Prioritized memory can explain the effect of value on category representation

Linas Nasvytis
Department of Psychology
Stanford University
linasmn@stanford.edu

Joshua Knobe
Department of Philosophy
Yale University
joshua.knobe@yale.edu

Fiery Cushman
Department of Psychology
Harvard University
cushman@fas.harvard.edu

Abstract

Category representations are often assumed to reflect the statistical distribution individual category members. However, recent work shows that people's category representations tend to be biased toward high-value category members. We propose that this bias stems from prioritized memory: when learning about a category, people devote more cognitive resources to remembering important or desirable items, leading to their overrepresentation in category-level representations. We test key predictions of this account behaviorally and computationally. Behaviorally, we find a strong correlation between the features people prioritize in memory and the features that dominate their spontaneous recall of category members. Computationally, we use variational autoencoders to show that when statistical learning prioritizes accuracy for certain items, these items are overrepresented when sampling from the learned category distribution. Together, these findings suggest that prioritized memory plays a key role in shaping category representations.

Keywords: category representation; memory; sampling; statistical learning; variational autoencoder

Introduction

People reason not just about individual things (Einstein; Tokyo; Citizen Kane) but also about the categories to which they belong (e.g., scientists; cities; films). How are category-level concepts represented and used?

One idea, amply documented by prior work, is that category representations are influenced by the distribution of items within the category, learned by a kind of unbiased statistical aggregation (Rosch & Mervis, 1975; Griffiths & Tenenbaum, 2006). Thus, for instance, our concept of a scientist might be a composite of the scientists we've met or learned about, designed to approximate the true "average" scientist. Many models of statistical learning illustrate how category representations might take this form.

Several lines of evidence suggest, however, that our category representations are not unbiased. These studies show that, in fact, our category representations are skewed towards more "valuable" or "ideal" category members (Barsalou, 1985; Foster-Hanson & Rhodes, 2019). Thus, for instance, when people are asked to describe a scientist, people do not report the most statistically common features, but instead those that are both relatively common and also especially good. A similar effect arises when people are asked to spontaneously mentally generate a novel example fitting the category (e.g., "picture the first scientist that comes to mind"): What people think of tends not to reflect the true statistical average, but rather a compromise between statistical frequency and value (Bear, Bensinger, Jara-Ettinger, Knobe, & Cushman, 2020). The scientist that comes to mind is not really an average scientist, but rather an especially good one.

Given this bias in the representation and use of category concepts, one natural possibility is that it reflects an upstream bias in the learning processes by which the category concept is acquired. That is the approach we pursue here, drawing on both behavioral experiments and a computational investigation of statistical learning.

Our proposal revolves around two key ideas. The first idea is that category representations are constructed, at least in part, through the process of encoding and remembering individual category members. Thus, for instance, we construct that category representation of "scientist" in part by encoding and remembering particular scientists we've met or heard about. Thus, the bias responsible for the effect of value on category representation may arise from how we learn about and remember specific scientists.

The second idea is that, when trying to learn about individual objects, people care much more about remembering some of them accurately, and care less about remembering others (Shohamy & Adcock, 2010; Mattar & Daw, 2018; Rouhani, Niv, Frank, & Schwabe, 2023). Since our capacity for memory is limited, there are tradeoffs in the accuracy with which different items can be remembered. Stated in terms of statistical learning, people might therefore adopt a *weighted loss function*. They might be willing to accept a large amount of additional inaccuracy in their representation of certain objects if that would allow them to be just slightly more accurate in their representation of other objects.

Putting these two claims together may explain the patterns observed in people's category representations (Knobe & Cushman, 2023). Category representations serve in part to enable people to represent individual objects, but different individual objects are not all equal in importance; it is more important to accurately represent some objects and less important to represent others. The core idea of our proposal is that it is precisely for this reason that people's category representations systematically deviate from the observed statistical distribution. Here, we take a two-pronged approach to assessing this proposal. First, we test a novel behavioral prediction. While prior work shows that people often over-represent *desirable* features within a category (e.g., more transformative scientists), our proposal predicts that they should over-represent *important-to-remember* features. Naturally, these

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

properties are often aligned. All else being equal, it is typically more important to remember good scientists than bad ones; better movies than worse ones; and more desirable cities than less desirable ones. In some cases, however, it can make sense to prioritize memory for items with *undesirable* features, like dangerousness—as, for instance, when allocating memory resources to wild animals, criminals, or diseases. In our behavioral study, then, we ask whether the overrepresentation of desirable vs. undesirable features in categories is correlated with the importance of remembering individuals with desirable vs. undesirable features.

Second, we use computational experiments to test the assumption that, when statistical learning prioritizes the accuracy of certain items over others, this can result in a category-level representation that overepresents prioritized items. We perform these computational experiments using variational autoencoders (VAEs). For our purposes, VAEs have three key advantages. First, they are a popular and well-understood architecture for machine learning that have, in some contexts, proved to be a useful tool for understanding human cognition (Bandyopadhyay et al., 2022; Keller, Gao, & Welling, n.d.). Second, by design, VAEs encode the details of specific items by building, and then drawing upon, structured category-level representations. Third, the construction of the VAE allows for a simple and interpretable operation in which one samples from a category-level representation, akin to asking, "describe the first scientist who comes to mind", or "tell me what a normal scientist is like". This allows us to ask whether a VAE that prioritizes certain items for accurate encoding also, like humans, tends to overrepresent items of this kind at the category level.

The basic approach that we pursue here can be contrasted with an alternative possibility: That the process of calling exemplars to mind from various categories is designed to overrepresent items—sometimes of high value, other times of low value—in order to efficiently guide decision-making (Lieder, Griffiths, & Hsu, 2018). These are not mutually exclusive possibilities, but they provide distinct explanations for the observation that items of extreme value tend to be overrepresented when making decisions by sampling (Madan, Ludvig, & Spetch, 2014) or in other areas of cognition (Horwath, Rouhani, DuBrow, & Murty, 2023).

Behavioral Study

Our behavioral study compared two measurements of a category, each taken from a separate group of participants. For a variety of categories (e.g., "Government agencies" or "Insects"), one group of participants was asked whether it is more important to remember details about especially good items or especially bad items. This served as a measure of prioritized memory. For the same variety of categories, a separate group of participants was asked to name the first category-member that comes to mind and then asked whether this item is above- or below-median in value. Our key question is whether there is a systematic relationship at the item level between these measures. That is, do categories for which it is important to remember especially good items also tend to generate good-biased samples of "what comes to mind", while categories for which it is important to remember especially bad items generate bad-biased samples?

We chose a variety of categories that we thought would be familiar to participants, and we attempted to choose some categories for which we thought good items would be prioritized in learning and memory (e.g., athletes; cars) and other categories for which we thought bad items would be prioritized (e.g., diseases; terrorist organizations).

Method

We recruited 400 participants via Amazon Mechanical Turk using the CloudResearch platform, and assigned them to one of two groups (200 each). Questions in each study involved 11 categories: Animals, Athletes, Cars, Diseases, Foods, Universities, Types of Weather, Government Agencies, Insects, Bodily Injuries, and Terrorist Organizations.

In Group 1, participants completed 11 questions, one for each category. For each question, they read a standardized prompt describing a trade-off between expertise in high-value items (those they valued the most or detested the least) and low-value items (those they valued the least or detested the most). The prompt framed the decision as a choice: participants could either be (A) highly knowledgeable about the items in the category they valued the most or detested the least, while knowing little to nothing about the items they valued the least, or (B) the reverse—becoming an expert on the items they valued the least or detested the most, while remaining largely ignorant of the items they valued most. After reading the prompt for each category, participants indicated which option they preferred—whether it was more important to them to have expertise in high-value or low-value items.

In Group 2, participants completed a questionnaire of a similar format. For each of the 11 categories, they were first asked to name the first item that came to mind ("Please name a [category name] that first comes to your mind"). After providing these responses, participants were asked to imagine listing all the different kinds of items that exist within that category (e.g. all the different kinds of animals). They were then instructed to mentally arrange these items in order of personal value or preference. At the top of the list would be the items they value the most or detest the least—these were referred to as the "better" items. Conversely, at the bottom of the list would be the items they value the least or detest the most—referred to as the "worse" items. After establishing this mental list, participants were reminded of the specific item they had previously generated for that category (e.g., the first animal that came to mind). They were then asked to determine where this item would fall within the top half or the bottom half of the ordered list.

Results

We asked whether the prioritization of high- versus low-value items for memory within each category predicted the value of

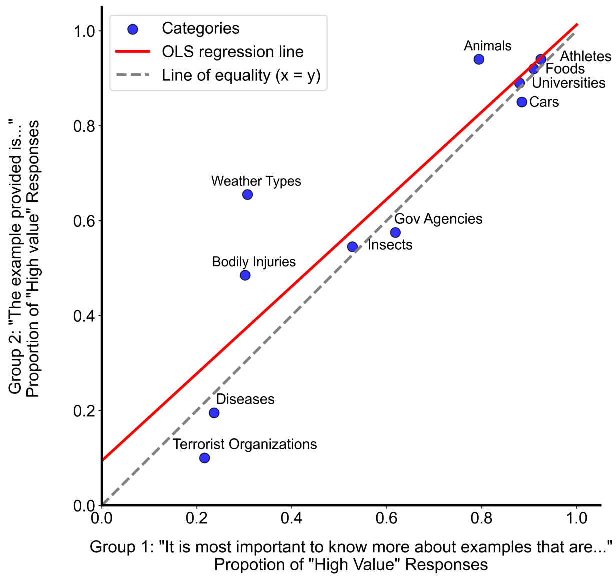

Figure 1: Proportion of participants who indicated prioritizing high-value items (x-axis) for each category, and proportion of items that came to mind to participants which were classified as high-value (y-axis) per each category.

items that come spontaneously to mind within the same category. For each category, we calculated (1) the proportion of participants who prioritized knowing and remembering more about the high value items (from Group 1), and (2) the proportion of items that came to mind that participants categorized as high-value (from Group 2). Figure 1 shows the relationship between these two proportions. We then conducted a linear regression on the resulting 11 pairs of summary statistics, predicting the latter from the former. This revealed a strong relationship between these measures, with a slope of 0.90 (95% CI: 0.58 to 1.23, $p < .001$), closely approximating the identity line (Figure 1, $R^2 = 0.81$).

Given that the effective sample size for this regression analysis was N = 11 categories, we validated this effect by conducting a non-parametric bootstrap analysis. Specifically, we resampled both category-level measurements with replacement 10,000 times, and then computed the standardized regression coefficient for their relationship. The proportion of bootstrapped slopes exceeding the observed effect in absolute value was 0.0004 (0.04%), confirming our findings.

Discussion

Our results illustrate a telling correspondence between people's explicit reports of prioritization in learning and memory for items, on the one hand, and their sampling from category-level representations when asked "what comes to mind?", on the other. Prior work shows that people often prioritize memory for high-value items. And, prior work also shows that when asked to spontaneously generate items from a given category, people tend to preferentially sample high-value items. What makes the present study especially powerful, however, is the correlation between the exceptions to these general principles. We show that there are some categories for which people say that it is important to prioritize learning and memory for bad items, rather than good ones: Categories like diseases and terrorist organizations. Then, we find that these very same categories also generate samples of "what comes to mind" biased towards bad, rather than good, items. Moreover, the 11 categories that we test present a graded variety of measures of each type, and we find that the best fitting regression line relating these graded measures falls very close to the identity line that would be predicted by the simplest version of our hypothesis. These results are consistent with a psychological coupling between learning and memory for individual items and resulting category-level representations.

One limitation of the current study is that it does not fully distinguish between internal memory prioritization and external exposure effects. People may encounter certain items more frequently—such as dangerous diseases or terrorist organizations—because others deem them important to remember. Thus, differences in recall may partly reflect societal prioritization during learning. Future work could address this by separately measuring or controlling for environmental exposure rates.

Computational Experiments

Experiment 1

Our behavioral study shows that when people think it is important not to be wrong about certain types of objects, those types of objects are overrepresented in the category members they call to mind. We now aim to provide a computational model of this phenomenon, drawing on existing computational methods for studying reconstructive memory. Specifically, we make use of variational autoencoders (VAEs). A VAE can both (a) use information about the distribution of objects within a category to accurately reconstruct individual token objects within that category and (b) generate new objects from that distribution. We treat this as a computational model of the human psychological process of "what comes to mind." In this computational experiment, we create a VAE with a weighted loss function, such that reconstruction loss is higher for certain types of objects than for others. We then ask whether introducing a weighted loss function for reconstruction has an impact on generation, i.e., whether a change in which objects it is most important to remember leads to a change in which objects "come to mind."

Stimuli For these computational experiments, we switch from the more complex stimuli used in the behavioral experiments (athletes, diseases, universities, etc.) to more simple visual stimuli. Specifically, we use images of handwritten digits (handwritten 0s or handwritten 7s).

We then assign these visual stimuli to have either a high loss penalty or a low loss penalty. Thus, if we assign handwritten 7s to have a high loss penalty, each of the different handwritten 7s will play the same role as one of the differ-

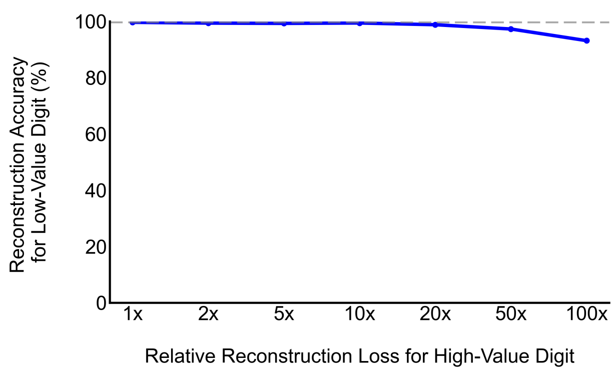

Figure 2: Reconstruction accuracy of low-value digits for different relative reconstruction loss penalties applied to high-value digits during training.

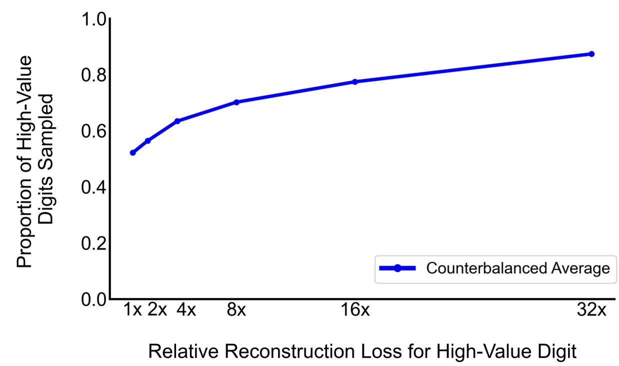

Figure 3: Proportion of high-value digits sampled as a function of relative reconstruction loss, averaged across counterbalancing of 0 and 7 as the high-value class.

ent extraordinary athletes in the behavioral experiment. Similarly, if we assign handwritten 0s to have a low loss penalty, each of the different handwritten 0s will play the same role as one of the different mediocre athletes in the behavioral experiment.

Images were sourced from the Modified National Institute of Standards and Technology database (MNIST) dataset, with 5,923 images per digit class. During training, we assigned either 0s or 7s to be ”high value” and applied a relatively larger penalty to its reconstruction loss. For example, ”2x” penalty doubles the reconstruction loss for high-value digits compared to low-value digits, and ”4x” penalty quadruples it.

Selecting penalties for reconstruction loss The first step was to select the reconstruction loss penalties used in Experiments 1 and 2. It was crucial to ensure that the VAE maintained good reconstruction performance for all input items, since later experiments—focused on sampling from the decoder—would be invalid if the model could not reliably reconstruct the low-value digit class. To address this, we trained VAEs across a range of penalty values applied to the high-value class, and then measured the accuracy of reconstructing low-value digits. Specifically, after training each VAE, we passed all low-value training images through the encoder and decoder, and then classified the outputs as 0s or 7s using a Convolutional Neural Network (CNN) trained on the MNIST dataset. The penalty values tested were 1x, 2x, 5x, 10x, 20x, 50x, and 100x, with 0s as the high-value class and 7s as the low-value class.

Results Figure 2 shows the reconstruction accuracy of low-value digits across different reconstruction loss penalties applied to high-value digits. Given these results, to ensure that the VAE maintained a reliable conceptual representation of both digit classes, we selected a cutoff of 32x as the maximum penalty applied to the high-value digit class in subsequent experiments. This value allows for a meaningful variation in penalty values while ensuring that the reconstruction accuracy of the low-value digit class remains sufficiently high.

Generating samples Next, we assess how the proportion of high-value digits sampled from the VAE changes as we increase the penalty applied to the reconstruction loss of high-value digits. To do so, we trained a series of VAEs with penalties for the high-value digit class of: 1x, 2x, 4x, 8x, 16x, 32x. We then used each VAE to generate 100,000 images of digits by first sampling 100,000 two-dimensional latent vectors from a standard normal distribution (the latent space prior), and then passing these vectors through the VAE's decoder to produce reconstructed images.

Results To classify the decoder-generated images as 0s or 7s, we similarly used a CNN trained on the MNIST dataset to classify each VAE-generated image. To ensure robustness, we counterbalanced the high-value designation across digit classes, resulting in 12 VAEs (6 with 0s as high-value, 6 with 7s as high-value). Figure 3 plots the proportion of high-value digits sampled as a function of their loss weight. Logistic regression confirmed that increasing loss weight significantly increases the likelihood of sampling a high-value digit (β = 0.0593, OR = 1.061, 95% CI: [1.061, 1.062], $p < .001$).

Latent Space Analysis Finally, to understand how the latent space representation evolves under differential reconstruction penalties, we analyzed the embeddings of training images. Specifically, we first passed all 11,846 training images through each VAE's encoder. Then, we collected the resulting two-dimensional latent mean vectors for each image, which encode the image's latent representation. Lastly, we plotted the resulting latent representations. Since there was no option to counterbalance the high-value digit class for the latent space plots, we used 0 as the high-value class and 7 as the low-value class. To quantify the changes in the latent space as the reconstruction penalty for one class increases, we also computed the average pairwise Euclidean distance between latent representations within each digit class across different reconstruction penalties.

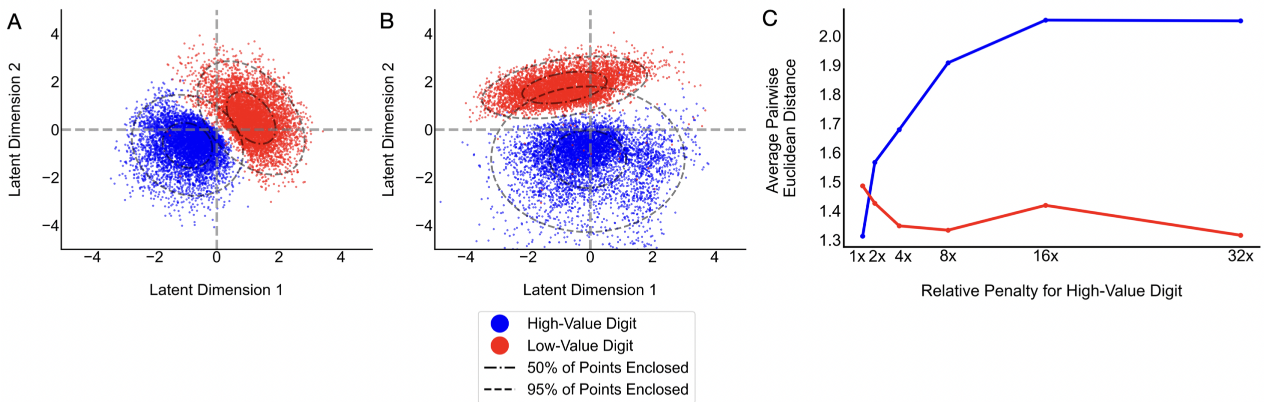

Figure 4: (A) Latent space visualization of high-value (blue) and low-value (red) digits under a 1x reconstruction penalty. (B) Same visualization under 16x penalty. Ellipses enclose 50% and 95% of points. (C) Average pairwise Euclidean distances for high-value and low-value digits across varying penalties.

Results Figure 4(A) illustrates the latent space when both digit classes have equal reconstruction penalties. The two classes are well-separated with similar distributions. In Figure 4(B), the reconstruction loss for high-value digits is 16 times higher (16x), causing high-value digits to spread more widely, while low-value digits cluster more tightly. This shift expands the representational space for high-value digits, as seen in the 50% and 95% contour lines. Figure 4(C) quantifies these changes: the average pairwise Euclidean distance among high-value digits increases with higher penalties (β = 0.0198, $p < .001$), while low-value digits become slightly more compressed (β = -0.0032, $p < .001$). These results suggest that prioritizing reconstruction accuracy for high-value digits reallocates latent space disproportionately toward them.

Experiment 2

In this next computational experiment, we analyze the functional form of the relationship between changes in the loss penalty and changes in what comes to mind. Existing behavioral studies with humans show something surprising about the precise relationship between changes in value and frequency and changes in what comes to mind. Specifically, they show that the impact of value and frequency on what comes to mind is not *additive* but rather *multiplicative*. That is, there is an interaction of value and frequency in determining what comes to mind. We now ask whether our computational model exhibits an analogous pattern.

Training setup Experiment 2 followed the same setup as Experiment 1 but introduced a second manipulation: the proportion of high-value digit images in the training set (hereafter, ”frequency”). We tested 21 frequency levels, ranging from 0 (only low-value digits) to 1 (only high-value digits) in increments of 0.05, along with six reconstruction penalty conditions (1x, 2x, 4x, 8x, 16x, 32x). To ensure robustness, we counterbalanced the high-value digit designation across both digit classes, resulting in 252 VAEs trained in total.

Generating samples As in Experiment 1, we sampled 100,000 images from each trained VAE and classified them using a CNN trained on the MNIST dataset.

Models To evaluate the functional form of the relationship between frequency, relative penalty for reconstruction loss, and the probability of generating high-value digits, we fitted two models to the results of Experiment 2. These models are based on theoretical frameworks developed in prior research by Bear et al. (2020). Both models assume that a digit's probability of being generated depends on its statistical frequency $F(x)$ and its reconstruction penalty $L(x)$.

Additive model The additive model assumes a weighted sum of frequency and loss:

$$\text{Add}(x;w) = wF(x) + (1-w)L(x)$$

where w is a free parameter that controls the relative influence of frequency vs. loss. The probability of generating a high-value digit $P(x_{\text{high}})$ is determined by normalizing across both classes:

$$P(x_{\text{high}}) = \frac{\text{Add}(x_{\text{high}};w)}{\text{Add}(x_{\text{high}};w) + \text{Add}(x_{\text{low}};w)}$$

where $\text{Add}(x_{\text{high}};w)$ and $\text{Add}(x_{\text{low}};w)$ refer to the additive scores for high- and low-value digits, respectively. Moreover, $\text{F}(x_{\text{high}})$ is the proportion of high-value digits in the training set, $\text{F}(x_{\text{low}}) = 1 - \text{F}(x_{\text{high}})$, and $\text{L}(x_{\text{low}}) = 1$.

Multiplicative model The multiplicative model assumes that frequency and loss interact multiplicatively, with a parameter B controlling the influence of loss:

$$\text{Mult}(x;B) = F(x)L(x)^B$$

The probability of generating a high-value digit is:

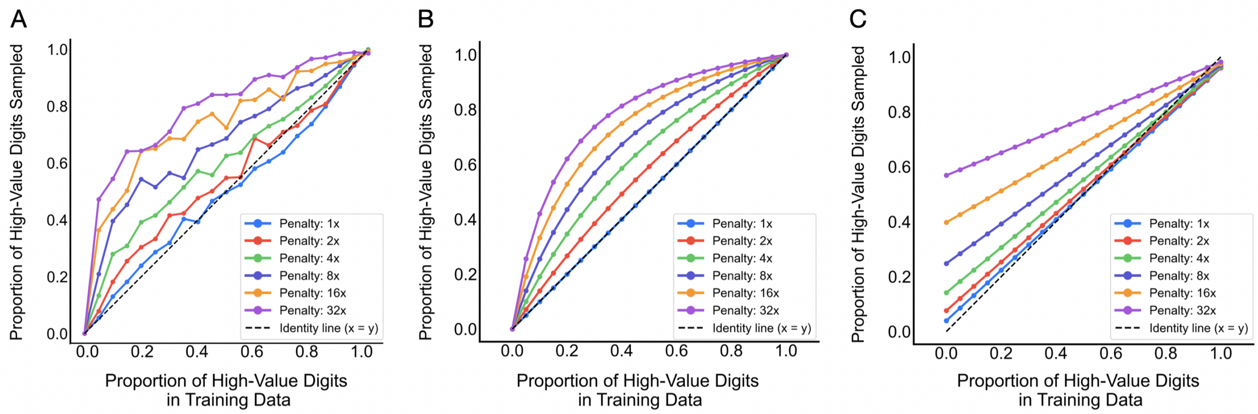

Figure 5: (A) Actual data: proportion of high-value digits sampled across different penalties. (B) Predictions from multiplicative model. (C) Predictions from additive model. Lines represent penalties (1x–32x), dashed line represents the identity line ($x = y$).

$$P_{\text{high}} = \frac{\text{Mult}(x_{\text{high}}; B)}{\text{Mult}(x_{\text{high}}; B) + \text{Mult}(x_{\text{low}}; B)}$$

where $\text{Mult}(x_{\text{high}}); B)$ and $\text{Mult}(x_{\text{low}}; B)$ refer to the multiplicative scores for high- and low-value digits, respectively, and $\text{F}(x_{\text{high}})$, $\text{F}(x_{\text{low}})$, $\text{L}(x_{\text{high}})$, $\text{L}(x_{\text{low}})$ are defined as before.

Model Fitting Procedure Both models were fit to the data using negative log-likelihood as the objective function. The parameters w and B were optimized using the L-BFGS-B method, a gradient-based optimization algorithm. The fit of each model was evaluated using the negative log-likelihood (NLL), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).

Results Table 1 presents fit statistics for each model. The multiplicative model achieves a lower NLL, AIC, and BIC, indicating a better overall fit compared to the additive model. Moreover, Figure 5 displays a comparison of actual data (Panel A) with the predictions of the multiplicative model (Panel B) and additive model (Panel C) using their optimal parameters, for the proportion of high-value digits sampled across different penalties. As is evident, the multiplicative model captures several features of the model results that the linear model does not.

Discussion

Our computational experiments yield several findings. When applying a weighted loss function during the training of a VAE, subsequent samples drawn from the latent space overrepresent the kinds of items subjected to greater loss. Visualization of the latent space reveals that these items tend to be embedded more centrally within the latent space, and also with greater dispersion. Furthermore, just as frequency and value interact multiplicatively to determine "what comes to mind" in humans, frequency and loss interact multiplicatively to determine the distribution of samples from the latent space of a VAE.

Table 1: Negative log-likelihood (NLL), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) scores for multiplicative and additive models' fit to the data.

Model	NLL	AIC	BIC
Multiplicative	12,094,470	24,188,930	24,188,950
Additive	12,453,470	24,906,950	24,906,960

General Discussion

Our findings help offer an explanation for when and why people's representations of a category are biased towards valuable items within the category. We propose that (1) valuable items are often more important to remember, (2) capacity-limited memory processes prioritize the items most important to remember, and (3) category representations generated by these memory systems overrepresent prioritized items.

This model predicts that valuable items (i.e., those that are desirable) will only be overrepresented in category representations when they are prioritized for memory. In cases where undesirable items are prioritized, these items should instead be overrepresented at the category level. Our behavioral experiment provides evidence for precisely this effect. As a practical matter, we conjecture that high value, ideal, or desirable items tend to be prioritized in memory more than low value or undesirable items. This might be especially true for categories of items that we can choose (like foods, friends, or artifacts), because it will be most important to accurately represent the strongest candidates for choice.

Meanwhile, our computational experiments use VAEs to show that, in a system where category-level representations aid the encoding and reconstruction of item-level representations, prioritizing memory for some items increases their influence on the category-level representation.

Together, these findings lend support to the proposed link between prioritized memory for items and the effect of value on category representations.

References

Bandyopadhyay, S., Dion, C., Libon, D. J., Price, C., Tighe, P., & Rashidi, P. (2022). Variational autoencoder provides proof of concept that compressing cdt to extremely low-dimensional space retains its ability of distinguishing dementia. *Scientific reports*, *12*(1), 7992.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of experimental psychology: learning, memory, and cognition*, *11*(4), 629.

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition*, *194*, 104057.

Foster-Hanson, E., & Rhodes, M. (2019). Is the most representative skunk the average or the stinkiest? developmental changes in representations of biological categories. *Cognitive psychology*, *110*, 1–15.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological science*, *17*(9), 767–773.

Horwath, E. A., Rouhani, N., DuBrow, S., & Murty, V. P. (2023). Value restructures the organization of free recall. *Cognition*, *231*, 105315.

Keller, T. A., Gao, Q., & Welling, M. (n.d.). Modeling category-selective cortical regions with topographic variational autoencoders. In *Svrhm 2021 workshop@ neurips.*

Knobe, J., & Cushman, F. (2023). The common effect of value on prioritized memory and category representation. *Trends in Cognitive Sciences*, *27*(10), 892–900.

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological review*, *125*(1), 1.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic bulletin & review*, *21*, 629–636.

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature neuroscience*, *21*(11), 1609–1617.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, *7*(4), 573–605.

Rouhani, N., Niv, Y., Frank, M. J., & Schwabe, L. (2023). Multiple routes to enhanced memory for emotionally relevant events. *Trends in Cognitive Sciences*, *27*(9), 867–882.

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in cognitive sciences*, *14*(10), 464–472.