifiable from their representations of the concrete-abstract axis. Each dot indicates identifiability within one parcel. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; n.s. $p > 0.05$.

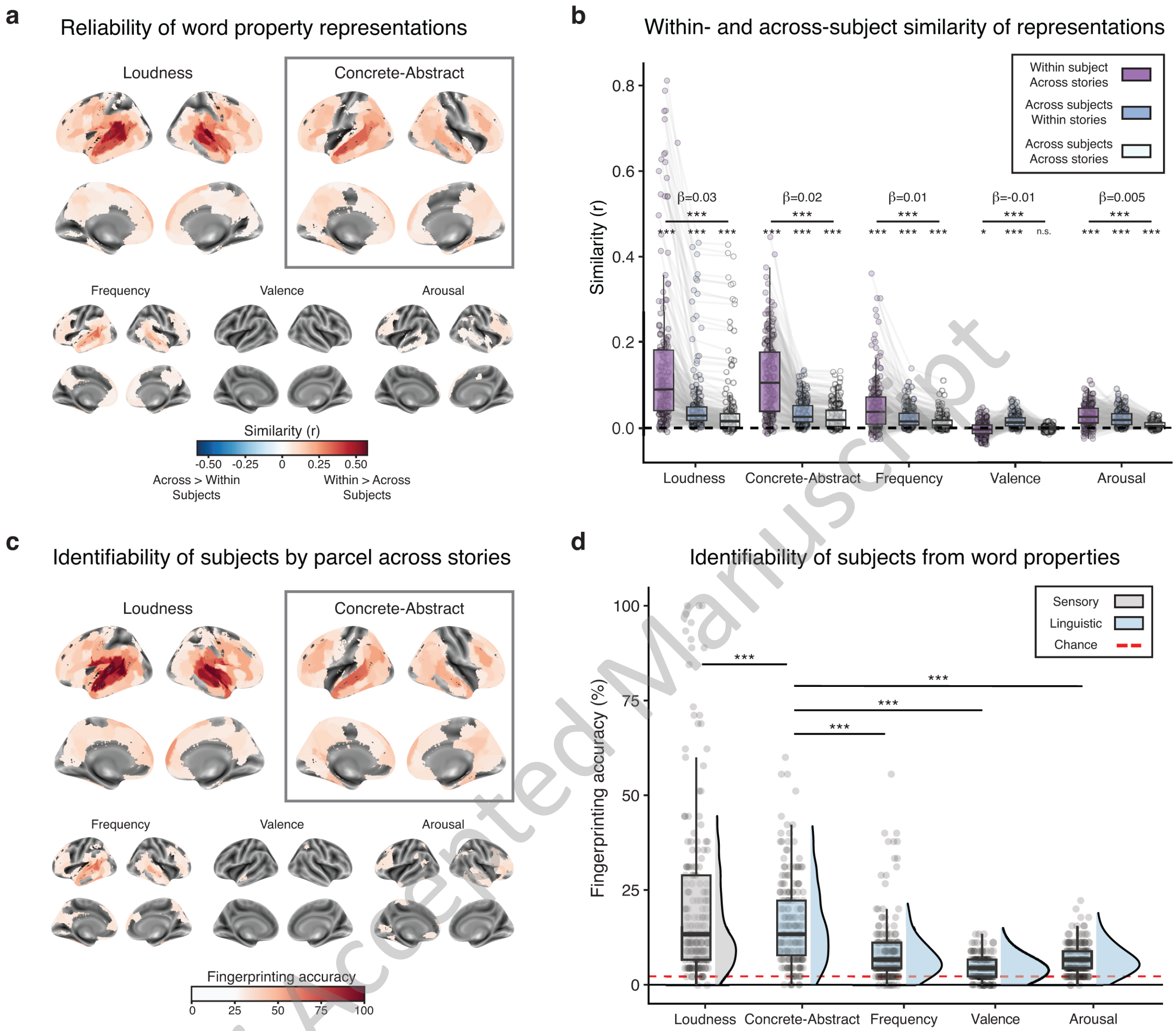
Figure 4. Within-subject reliability of neural representations of concrete and abstract words. (a) We selected concrete and abstract words as the top/bottom 30% of nouns within the concrete-abstract axis and estimated neural responses to each set of words in a second GLM analysis. **(b)** While both concreteness and abstractness exhibited reliable representations within subjects across stories, representations of concreteness were more reliable than representations of abstractness across much of cortex ($q_{FDR} < 0.05$, null = 10,000 permutations). **(c, d)** Representations of concreteness provided a greater ability to identify subjects than representations of abstractness ($q_{FDR} < 0.05$, null = 10,000 permutations).

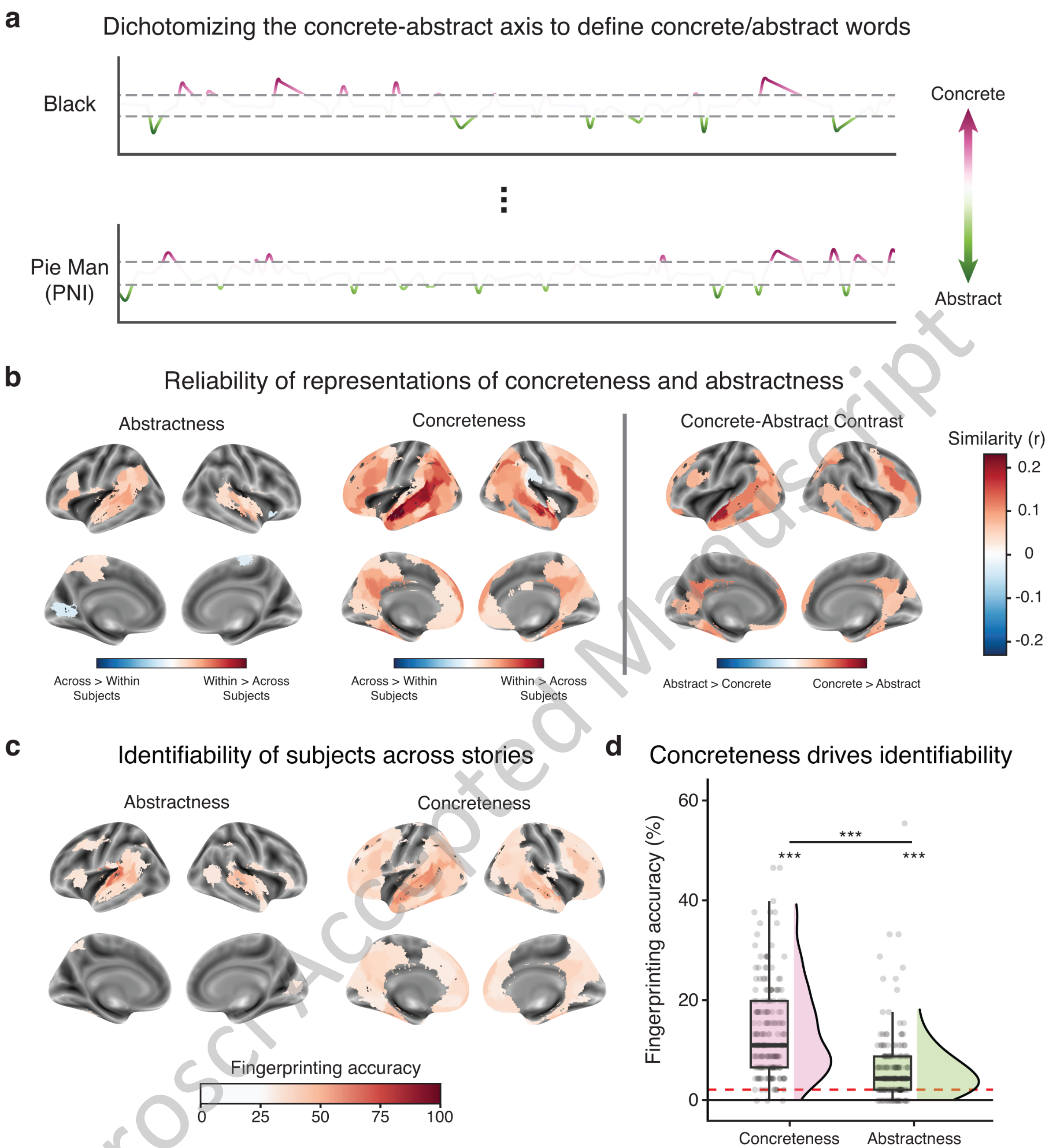
Figure 5. Stability of concrete and abstract concept cluster representations within and across subjects. (a) We clustered the top 30% concrete and top 30% abstract words within a high-dimensional semantic space (GloVe). We then estimated voxel-wise beta values for each of six clusters (3 concrete, 3 abstract) within each subject and story. Next, within each parcel (200 total), we correlated beta values between all sets of clusters across stories and averaged the across-story similarity of clusters. **(b)** Visualization of concept clusters within a 2-dimensional projection using UMAP, plus example words from each cluster. **(c)** Within semantic space, words within abstract clusters were more similar (i.e., less distant) than words within concrete clusters. Each dot represents the average similarity of a given word to other words within a given comparison. In contrast, **(d)** within-subject neural representations of concrete clusters were more similar across stories than representations of abstract clusters. Each dot indicates the average similarity of one subject's concept cluster representations within a given comparison. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; n.s. $p > 0.05$.

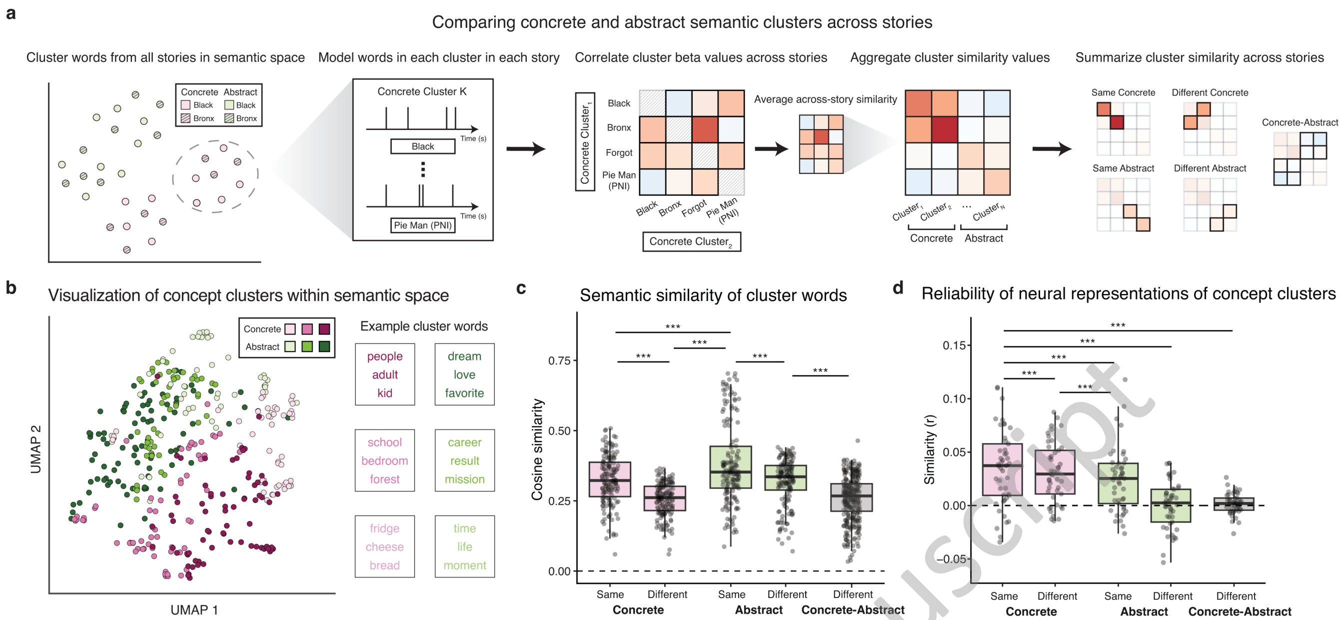

a
Naturalistic auditory narratives
45 subjects
4 stories each
and put the needle down on side one caring not to scratch it and then let part one play
Time (s)
0 100 200 300 400 500 600 700 800
TRs
0 50 100 150 200 250 300 350 400 450 500 550
b
Human ratings of words along the concrete-abstract axis
Black
needle
moment
Bronx
runner
chance
Forgot
head
hope
Pie Man (PNI)
mask
becomes
c
Hypothetical across-story patterns
Reliable
Unstable
Black
Bronx
Subjects
d
Evaluating neural signatures of sensory and linguistic properties across stories
Voxel-wise beta values (GLM)
Correlate representations across subjects
Black
Bronx
Forgot
Pie Man (PNI)
45 Subjects
Across-story similarity matrices
6 unique pairs
Average across stories
Reliability
Within-subject
Average across-subject
Identifiability
Across-subject
Accuracy



Sensory and linguistic properties exhibit consistent group-level activations
Loudness
Concrete-Abstract
Frequency
Valence
Arousal
Beta values
0.04
0.02
0
-0.02
-0.04
Quiet
Loud
Abstract
Concrete
Rare
Common
Negative
Positive
Low
High


a
Reliability of word property representations
Loudness
Concrete-Abstract
Frequency
Valence
Arousal
Similarity (r)
-0.50 -0.25 0 0.25 0.50
Across > Within Subjects
Within > Across Subjects
b
Within- and across-subject similarity of representations
Within subject Across stories
Across subjects Within stories
Across subjects Across stories
Similarity (r)
β=0.03
β=0.02
β=0.01
β=-0.01
β=0.005
n.s.
Loudness
Concrete-Abstract
Frequency
Valence
Arousal
c
Identifiability of subjects by parcel across stories
Loudness
Concrete-Abstract
Frequency
Valence
Arousal
Fingerprinting accuracy
0 25 50 75 100
d
Identifiability of subjects from word properties
Sensory
Linguistic
Chance
Fingerprinting accuracy (%)
Loudness
Concrete-Abstract
Frequency
Valence
Arousal


a
Dichotomizing the concrete-abstract axis to define concrete/abstract words
Black
Pie Man (PNI)
Concrete
Abstract
b
Reliability of representations of concreteness and abstractness
Abstractness
Concreteness
Concrete-Abstract Contrast
Similarity (r)
0.2
0.1
0
-0.1
-0.2
Across > Within Subjects
Within > Across Subjects
Abstract > Concrete
Concrete > Abstract
c
Identifiability of subjects across stories
Fingerprinting accuracy
0
25
50
75
100
d
Concreteness drives identifiability
Fingerprinting accuracy (%)
60
40
20
0
***
Concreteness
Abstractness



a
Comparing concrete and abstract semantic clusters across stories
Cluster words from all stories in semantic space
Model words in each cluster in each story
Correlate cluster beta values across stories
Aggregate cluster similarity values
Summarize cluster similarity across stories
Concrete Cluster K
Average across-story similarity
Same Concrete
Different Concrete
Concrete-Abstract
Same Abstract
Different Abstract
b
Visualization of concept clusters within semantic space
Example cluster words
people adult kid
dream love favorite
school bedroom forest
career result mission
fridge cheese bread
time life moment
UMAP 1
UMAP 2
c
Semantic similarity of cluster words
Cosine similarity
d
Reliability of neural representations of concept clusters
Similarity (r)