bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Main Manuscript for

The Features Underlying the Memorability of Objects

Max A. Kramer[1], Martin N. Hebart[2], Chris I. Baker[3], Wilma A. Bainbridge[1*]

1 – Department of Psychology, University of Chicago

2 – Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences

3 – Laboratory of Brain and Cognition, National Institute of Mental Health

*Corresponding Author: Wilma A. Bainbridge

Email: wilma@uchicago.edu

Competing Interest Statement: The authors do not have any competing interests.

Keywords: Memory, Feature Space, Typicality, Semantics

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Despite decades of study of memory, it remains unclear what makes an image memorable. There is considerable debate surrounding the underlying determinants of memory, including the roles of semantic (e.g., animacy, utility) and visual features (e.g., brightness) as well as whether the most prototypical or most atypical items are best remembered. Prior studies have relied on constrained stimulus sets, limiting any generalized view of the features that may contribute to memory. Here, we collected over one million memory ratings (N=13,946) for THINGS (Hebart et al., 2019), a naturalistic dataset of 26,107 object images designed to comprehensively sample concrete objects. First, we establish a model of object features that is predictive of image memorability, capturing over half of the explainable variance. For this model, we find that semantic features have a stronger influence than visual features on what people will remember. Second, we examined whether memorability could be accounted for by the typicality of the objects, by comparing human behavioral data, object feature dimensions, and deep neural network features. While prototypical objects tend to be the most memorable, the relationship between memorability and typicality is more complex than a simple positive or negative association and typicality alone cannot account for memorability.

SIGNIFICANCE STATEMENT

Why is it that we seem to remember and forget the same things? Our lived experiences differ, but there is remarkable consistency in what is remembered across people. Here, we collected memory performance scores for a comprehensive and diverse collection of natural object images to identify which properties determine our ability to remember. We created a model for predicting memory from object features showing that semantic properties more than visual properties contribute primarily to memorability. Further, we find that it is neither the most prototypical or atypical images that are best remembered, which suggests that typicality alone cannot account for memorability. Our findings challenge decades of prior research that suggest that the most atypical items are most memorable, informing our understanding of the features and organizational principles of memory.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

What is it that makes something memorable? Researchers have been struggling for decades to understand the determinants of memory and how information is encoded, processed, and retrieved in the brain. The majority of research in memory uses a subject-centric framework, attempting to understand the underlying processes of memory and individual differences across people. This subject-centric framework is motivated by the highly personal nature of memory, as everyone has their own experiences that influence what they will later remember. However, a complementary stimulus-centric framework has arisen out of the surprising finding that, despite our individual experiences, we largely remember and forget the same images (Isola et al., 2011; Bainbridge et al., 2013). This new stimulus-driven perspective allows for a targeted examination of *what* we remember, and *why*.

This stimulus-driven perspective has revealed that images have an intrinsic *memorability*, defined for a stimulus as the likelihood that any given person will remember that stimulus later (Bainbridge et al., 2019). By using aggregated task scores for each stimulus rather than individual participant responses, memorability for a given stimulus can be quantified, repeatedly demonstrating a high degree of consistency in what people remember (Isola et al., 2011; Bainbridge et al., 2019) across stimulus types (see Isola et al., 2011; Bainbridge et al., 2013; Borkin et al., 2013; Xie et al., 2020). These memorability scores can account for upwards of 50% of variance in memory task performance (Bainbridge et al., 2013) and demonstrate remarkable resiliency across tasks and robustness to attention and priming (Bainbridge, 2020). This high consistency allows one to make honed predictions about what people will remember, which could have far-reaching implications for fields including advertising, marketing, public safety (Bainbridge et al., 2019), patient care (Bainbridge, Berron, et al., 2019), and computer vision (Needell & Bainbridge, 2022). However, in spite of these high consistencies in what individuals remember, what specific factors determine the memorability of an image is still largely unknown.

Prior research has sought to explain memorability as either a proxy for a given stimulus feature such as attractiveness or brightness, while others have attempted to reduce memorability to a linear combination of features in a constrained stimulus set (Bainbridge et al., 2013; Isola et al., 2014). These studies mostly utilize faces or scenes as stimuli, and none of them have explained the majority of variance in memorability using these models. More recently, researchers have emphasized the importance of considering items in a multidimensional representational space, with memorability arising from the relative location of an item within that space (Lukavský & Děchtěrenko, 2017; Bainbridge, 2019; Koch et al., 2020). This theoretical framework has sparked debate about the roles of low-level visual features such as color and shape and semantic information such as animacy in determining what we remember and what we forget (Khosla, 2015; Jaegle et al., 2019; Madan, 2020; Xie et al., 2020). Additionally, researchers disagree on whether the most memorable items are the most prototypical items (Bainbridge, Dilks, & Oliva, 2017; Bainbridge & Rissman, 2018) or the most atypical items (Bylinskii et al., 2015; Lukavský & Děchtěrenko, 2017; Mosenzadeh et al., 2019). Thus, there is

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a lack of consensus surrounding the roles of visual and semantic features as well as typicality with regards to what we remember, necessitating a much broader and detailed investigation.

Here, we provide a comprehensive characterization of visual memorability across an exhaustive set of picturable object concepts in the American English language (THINGS database, Hebart et al., 2019). Specifically, we determine the object features that drive our memories. We collected over 1 million memory scores for all 26,107 images in the THINGS database, which we have made publicly available (https://osf.io/5a7z6/?view_only=675e901c176c4bec9c2540fc4981e5fe). We then leveraged three complementary measures—human judgments, multidimensional object features, and predictions from a deep convolutional neural network (CNN)—to examine the relationship of memorability to object typicality. We construct a feature model that is able to predict a majority of the variance in image memorability. Among those features, our results uncover a primacy of semantic over visual dimensions in what we remember. Further, while we find evidence of the most prototypical items being best remembered, our discovery of high variance in the relationship between memorability and typicality at multiple levels suggest that typicality alone cannot account for memorability.

RESULTS

To explore memorability across concrete objects, we collected memorability scores for the entire image corpus of the THINGS database of object images (Hebart et al., 2019) and uncovered a dispersion of memorability across the hierarchical levels of THINGS. We examined the roles of semantic and visual information by predicting memorability from semantic and visual features using multivariate regression, revealing that semantic dimensions contribute primarily to object memorability. We then analyzed multiple measures of object typicality along with the memorability scores and found a small but robust effect of the most prototypical items being best remembered.

THINGS is a hierarchically structured dataset containing 26,107 images representing 1,854 object *concepts* (such as *aardvark*, *tank*, and *zucchini*) derived from a lexical database of picturable objects in the English language (see Methods), 1,619 of which are assigned to 27 higher *categories* (such as *animal*, *weapon*, and *food*). The concepts were assigned to categories in prior work through a two-stage process where one group of participants proposed categories for a given concept while a second group narrowed the potential categories further, with the most consistently chosen category becoming the assigned category for the concept (Hebart et al., 2020). The concepts and images are also characterized by an object space consisting of 49 dimensions that capture 92.25% of the variance in human behavioral similarity judgments of the objects (Hebart et al., 2020). Each concept and each image thus can be described by a 49-dimensional embedding that corresponds to the representation of that item in the object space. This overall dataset structure enables the analysis of memorability at the image, concept, category, and dimensional levels.

Memorability is Highly Variable Across Objects

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In order to quantify memorability for all 26,107 images in THINGS, we conducted a continuous recognition memory task (N = 13,946) administered over the online experiment platform Amazon Mechanical Turk (AMT) wherein participants viewed a stream of images and were asked to press a key when they recognized a repeated image that occurred after a delay of at least 60 seconds. Memorability was quantified as the corrected recognition (CR) score for a given image, calculated as the proportion of correct identifications of the image minus the proportion of false alarms on that image (Bainbridge & Rissman, 2018). The overall pattern of results remains unchanged when corrected recognition is instead substituted with hit rate or false alarm rate (Supplementary Material). To test if we observe consistency across people in what they remember and forget, we conducted a split-half consistency analysis across 1,000 iterations and found significant agreement in what independent groups of participants remembered (Spearman-Brown corrected split-half rank correlation, mean $\rho = 0.449$, $p < .001$), which is striking given the diversity of the THINGS images. This consistency in memory performance demonstrates that memorability can be considered an intrinsic property of these stimuli.

When assessing memorability at the concept level (e.g. *candy bars*, *windshields*), we observe that memorability varied widely across the concepts (Figure 1a). This dispersion of CR suggests that not all concepts in THINGS are equally memorable. For example, *candy bars* were highly memorable overall with a maximum CR of 1, a mean of 0.873, and a minimum of 0.756 (range = 0.127), while *windshields* were less memorable with a maximum CR of 0.756, a mean of 0.649, and a minimum of 0.404 (range = 0.352). We observe a similar diversity of memorability patterns at the higher category level (e.g. *dessert*, *part of car*; Figure 1b). The average CR across the THINGS categories is 0.793, with some categories demonstrating a higher average memorability than others; *body parts* attained the highest average memorability at 0.855 while *part of car* had the lowest average memorability of 0.753. These measures highlight the rich variation present within the THINGS database as it relates to memorability.

The previously reported embeddings along 49 dimensions for each of the object concepts (Hebart et al., 2020) allow us to determine if certain dimensions are more strongly reflected in memorable stimuli (Figure 1c). Specifically, we examined Spearman rank correlations between the memorability of the THINGS concepts and the concepts' embedding values for each of the 49 dimensions. We found that 36 dimensions showed a significant relationship to memorability (FDR-corrected $q < 0.01$), of which 9 were positive and the remaining 27 were negative. These correlations reveal that some properties used to characterize an object do show a relationship to memorability. For example, the positive relationship for the *body / body part* dimension ($\rho = 0.257$, $p = 1.873 \times 10^{-29}$) indicates that stimuli related to body parts tend to be more memorable, while a negative correlation like *metal / tools* ($\rho = -0.323$, $p = 1.689 \times 10^{-15}$) implies that stimuli made of metal tend to be less memorable.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

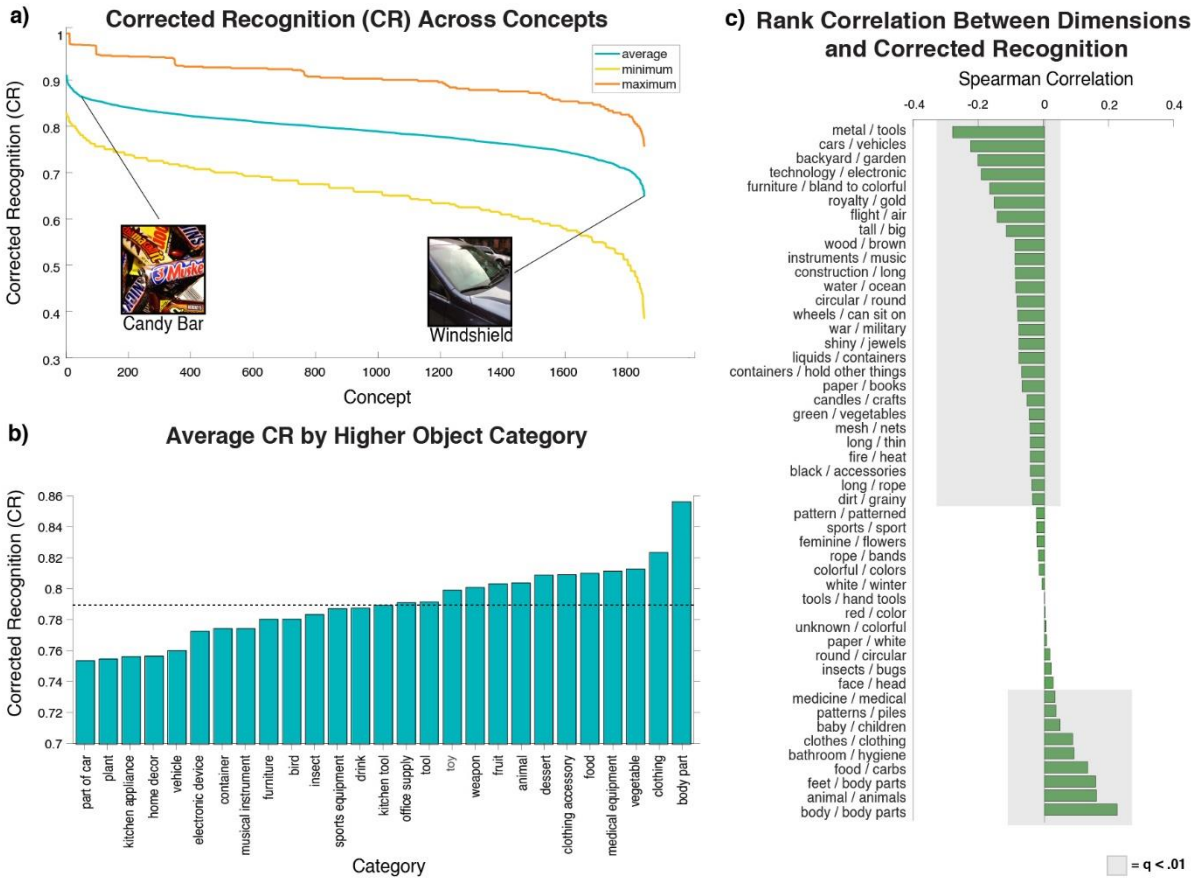


Figure 1. Descriptive analyses of memorability across the concept and category levels of the THINGS database as well as the 49 object dimensions. (A) The spread of corrected recognition (CR) across the 1,854 object concepts revealed that not all concepts are equally memorable. For concepts like *candy bars*, the entire range of component image memorability values were contained above the average value for a concept like *windshields*. (B) Visualizing the same spread across higher order categories revealed variation in average memorability across the 27 categories, with some categories including *part of car* displaying a CR score below the overall average memorability of 0.793 represented by the dotted horizontal line while others like *body parts* displayed a score above the average. (C) This high variability in memorability continues when examining the correlation between memorability and embeddings along the object dimensions. 36 out of 49 dimensions displayed a significant association with memorability (shaded bars, FDR-corrected $q < 0.01$), with 9 showing a positive relationship (i.e., *body / body parts* being more memorable), and 27 showing a negative relationship (i.e., *metal / tools* being less memorable).

Having explored memorability across the structure of THINGS, we can readily observe that memorability varies at the exemplar, concept, higher category, and dimensional levels. With this understanding, the question becomes: what determines some concepts/categories/dimensions to be more memorable than others?

Semantic Information Contributes Most to Memorability

To examine which object features are most important for explaining what is remembered and what is forgotten, we used the object space dimensions to predict the average memorability scores of the THINGS concepts (Table 1). Our regression model utilized the 49-dimensional

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

embedding of each concept to predict the average CR score for the concept. Overall, the model explained 38.52% of the variance in memorability (Figure 2b). Because memorability scores contain some noise, we also calculated performance of this model in comparison to a noise ceiling estimated by predicting split halves of the memory data across 100 iterations (see Methods). We found our model explained 61.66% of the variance given the noise ceiling, implying that these dimensions capture a majority of variance in memorability.

The explanatory power of our model serves as a strong starting point for an analysis of the types of dimensions that contribute most to memorability. We sorted the dimensions into two main categories: visual and semantic dimensions. Dimension names were determined in a prior study, as the top two-word phrases selected by naïve observers for sets of the most heavily weighted images on those dimensions (see Methods; Hebart et al., 2020). We defined visual dimensions of an image to be those concerned primarily with color and shape information, such as “red / color”, “long / thin”, “round / circular”, and “pattern / patterned” (Table 1). We defined semantic dimensions as categorical information that did not include references to color or shape, such as “food / carbs”, “technology / electronic”, and “body / body parts”. Any dimensions that contained both semantic and visual information as defined above were classified as mixed, such as “green / vegetables”, “black / accessories”, and “white / winter”.

Table 1. Categorization of THINGS object space dimensions across semantic, visual, and mixed dimensions. Dimension names were derived from naïve observers viewing the highest weighted images on each dimension. Dimensions are listed in order of highest to lowest correlation with memorability score.

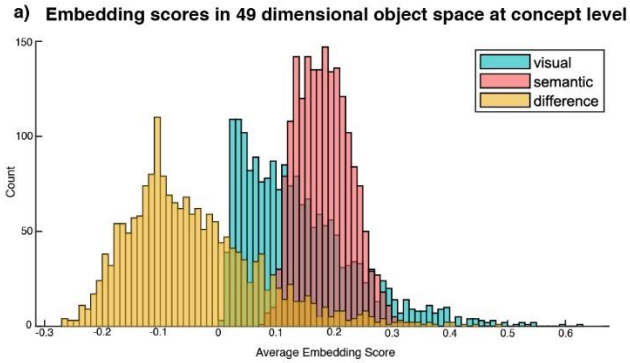
Semantic	**Visual**	**Mixed**
Metal / Tools	Colorful / Colors	Furniture / Bland to Colorful
Food / Carbs	Circular / Round	Green / Vegetables
Animal / Animals	Patterns / Piles	Wood / Brown
Clothes / Clothing	Long / Thin	Royalty / Gold
Backyard / Garden	Red / Color	Dirt / Grainy
Cars / Vehicles	Round / Circular	Black / Accessories
Body / Body Parts	Pattern / Patterned	Long / Rope
Technology / Electronic	Tall / Big	Paper / White
Sports / Sport	Mesh / Nets	Rope / Bands
Tools / Hand Tools		Construction / Long
Paper / Books		Unknown / Colorful
Liquids / Containers		White / Winter
Water / Ocean		Shiny / Jewels
Feminine / Flowers		
Bathroom / Hygiene		
War / Military		
Instruments / Music		
Flight / Air		
Insects / Bugs		
Feet / Body Parts		
Fire / Heat		
Face / Head		
Wheels / Can Sit On		
Containers / Hold Other Things		
Baby / Children		
Medicine / Medical		
Candles / Crafts		

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

With these dimensions labelled, we can differentiate the contributions of primarily semantic and primarily visual dimensions to memorability. By analyzing the embeddings of each concept in the multidimensional object space, we revealed that 70.44% of the concepts were more heavily embedded in dimensions classified as semantic than dimensions classified as visual (Figure 2a). We ran a regression model that predicted memorability only from the dimensions strictly classified as either semantic or visual (excluding mixed dimensions). The resulting 36-dimensional model (27 semantic, 9 visual) explained 35.16% of the variance in memorability, and the semantic dimensions contributed 31.22% of the variance while visual dimensions only accounted for 1.62% with a shared variance of 2.32% (Figure 2c). This result suggests a clear dominance of semantic over visual properties in memorability. To examine the effects of dimensions labelled as mixed, we also broke down the unique and shared variance contributions from semantic, visual, and mixed dimensions in the full 49-dimensional model, demonstrating that mixed dimensions contributed 1.03% of variance in memorability (see Supplementary Material).

However, since there are also a larger number of semantic dimensions than visual dimensions in that model, we conducted a follow-up analysis with a model using just the top 9 highest weighted semantic dimensions and top 9 highest weighted visual dimensions. This model accounted for 19.15% of variance in memorability, with the top 10 semantic dimensions contributing 15.21% of variance while the top 10 visual dimensions contributed 1.87% of variance with a shared variance of 2.07% (Figure 2d). A summary of all regression results is displayed in Figure 2b.

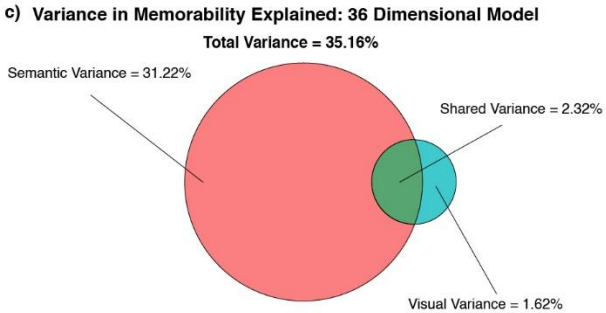
bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



b) Summary of Regression Results

Model	R^2	F	p	error
All dimensions	0.3852	23.0505	2.13e-154***	0.0011
Semantic	0.3354	34.113	9.37e-141***	0.0012
Visual	0.0394	8.3998	2.25e-12***	0.0017
Top Semantic	0.1728	42.7848	5.25e-70***	0.0014
Top Visual	0.0394	8.3998	2.25e-12***	0.0017

*** = $p < .001$



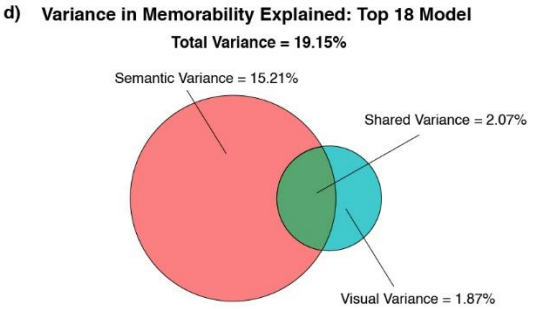


Figure 2. Analyses of relative contributions of semantic and visual properties to memorability. (A) Histogram of averaged embedding values in semantic (red) and visual (blue) dimensions across concepts. The yellow histogram represents the difference between the visual and semantic embeddings (blue - red). The embeddings of the 1,854 concepts in the object space reveal that 70.44% of the concepts are more heavily embedded in semantic dimensions than in visual dimensions. (B) Table of regression models. The semantic and visual models utilize all 27 semantic and 9 visual dimensions respectively to predict memorability and captured 38.52% of the variance in memorability. The top models utilized only the 9 most heavily embedded semantic and visual dimensions, to balance the number of semantic and visual dimensions in the model. Across models, the majority of variance was captured by semantic dimensions. (C) Venn diagram displaying the unique contributions to memorability from semantic and visual dimensions. For the model using all non-mixed dimensions, the majority of variance is captured by the 27 semantic dimensions, with a smaller contribution from the 9 visual dimensions. Note the larger shared variance than visual variance, suggesting that most of the contribution of visual dimensions may be contained in shared variance with semantic dimensions. (D) The same type of Venn diagram as in (C) but with a model including equal numbers of semantic and visual dimensions (9 regressors each). Again, the majority of explained variance comes from semantic dimensions.

Taken together, our results indicate that semantic properties contribute far more than visual properties towards the memorability of an image. While the results reveal contributions of visual properties, these contributions are largely captured by shared variance with semantic properties.

Memorability is More than Just Typicality

While we have determined that semantic features are the most predictive dimensions of the object space for memorability, there is still the question of whether it is the most prototypical or most atypical items that are best remembered along these dimensions. In terms of the object feature space, items that are clustered closely together are the most prototypical items, while

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

items spaced further apart are the most atypical items. The relationship between typicality and memory has been studied extensively in face processing, scene recognition, and related fields (Lee et al., 2000; Bylinskii et al., 2015; Lukavský & Děchtěrenko, 2017), with some studies using memorability interchangeably with atypicality or distinctiveness (e.g., Bruce et al., 1994). A recent body of work has suggested three different hypotheses, where the relationship between typicality and memorability is either always negative (Lukavský & Děchtěrenko, 2017), always positive (Bainbridge & Rissman, 2018), or a specific combination of the two (Koch et al., 2020). Here, we leverage the scale of THINGS to determine this relationship utilizing converging methods for defining typicality based on the multidimensional object space derived from human similarity judgments, a deep neural network for object recognition, and behavioral ratings. These three complementary approaches allow for testing a wide range of hypotheses concerning whether the most prototypical or atypical items are most often remembered.

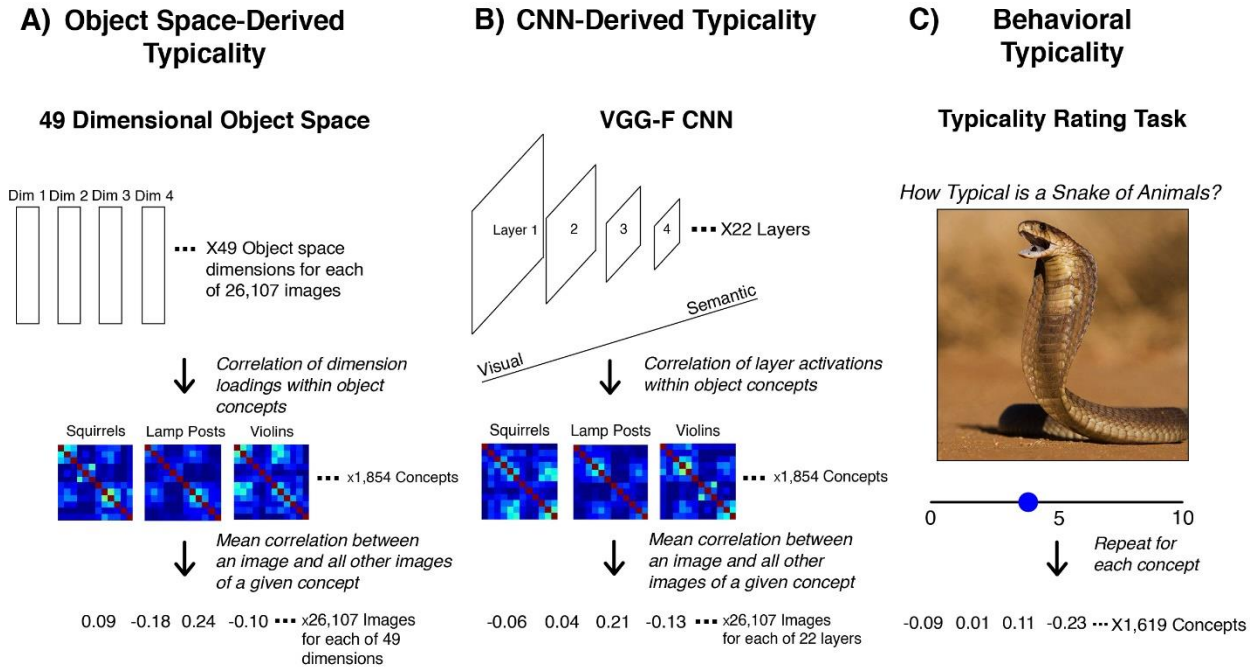


Figure 3. Generating typicality scores from object space dimensions, CNN activations, and behavior. (A) To generate typicality scores from the object space dimensions, we begin with loadings on each of the 49 dimensions of the object space for each of the 26,107 THINGS images. Correlating the resulting dimension loadings within each of the object concepts allowed for the generation of similarity matrices for each object concept. From these matrices, we compute the typicality of each image as the mean correlation between that image and all other images of a given object concept, resulting in a typicality score for every image in relation to its concept. (B) The procedure for generating typicality scores from a CNN is largely the same as the process for the object space dimensions but relying instead on layer activations at each of the 22 layers of the VGG-F network as the representations for each image, which were then correlated to form similarity matrices. (C) For behavioral typicality, participants on Amazon Mechanical Turk used a 0-10 Likert scale to assess the typicality of a given object concept (snake) to its higher category (animals). These typicality scores were then aggregated across all the concepts under a given higher category to generate a typicality score for that category.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Object Space Typicality

Our first measure of typicality we dub "object space typicality", and it is derived from the object space employed in the previous analysis of the visual and semantic dimensions (Figure 3a). Specifically, we quantify an object's typicality as the average similarity of a given example image (e.g. a particular example of a squirrel) to all other examples of that image's concept (e.g. all images of *squirrels* in THINGS; see Methods). The 49-dimensional space has been demonstrated to capture human behavior in excess of 90% of the noise ceiling (Hebart et al., 2020) and we have just shown it is able to predict memorability with high accuracy.

We first tested the overall relationship between corrected recognition and typicality scores for the 26,107 image corpus of THINGS. We found a significant positive relationship between image typicality and memorability across the THINGS dataset ($r = 0.309$, $p = 6.131 \times 10^{-7}$). This suggests that more memorable images tend to be more prototypical of their concept in their representations across these dimensions, arguing against a general primacy of atypicality in memorability. We also analyzed the relationship between object space typicality and memorability within each of the 1,854 concepts in THINGS by correlating memorability and typicality values across the exemplar images of each concept. In other words, within each concept, what is the relationship between typicality and memorability? We determined that overall, the concepts were more likely to display a relationship where more prototypical images tended to be more memorable (one sample t-test: $t(1852) = 2.074$, $p = 0.038$).

While this finding of most object concepts showing a positive relationship between object space typicality and memorability seems to provide evidence for memorability corresponding to object prototypicality, it is important to note that many object concepts (917) show the opposite relationship where more atypical images are more memorable. For example, for *coats*, more prototypical images were more memorable ($r = 0.857$, $p = 3.66 \times 10^{-4}$), but for other concepts such as *handles*, more atypical images were more memorable ($r = -0.798$, $p = 0.001$).

Additional mixed evidence is also apparent when relating the typicality and memorability of concepts within each of the 27 higher categories present in THINGS, in contrast to the previously described analyses that tested the typicality of images in relation to their concepts. For any given concept, the category typicality score reflects the typicality of that concept (e.g. *squirrels*) relative to all other concepts of its higher category (e.g. *animals*).

When examining the relationship between memorability and object space typicality at the category level, we observed that certain categories, such as *containers* ($r = -0.213$, $p = 0.029$) and *electronic devices* ($r = -0.232$, $p = 0.047$) showed negative relationships (e.g., more atypical containers are more memorable), while *animals* ($r = 0.159$, $p = 0.034$) and *body parts* ($r = 0.473$, $p = 0.005$) demonstrated positive relationships. Overall, across all high-level categories, there were an equal number of positive and negative significant relationships, demonstrating further mixed evidence within the THINGS dataset.

CNN-Based Typicality

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We term our second measure of typicality "CNN-based typicality", as it employs the VGG-F deep CNN to compute similarity ratings across the 22 layers of the network (Figure 3b). Deep neural network models have demonstrated success in predicting the neural responses of different regions in the visual system (Yamins et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014). A critical insight from these studies suggests that earlier layers in the network represent low-level visual information such as edges, while later layers represent more complex and semantic features like categorical information (Güçlü & van Gerven, 2015). Unlike the object space derived scores, these typicality values are directly computed from image features, rather than based on behavioral similarity judgments in response to the images themselves.

Recent analyses using CNNs have suggested that the relationship between typicality and memorability may differentially depend on similarity across semantic and visual features; for example, for a set of scene images, images that were the most visually atypical (i.e., atypical at early layers) but semantically prototypical (i.e., prototypical at late layers) tended to be most memorable (Koch et al., 2020). We can directly address this hypothesis using our CNN-based typicality measure, as we can directly compare typicality values at both early and late layers of the VGG-F network. If the pattern displayed in Koch et al. (2020) holds true, we would expect to see a strong negative correlation between memorability and early layer typicality (i.e., visually atypical items are best remembered) and a strong positive correlation with late layer typicality (i.e., semantically prototypical items are best remembered).

We test this hypothesis by producing two correlations for each object concept – the correlation between CR and early layer (2) typicality, and the correlation between CR and late layer (20) typicality. This produces a pair of correlations for each of the 1,854 object concepts. We then visualize these correlation pairs (Figure 4a) and provide a best fit line, which demonstrates the relationship between each correlation pair. The resulting correlation ($r = 0.253$, $p = 2.504 \times 10^{-28}$) suggests that in general, visual and semantic features (as represented in early and late layers) show similar correlations with memorability across the object concepts.

We also segment the concepts into quadrants, which represent four potential patterns for the correlation pairs for a given object concept. The first quadrant contains concepts that display positive correlations for both early and late layers (i.e., visually and semantically prototypical items are best remembered). The second quadrant contains concepts that have positive early layer correlations but negative late layer correlations (i.e., visually prototypical and semantically atypical items are best remembered). The third quadrant contains concepts with negative correlations for both early and late layers (i.e., visually and semantically atypical items are best remembered), and the fourth quadrant contains concepts with negative early layer and positive late layer correlations (i.e., visually atypical and semantically prototypical items are best remembered). This fourth quadrant can be considered a representation of Koch et al.'s (2020) hypothesis, as it corresponds to visually atypical and semantically prototypical items being best remembered.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A chi-square analysis on each quadrant revealed that significantly more concepts than chance showed a pattern where the most memorable items were prototypical in terms of both early and late layer features ($\chi^2 = 38.046$, $p = 6.909 \times 10^{-10}$). In contrast, we find significantly fewer concepts than chance show a mixed pattern, where memorable items were determined by early layer prototypicality and late layer atypicality ($\chi^2 = 8.454$, $p = 0.004$), or the opposite pattern of early layer atypicality and late layer prototypicality ($\chi^2 = 20.286$, $p = 6.668 \times 10^{-6}$). Finally, there was no difference from chance in the proportion of concepts that showed a pattern where the most memorable items were the most atypical items for both early and late CNN layers ($\chi^2 = 8.3993$, $p = 0.553$). These results suggest that in general, memorable images tend to be those that are both visually and semantically prototypical of their object concept, although there are also concepts for which memorable images may tend to be either visually or semantically atypical.

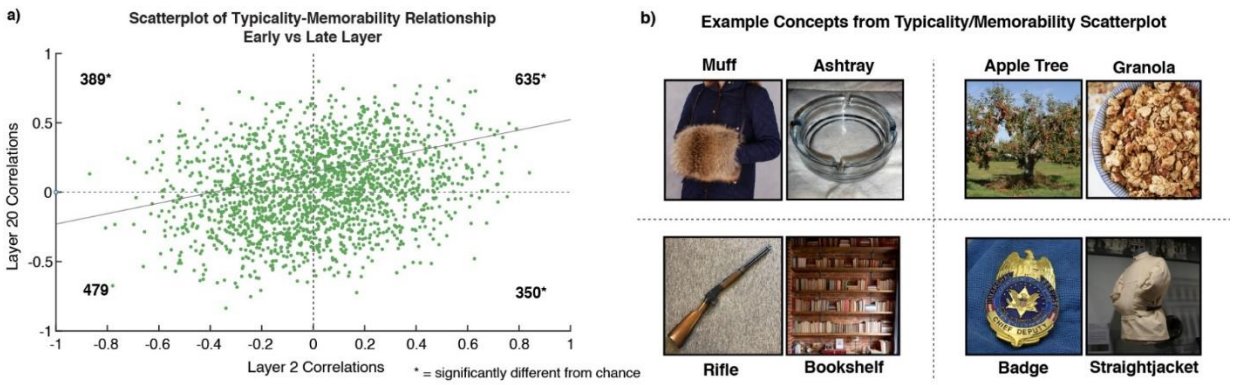


Figure 4. Examining relationships between typicality, memorability, and semantic and visual content. (A) Visualizing the correlation of CNN-based typicality and memorability for all 1,854 concepts in terms of an early layer (layer 2) and late layer (layer 20) allows for the observation of an overall positive relationship between early and late layer typicality scores across the concepts ($r = 0.253$, $p = 2.504 \times 10^{-28}$). A chi-square analysis of the four quadrants of the scatterplot demonstrated significantly more concepts than chance showed a pattern where the most memorable items were prototypical in terms of both early and late layer features ($\chi^2 = 38.046$, $p = 6.909 \times 10^{-10}$). Contrastingly, we find significantly fewer concepts that demonstrate "mixed" patterns where more memorable items demonstrated early layer prototypicality and late layer atypicality ($\chi^2 = 8.454$, $p = 0.004$), or the opposite pattern ($\chi^2 = 20.286$, $p = 6.668 \times 10^{-6}$). We found no significant difference from chance for concepts where the most memorable items were atypical across both early and late layer features ($\chi^2 = 8.3993$, $p = 0.553$). This suggests that, in general, memorable concepts tend to be both visually and semantically prototypical. (B) Example concepts that fell into each quadrant of the scatterplot seen in A.

Behavioral Typicality

Our third and final measure of typicality, referred to as "behavioral typicality", consists of behavioral ratings derived from a concept to category matching task (Hebart et al., 2020) (Figure 3c). In this prior study, participants on Amazon Mechanical Turk used a 0-10 Likert scale to assess the degree to which a given concept was typical of a category (e.g., how typical is a snake of animals?). These ratings allow us to capture human intuition regarding typicality.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A correlation between CR scores and behavioral typicality scores across all higher categories showed no significant relationship between typicality and memorability ($r = 0.139$, $p = 0.576$). When examining the distribution of correlations between typicality and memorability within the higher categories, we observed a marginal effect of more atypical (rather than prototypical) categories being more memorable ($t(26) = -2.022$, $p = 0.054$). When examining the correlations for each of the 27 categories separately (see Supplementary Material), we found that *home décor* ($r = -0.384$, $p = 0.009$), *office supplies* ($r = -0.430$, $p = 0.032$), and *plants* ($r = -0.429$, $p = 0.003$) showed significant negative relationships, implying that more memorable examples of each category were more atypical. In contrast, *animals* ($r = 0.176$, $p = 0.020$), *food* ($r = 0.115$, $p = 0.050$) and *vegetables* ($r = 0.317$, $p = 0.041$) had positive relationships, implying that more memorable examples were more prototypical. Overall, when examining typicality using behavioral ratings, we find additional evidence suggesting that memorability is not accounted for by either object prototypicality or atypicality.

Together, our results demonstrate that memorability cannot be considered synonymous with either prototypicality or atypicality, as has been suggested in previous studies (e.g., Valentine et al., 1991; Bylinskii et al., 2015; Bainbridge et al., 2017). Certain results collected using both object space-derived and CNN-derived typicality scores suggest a trend towards more prototypical stimuli being more often remembered, but the large number of counterexamples present across the different typicality scores and levels of analysis suggest that the relationship between memorability and typicality is likely more complex than a simple positive or negative association, varying strongly from concept to concept.

DISCUSSION

We acquired and analyzed a large dataset of memory ratings for a representative object image database to uncover what makes certain objects more memorable than others. Specifically, we investigated the roles of semantic and visual features and revealed that semantic properties more strongly influence what is remembered than visual properties. We leveraged three complementary measures of object typicality to determine whether the most prototypical or most atypical images are best remembered and uncovered some evidence suggesting more prototypical items are more memorable, but also a high degree of variance across concepts and categories, suggesting that memorability is not just a measure of the typicality of an object or image. These findings shed new light on the determinants of what we remember and stand in contrast to previous studies that have claimed both that semantic information is not required to determine memorability (Lin et al., 2021) and that memorability is synonymous with atypicality (Bruce et al., 1994).

Semantic Primacy of Memorability

We analyzed the contributions of semantic and visual dimensions to memorability to determine if the two types of information contribute differentially to the THINGS stimuli. Our results reveal a primacy of semantic dimensions in explaining memorability, based on multiple regressions comparing the relationship of the entire object dimensional space to memorability.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Even after equalizing the number of semantic and visual dimensions inputted to the model, 88.02% of the variance in memorability captured by the space was exclusively from the top 9 semantic dimensions.

Previous findings of the ability of CNNs (Khosla et al., 2015) and monkeys (Jaegle et al., 2019) to predict human performance on memorability tasks and examples of memory performance robust to semantic degradation (Lin et al., 2021) have led to the assertion that semantic knowledge is not required to make an image memorable. However, recent research has demonstrated that semantic similarity is predictive of memorability and lexical stimuli also display intrinsic memorability despite a lack of rich visual information (Xie et al., 2020; Madan et al., 2021). More recently, other studies have demonstrated that both visual and semantic features contribute differentially with regards to both object memory (Hovhannisyan et al., 2021) as well as the typicality-memorability relationship, where visually atypical but semantically prototypical scene images may be the most memorable (Koch et al., 2020). Additionally, recent research in memorability prediction suggests that adding semantic information to a deep neural network improves the prediction of memorability scores (Needell & Bainbridge, 2022). Our results demonstrate a strong semantic primacy in memory which lends additional support to recent findings demonstrating the importance of semantic information in determining what we remember.

Beyond behavior, our findings align with the results from recent neuroimaging studies that have examined the neural correlates of memorability. One such study found a lack of memorability-related activation in the Early Visual Cortex (EVC), suggesting that areas involved in lower-level perception may not be sensitive to memorability (Bainbridge et al., 2017). This result, coupled with a study demonstrating faster neural reinstatement for highly memorable stimuli in the Anterior Temporal Lobe (ATL), an area typically associated with semantic processing (Xie et al., 2020), could potentially reflect a neural signature of the observed outsize influence of semantic features in determining what is best remembered. In this study, memorability for word stimuli could be significantly predicted by the semantic connectedness of these words, where words that exist at the roots of a semantic structure tended to be more memorable (Xie et al., 2020). This suggests that memorability could reflect our semantic organization of items in a memory network. Other work has also found sensitivity to memorability in late perceptual areas, such as the Fusiform Face Area (FFA) and the Parahippocampal Place Area (PPA) (Bainbridge et al., 2017; Bainbridge & Rissman, 2018), often associated with the patterns seen in late CNN layers (Yamins et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014).

Our findings are particularly surprising given the fact that the object space dimensions explained 61.66% of the variance in memorability. Unlike previous studies of memorability using single attributes (Bainbridge et al., 2017; Isola et al., 2014) or linear combination models with constrained stimulus sets (Bainbridge et al., 2013), we are able to explain a large degree of the variance in memorability, further highlighting the importance of semantic properties. The success of this model also means that this same model can be applied to selecting stimulus sets

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

intended to drive memory in specific ways; given an object's feature space, we can predict which items are likely to be remembered or forgotten. However, given the remaining unexplained variance, it is clear that there are still lingering questions about the determinants of what we remember and what we forget.

Typicality as it Relates to Memorability

Here, we observe that across our images, concepts, and categories, there are some by which the most prototypical are the most memorable, while there are others where the most atypical are the most memorable. These results suggest that memorability does not just reflect an object's typicality, and it is not merely that memorable items are the most distinctive, atypical items. In fact, across multiple levels of analysis, we observe the opposite, where in general more prototypical items tend to be the most memorable.

This is surprising, given that typicality has long been thought to encapsulate the effect of memorability based on evidence from faces (Valentine, 1991; Bruce et al., 1994) and scenes (Bylinskii et al., 2015), whereby more atypical items are thought to be easier to remember. Other studies have rebutted this claim by demonstrating that semantic similarity is predictive of memorability (Xie et al., 2020). Furthermore, late visual areas regions show neural patterns reflective of our current behavioral findings, where memorable face and scene images show more similar neural patterns to each other (i.e., have more prototypical patterns), while forgettable images have more dissimilar neural patterns (i.e., more atypical patterns; Bainbridge et al., 2017; Bainbridge & Rissman, 2018). Further, Koch and colleagues (2020) found a complex relationship with typicality, where visually distinct and semantically similar images were most often remembered in an indoor-outdoor classification task. Our divergent findings could possibly be explained by the constrained stimulus sets utilized in prior studies. While prior work focused on narrow stimulus sets such as faces or a smaller sampling of scene images, our study examines a comprehensive, representative set of object images across the human experience. Our divergent findings from these earlier studies may suggest that while previous findings are reasonable extrapolations from the stimuli domains examined, they are not characteristic of memorability as a whole. When assessed at a global scale, it is neither prototypicality nor atypicality of an item that makes it memorable.

The observation of variability in the typicality-memorability relationship may have important ramifications for neuroimaging research examining the neural correlates of memorability and memory more broadly. Observations of prototypicality in neuroimaging research reference a phenomenon called pattern completion as a means by which the hippocampus retrieves a complex representation from a given cue (LaRocque et al., 2013). This process depends on another hippocampal phenomenon termed pattern separation, where similar inputs are assigned distinct representations to facilitate the mnemonic discrimination required in memory (Ngo et al., 2020). Whole-brain fMRI analyses have revealed that different areas involved in memory utilize separated and overlapping information to facilitate memory (LaRocque et al., 2013), suggesting a potential role for both prototypicality (as represented by

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pattern completion) and atypicality (as represented by pattern separation) in facilitating memory. Future neuroimaging research could identify potential neural markers of prototypicality and atypicality and determine if the effects of semantic and visual information are dissociable at a neural level.

Conclusion

Here, we have created the best performing model to date of the object features that are predictive of image memorability. From this model, we have observed a primacy of semantic properties in determining what we remember. This underscores recent findings of the important role of semantic information in memory (Xie et al., 2020) and emerging work with CNNs that demonstrate a classification performance benefit when including semantic information into their models (Needell & Bainbridge, 2022).

Beyond highlighting the roles of semantic and visual dimensions, our results demonstrate that neither prototypicality nor atypicality fully explains what makes something memorable, and if anything, prototypical items tend to be the most memorable. Our findings challenge decades of prior research suggesting we best remember more atypical items (Valentine, 1991; Vokey & Read, 1992; Lee et al., 2001; Bylinskii et al., 2015; Lukavský & Děchtěrenko, 2017). This trend towards prototypicality is reflected in recent neuroimaging studies (Bainbridge et al., 2017; Bainbridge & Rissman, 2018; Xie et al., 2020), suggesting that prototypicality may be related to the underlying neural mechanisms governing memory.

Our findings shed new light on the features and organizational principles of memory, opening up a wide variety of potential follow-up studies. In fact, with this large-scale analysis, we have identified the stimulus features that govern memorability within and across a comprehensive set of objects, and make this data publicly available for use (https://osf.io/5a7z6/?view_only=675e901c176c4bec9c2540fc4981e5fe). This will allow researchers to make honed predictions of memory within these categories, or use these dimensions to design ideal stimulus sets. For example, our analysis found that animal images are highly memorable, while manmade, metal images are highly forgettable, and so memorability is an important factor to consider in studies looking at visual perception of animacy (Konkle & Caramazza, 2013). Further, given the success of our feature model in predicting memorability, this model could be potentially used to identify memorable images in other image datasets. While THINGS representatively samples concrete object concepts, there are additional stimulus domains beyond objects including dynamic stimuli such as movies, scenes, and non-visual stimuli that could be analyzed in the context of our results. With the understanding that neither prototypicality nor atypicality alone fully characterizes the relationship between typicality and memorability, there is the question of what biases certain stimuli towards one or the other.

We uncover both a semantic primacy in explaining memorability and determine that the relationship between typicality and memorability is more complex than either prototypicality or atypicality alone. We provide this comprehensive characterization in pursuit of a nuanced understanding of the underlying determinants of memorability, and memory more broadly.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Developing this understanding further will have implications far beyond cognitive neuroscience in realms such as advertising, patient care, and computer vision. With the development of generative models of stimulus memorability, it is more important than ever before to ground these models in an empirical understanding of what makes something memorable.

METHODS

Participants

13,946 unique participants completed a continuous recognition repetition detection task on the THINGS images over AMT (see "*Obtaining Memorability Scores for THINGS"*). All online participants acknowledged their participation and were compensated for their time, following the guidelines of the National Institute of Health Office for Human Subjects Research Protections (OHSRP). Participants had to be located within the United States and have participated in at least 100 tasks previously on AMT with at least a 98% approval rating overall to be recruited for the experiment. Participants who made no responses on the task were removed from the data sample.

Stimuli: THINGS

To examine memorability across a broad range of object concepts, we utilized the entire 26,107 image corpus of the THINGS database (Hebart et al., 2019, https://osf.io/jum2f/) for all of our experiments. The THINGS concepts span the wide range of concrete objects, including animate and inanimate, as well as manmade and natural concepts, such as *aardvarks*, *goalposts*, *tanks*, and *boulders*. These 1,854 concepts were generated from the WordNet lexical database through a multilevel web scraping process (Hebart et al., 2019). Each concept has a minimum of 12 exemplar images, though some have as many as 35. These concepts were sorted into 27 overarching categories including *animal-related*, *food-related*, and *body parts*. These higher categories were generated using a two-stage AMT experiment.

At the concept level, we utilized the representational embedding of each concept supplied by THINGS as the multidimensional space for our analyses (Hebart et al., 2020). The original 49-dimensional behavioral similarity embeddings (Hebart et al., 2020) had been generated based on the 1,854 object concepts. Dimension names were generated by two pools of naïve observers in a categorization task (Hebart et al., 2020). The first pool of observers viewed the most heavily reflected images along a given dimension of the space and generated potential labels from the images. The second pool of observers then narrowed down the list of labels until the top two labels remained for each dimension, which was then assigned as the name for that dimension. To derive 49-dimensional embeddings for each of the 26,107 images in the THINGS database, we used predictions from a deep neural network as a proxy. The prediction was carried out for each dimension separately using Elastic Net regression based on the activations of object images in the penultimate layer of the CLIP Vision Transformer (ViT, Radford et al., 2021), which has been shown to yield the most human-like behavior of all available CNN models in a range of

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tests (Geirhos et al., 2021). The Elastic Net hyperparameters were tuned and evaluated using nested 10-fold cross-validation, yielding high predictive performance in most dimensions (mean Pearson correlation between predicted and true dimension scores: r > 0.8 in 20 dimensions, r > 0.7 is 32 dimensions, r > 0.6 in 44/49 dimensions). We then tuned the hyperparameters on all available data using 10-fold cross-validation and applied the regression weights to the CNN representations of THINGS images, yielding 49-dimension scores for all 26,107 images.

Obtaining Memorability Scores for THINGS

In order to examine memorability in the context of the THINGS space, we collected memorability scores for all 26,107 images (publicly available in an online repository: https://osf.io/5a7z6/?view_only=675e901c176c4bec9c2540fc4981e5fe). To quantify the memorability of each stimulus, each participant viewed a stream of images on their screen and was instructed to press the R key whenever they saw a repeated image. Each image was presented for 500ms, and the interstimulus interval was 800ms. For each repeated stimulus, there was a minimum 60-second delay between the 1st and 2nd presentation of that image, although this delay was jittered so that repetitions could not be predicted based on timing. The task also included easier "vigilance repeats" of 1-5 images apart, to ensure participants were paying attention to the task. The presentation of images was such that approximately 40 participant responses were gathered per image. Of the 1,854 concepts in THINGS, each concept was either represented with a single exemplar or not represented at all during a participant's set of trials in order to control for within-concept competition effects on memory performance. To avoid familiarity effects, participants were only allowed to participate again after a minimum delay of 2 weeks.

Memorability was quantified in THINGS using corrected recognition (CR) scores for each image. Corrected recognition is calculated by subtracting the false alarm rate for a given stimulus from the hit rate for the same stimulus. Hit rate is defined as the proportion of correct repetition detections, whereas false alarm rate is defined as the proportion of incorrect detections. CR allows for a single metric that integrates information about both hit rate and false alarm rate. However, we also replicate all results using hit rate and false alarm rate separately (Supplementary Material).

We ran a split-half consistency analysis to determine if participants were consistent in what they remembered. The analysis randomly partitioned participants into two halves and calculated a Spearman rank correlation between the CR scores for all images, as defined by the two random halves of participants. In other words, this analysis determines how similar the memory performance is for each image between these two independent halves of participants. This process was repeated across 1,000 iterations and an average correlation *rho* was calculated. This *rho* was then corrected using the Spearman-Brown correction formula for split-half correlations. If there is no consistency in memory performance across participants, we would expect a zero value for *rho*, whereas a high value would suggest that what one-half of participants remembered, so did the other. To estimate chance, we correlated one half of

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

participants' scores with those for a shuffled image order of the other participant half, across 1,000 iterations. The p-value was calculated as the proportion of shuffled correlations higher than the mean consistency between halves.

Semantic/Visual Contribution and Regression Model Analyses

With memorability scores at the image level available, we can relate the memorability of THINGS stimuli with the associated representational space and determine the relative contributions of semantic and visual dimensions to memorability. To accomplish this, we analyzed the embeddings of the 1,854 concepts in the 49 dimensions and separated them into semantic and visual dimensions. Of the 49 dimensions, 27 were identified as semantic, 9 as visual, and the remaining 13 as mixed (Table 1).

To determine the effects of semantic and visual dimensions on memorability, we ran a series of multiple regression models. We began with an omnibus model predicting average memorability for each of the 1,854 concepts using the full set of 49 dimensions. This model assessed the total variance in memorability explained by the dimensions. We then utilized a model predicting memorability from the 36 dimensions classified as either semantic or visual to determine the differential contributions of each type of information. As there were more semantic dimensions than visual dimensions, we also ran a model that only used the 9 most heavily reflected semantic and 9 most heavily reflected visual dimensions to control for the overrepresentation of semantic information. In order to assess the potential variance explained by dimensions classified as mixed, we also break down the unique variance contributed by mixed dimensions to the full 49-dimensional model (see Supplementary Material). In all models we also analyzed the unique and shared variance contributions of the two types of dimensions to memorability using variance partitioning. Unique semantic variance was calculated as the overall R^2 value for the full model minus the R^2 value for a model containing only the visual dimensions and vice versa for visual variance. The shared variance was calculated as the overall model R^2 minus both the unique semantic and unique visual variance.

In order to compare the performance of the omnibus model (all 49 dimensions) to the noise ceiling, we conducted a split-half regression analysis. Across 100 iterations, the participant sample was split into two random halves, and we ran two models. For the first model, we looked at the ability of the 49 dimensions to predict the memorability scores derived from the first half of participants. For the second model, we included an additional 50th predictor which was the memorability scores derived from the second half of participants, for the same images. This second model serves as a noise ceiling of memorability from which we can compare the first model. To see the proportion of variance explained in comparison to this noise ceiling, we then averaged the ratio of the R^2 of the first model to the second model, across iterations.

Memorability-Typicality Relationship Analyses

To determine if memorability is reflective of object prototypicality or atypicality, we assessed the relationship between typicality and memorability of the THINGS images. We

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

conducted these analyses at two levels: mapping images to concepts, and mapping concepts to categories. We utilized typicality scores from the object space dimensions, the VGG-F Convolutional Neural Network (CNN), and behavioral ratings of typicality.

To create our object space typicality scores, we leveraged the 49-dimensional object space and embeddings of all 26,107 images within that space. For each concept, we generated a similarity matrix containing the embedding values of the component images of that concept along all 49 dimensions. From that matrix, we can extract a single value for each image that is the average similarity (Pearson correlation) between that image's dimensional embeddings and those of the other images of that concept, which we define as the typicality of that image. In other words, a low mean correlation would imply a highly atypical stimulus (distinct from other exemplars of the same concept), while a high mean correlation would imply a highly prototypical stimulus (very similar to exemplars of the same concept). We utilize the same paradigm to generate typicality values for each concept in relation to other concepts under a given category using an embedding of each concept in the object space and comparing its similarity to the embedding of all other concepts within the same category.

For our CNN-based typicality scores, we leveraged the VGG-F CNN object classification network to compute typicality directly from image features. Early layers of CNNs are more sensitive to low-level image features, such as edges, while later layers are more sensitive to higher-level and semantic features, such as animacy (Güçlü & van Gerven, 2015). We can therefore extract information at these various points in the network to test the separate contributions of visual and semantic typicality. The paradigm for extracting typicality values was similar to the object space typicality values: for each concept, similarity matrices were generated based on the flattened layer output values for all component images. The typicality for each exemplar was then calculated as the mean of its similarity (Pearson correlations) with all other exemplars in the concept. This measure tells us how similar a given exemplar is to all other exemplars in terms of its CNN-predicted features. This procedure is repeated for every layer in VGG-F, resulting in 21 typicality values for each image in relation to its object concept, one for each layer of VGG-F.

For our behavioral typicality scores, we employed the ratings collected as part of the THINGS database (Hebart et al., 2020). These ratings were collected for each of the 1,854 THINGS concepts and represent the typicality of the concept in relation to its higher category on a scale of 0 to 10. For example, the typicality rating for *stomach* under the higher category *body parts* reflects how typical a stomach is as a body part (considering other body parts like legs or shoulders).

To analyze the relationships between typicality and memorability across the THINGS dataset, we use our object space, CNN-based, and behavioral typicality scores at two different levels of analysis: image level and concept level. At the image level, we analyze the object space and CNN-derived typicality values to examine their relationship to memorability across all

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26,107 images in THINGS, which gives a single value for the overall typicality-memorability relationship of the THINGS images. Beyond the overall trend, we also examine the relationship within each of the 1,854 image concepts by correlating the typicality scores and memorability scores of their component images. This allows for the visualization of more nuanced relationships between the THINGS concepts. At the concept level, we perform a correlation between the behavioral typicality scores and CR scores and examine the resulting distribution of the relationships for each of the 27 higher categories.

ACKNOWLEDGEMENTS

The researchers would like to thank Coen Needell and Deepasri Prasad for their helpful comments on the manuscript and Sara Hedberg for assistance in generating figures. This research was funded by the Intramural Research Program of the National Institutes of Health (ZIA-MH-002909), under National Institute of Mental Health Clinical Study Protocol 93-M-1070 (NCT00001360).

REFERENCES

1. Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? Journal of Vision, 11(11), 1282–1282. https://doi.org/10.1167/11.11.1282
2. Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, *142*(4), 1323–1334. https://doi.org/10.1037/a0033872
3. Bainbridge, W. A. (2019). Memorability: How what we see influences what we remember. Psychology of Learning and Motivation Knowledge and Vision, 1–27. https://doi.org/10.1016/bs.plm.2019.02.001
4. Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What Makes a Visualization Memorable? IEEE Transactions on Visualization and Computer Graphics, 19(12), 2306–2315. https://doi.org/10.1109/tvcg.2013.234
5. Xie, W., Bainbridge, W. A., Inati, S. K., Baker, C. I., & Zaghloul, K. A. (2020). Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. Nature Human Behaviour, 4(9), 937–948. https://doi.org/10.1038/s41562-020-0901-2
6. Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. Neuropsychologia, 141, 107408. https://doi.org/10.1016/j.neuropsychologia.2020.107408
7. Bainbridge, W. A., Berron, D., Schütze, H., Cardenas-Blanco, A., Metzger, C., Dobisch, L., … Düzel, E. (2019). Memorability of photographs in subjective cognitive decline and mild cognitive impairment: implications for cognitive assessment. https://doi.org/10.1101/660365
8. Needell, C. D., & Bainbridge, W. A. (2022). Embracing new techniques in deep learning for estimating image memorability. *Computational Brain & Behavior*. https://doi.org/10.1007/s42113-022-00126-5

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

9. Bainbridge, W. A., Dilks, D. D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. NeuroImage, 149, 141–152. https://doi.org/10.1016/j.neuroimage.2017.01.063
10. Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What Makes a Photograph Memorable? IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7), 1469–1482. https://doi.org/10.1109/tpami.2013.200
11. Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. The Quarterly Journal of Experimental Psychology Section A, 43(2), 161–204. https://doi.org/10.1080/14640749108400966
12. 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). https://doi.org/10.3115/v1/d14-1162
13. Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and Predicting Image Memorability at a Large Scale. 2015 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2015.275
14. Jaegle, A., Mehrpour, V., Mohsenzadeh, Y., Meyer, T., Oliva, A., & Rust, N. (2019). Population response magnitude variation in inferotemporal cortex predicts image memorability. ELife, 8. https://doi.org/10.7554/elife.47596
15. Lin, Q., Yousif, S. R., Chun, M. M., & Scholl, B. J. (2021). Visual memorability in the absence of semantic content. Cognition, 212, 104714. https://doi.org/10.1016/j.cognition.2021.104714
16. Madan, C. R. (2020). Exploring word memorability: How well do different word properties explain item free-recall probability? Psychonomic Bulletin & Review, 28(2), 583–595. https://doi.org/10.3758/s13423-020-01820-w
17. Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature Effects, Distinctiveness, and Identification: Testing the Face-Space Framework. Psychological Science, 11(5), 379–385. https://doi.org/10.1111/1467-9280.00274
18. Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. Vision Research, 116, 165–178. https://doi.org/10.1016/j.visres.2015.03.005
19. Lukavský, J., & Děchtěrenko, F. (2017). Visual properties and memorising scenes: Effects of image-space sparseness and uniformity. Attention, Perception, & Psychophysics, 79(7), 2044–2054. https://doi.org/10.3758/s13414-017-1375-9
20. Mohsenzadeh, Y., Mullin, C., Oliva, A., & Pantazis, D. (2019). The perceptual neural trace of memorable unseen scenes. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-42429-x
21. Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Wicklin, C. V., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. https://doi.org/10.1101/545954
22. Koch, G. E., Akpan, E., & Coutanche, M. N. (2020). Image memorability is predicted by discriminability and similarity in different stages of a convolutional neural network. Learning & Memory, 27(12), 503–509. https://doi.org/10.1101/lm.051649.120

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

23. Bainbridge, W. A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-26467-5
24. LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global Similarity and Pattern Separation in the Human Medial Temporal Lobe Predict Subsequent Memory. Journal of Neuroscience, 33(13), 5466–5474. https://doi.org/10.1523/jneurosci.4293-12.2013
25. Ngo, C. T., Michelmann, S., Olson, I. R., & Newcombe, N. S. (2020). Pattern separation and pattern completion: Behaviorally separable processes? Memory & Cognition, 49(1), 193–205. https://doi.org/10.3758/s13421-020-01072-y
26. Contier, O., Hebart, M. N., Dickter, A. H., Teichmann, L., Kidder, A., Corriveau, A., Zheng, C., Vaziri-Pashkam, M., Baker, C. I. (2021). THINGS-fMRI/MEG: A large-scale multimodal neuroimaging dataset of responses to natural object images.
27. Hebart, M. N., Zhang, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. Nature Human Behavior. https://doi.org/https://doi.org/10.1038/s41562-020-00951-3
28. Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *Proceedings of the British Machine Vision Conference 2014*. https://doi.org/10.5244/c.28.6
29. Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences, 111(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111
30. Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Computational Biology, 10(11). https://doi.org/10.1371/journal.pcbi.1003915
31. Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/jneurosci.5023-14.2015
32. Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. Memory & Cognition, 20(3), 291–302. https://doi.org/10.3758/bf03199666
33. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. http://arxiv.org/abs/2103.00020
34. Geirhos, R., Narayanappa, K., Mitzkus, B., Thieringer, T., Bethge, M., Wichmann, F. A., & Brendel, W. (2021). Partial success in closing the gap between human and machine vision. http://arxiv.org/abs/2106.07411
35. Madan, C.R. (2021). Exploring word memorability: How well do different word properties explain item free-recall probability?. *Psychon Bull Rev* 28**,** 583–595 https://doi.org/10.3758/s13423-020-01820-w

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.29.490104; this version posted April 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

36. Konkle, T., & Caramazza, A. (2013). Tripartite Organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, *33*(25), 10235–10242. https://doi.org/10.1523/jneurosci.0983-13.2013
37. Bruce, V., Burton, M. A., & Dench, N. (1994). What's distinctive about a distinctive face? *The Quarterly Journal of Experimental Psychology Section A*, *47*(1), 119–141. https://doi.org/10.1080/14640749408401146
38. Hovhannisyan, M., Clarke, A., Geib, B., Cicchinelli, R., Monge, Z., Worth, T., Szymanski, A., Cabeza, R., & Davis, S. (2020). The visual and semantic features that predict object memory: concept property norms for 1000 object images. https://doi.org/10.31234/osf.io/nqmjt