Hybrid-Similarity Exemplar Model of Context-Dependent Memorability

Robert M. Nosofsky (nosofsky@indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
Bloomington, IN 47405 USA

Adam F. Osth (adam.osth@unimelb.edu.au)
School of Psychological Sciences, University of Melbourne
Victoria 3010 Australia

Abstract

We conduct tests of a hybrid-similarity exemplar model on its ability to account for the context-dependent memorability of items embedded in high-dimensional category spaces. According to the model, recognition judgments are based on the summed similarity of test items to studied exemplars. The model allows for the idea that "self-similarity" among objects differs due to matching on highly salient distinctive features. Participants viewed a study list of rock images belonging to geologically defined categories where the number of studied items from each category was manipulated, and their old-new recognition performance was then tested. With a minimum of parameter estimation, the model provided good accounts of changing levels of memorability due to contextual effects of category size, within- and between-category similarity, and the presence of distinctive features. We discuss future directions for improving upon the current predictions from the model.

Keywords: recognition memory, memorability, computational modeling, exemplar model, high-dimensional similarity space

Introduction

Modern research indicates that images of individual real-world objects vary considerably in their memorability: some objects appear to be relatively easy to remember, whereas memories of other objects are difficult to maintain. This differential memorability is often studied in tasks of old-new recognition, in which observers are presented with lists of items to study, followed by test lists composed of old and new items. Certain old items are consistently recognized with higher accuracy than are others; and people also tend to false-alarm to certain new items more than others (Bainbridge, 2019; Kramer et al., 2023).

There is a modern computer-science literature in which researchers have reported successful attempts to build deep-learning networks that can predict which images will and will not be memorable (Bylinskii et al., 2022; Dubey et al., 2015; Khosla et al., 2015; Needel & Bainbridge, 2022). Although these deep-learning networks have provided impressive predictions of individual-item memorability, a limitation is that they don't provide insights into the psychological reasons why certain images are more memorable than others.

A more fundamental limitation is that these deep learning models assume that variations in memorability are entirely due to the stimulus itself. An extensive cognitive-psychology literature has shown, however, that memory is highly dependent on the context of the study list in which the items are embedded - a stimulus can be memorable or non-memorable depending on the other stimuli that accompany it on the study list. One example is the demonstration of category size effects, where increasing the number of items from the same category on the study list greatly increases false alarm rates to novel members of the same category (e.g., Robinson & Roediger, 1997; Shiffrin, Huber, & Marinelli, 1995). We discuss other examples below.

There is a previous cognitive-psychology modeling literature that has also attempted to account for differences in individual-item memorability. One approach that has shown some success involves the application of summed-similarity exemplar models (Nosofsky, 1991; Nosofsky et al., 2011). In this approach, studied items are presumed to be represented as individual exemplars in memory. The exemplars are embedded in multidimensional similarity spaces. Test probes are assumed to give rise to a global activation of memory via a summing of similarity to the exemplars in the space, where similarity between any pair of exemplars is a decreasing function of their distance in the space. The greater the summed similarity, the greater is a "familiarity signal" to which the test probes gives rise; and the greater the familiarity, the higher is the probability that the test probes will be judged to be old (cf. Gillund & Shiffrin, 1984; Osth & Dennis, in press).

The summed-similarity signal can be conceptualized as being composed of two components. For old test probes, one component is the self-match of the test probe to its own representation in memory. And for both old and new test probes, the second component is inter-item similarity: the similarity of a test probe to the memory representations of other exemplars besides itself. In general, summed-similarity exemplar models predict higher recognition rates for old test probes (hit rates) than for new test probes (false-alarm rates) because the self-match of an old test probe to its own exemplar trace provides a significant boost to the summed-similarity signal. However, the models also predict high false-alarm rates for new test probes that are highly similar to old exemplars stored in memory (such as category prototypes), because such items will also give rise to a large-magnitude summed-similarity signal. Previous studies have been reported that illustrate that the summed-similarity exemplar model can indeed yield very accurate predictions of

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

individual-item hit and false-alarm rates (e.g., Nosofsky, 1991; Nosofsky et al., 2011).

A major limitation of original versions of summed-similarity exemplar models, however, is that they failed to account for well-known phenomena from the face-recognition literature in which face targets with highly distinctive features (such as scars) often have higher hit rates than do more typical faces (e.g., Busey & Tunnicluff, 1999; Valentine & Ferrara, 1991). Because distinctive faces are presumably located in isolated regions of multidimensional similarity space, standard summed-similarity exemplar models predicted that distinctive faces would have *lower* summed similarity than typical ones, so they failed to predict the hit-rate advantage for old distinctive faces.

To address this limitation, Nosofsky and Zaki (2003) proposed and tested a hybrid-similarity exemplar model. As in classic versions of the models, test probes are assumed to give rise to a global activation of memory via a summing of similarity to studied exemplars, where similarity is assumed to be a decreasing function of distance in a multidimensional similarity space. However, unlike in the classic versions of the model, a key idea was that items with highly distinctive features may give rise to a greater degree of *self-match* than do items without such features (see below for formal details). This high degree of self-match can potentially dominate the summed-similarity signal, leading to high recognition hit rates for old items with such features. Nosofsky and Zaki (2003) demonstrated that the hybrid-similarity exemplar model provided good quantitative accounts of individual-item old-new recognition in experiments in which artificial distinctive features were manually added to highly controlled stimuli embedded in low-dimensional similarity spaces (e.g., color patches varying in lightness, saturation and hue).

Despite the successes briefly reviewed above, a fundamental limitation of experimental tests of the cognitive-modeling approach to individual-item memorability is that the experiments all involved use of highly simplified perceptual stimuli and artificial category structures, rather than the real-world, high-dimensional objects that are the focus of modern work. To begin to address this limitation, in recent research Nosofsky and Meagher (2022) and Meagher and Nosofsky (2023) applied the hybrid-similarity exemplar model to predict individual-item old-new recognition for a set of rock-image stimuli. They used multidimensional scaling methods to embed the rock stimuli in a high-dimensional similarity space. In addition, they collected ratings of the extent to which individual rock images contained highly distinctive features that made them stand out from other items in the set. Combining these sources of information, Nosofsky and Meagher (2022) and Meagher and Nosofsky (2023) were able to apply the hybrid-similarity exemplar model to account reasonably well for hit- and false-alarm rates associated with the individual rock images in their experiments.

The main purpose of the present work was to pursue Meagher and Nosofsky's (2023) recent investigations of the hybrid-similarity exemplar model in this real-world high-dimensional rocks domain with still more challenging tests. A critical new question concerned the extent to which the model could capture the *context-dependent* nature of memorability. We conducted an experiment in which participants first studied a list of rock images organized into categories and were then tested on their ability to judge whether test probes from the categories were old or new. Following previous research, a key variable that we manipulated was category size: some categories were composed of large numbers of studied items and other categories were composed of small numbers of studied items (see also Konkle, et al., 2010). The studied categories also varied in their degrees of within- and between-category similarity. Finally, the rock images varied in the extent to which they contained distinctive features that were likely to make them stand out from other objects in the set. The goal was to test the ability of the hybrid-similarity model to capture quantitatively how old-new recognition performance varied as a function of all these variables.

Previously, such models have been demonstrated to capture increases in false-alarm rates with increasing category size because increasing the number of similar studied items increases the summed similarity between the probe and the contents of memory (Hintzman, 1988; Nosofsky et al., 2011; Shiffrin et al., 1995). However, to our knowledge, no models have simultaneously accounted for the joint effects of category size, within- and between-category similarity, and presence of distinctive features on variations in memorability for individual old and new items.

We organize the remainder of our article as follows. First, we report the experimental method. Next, we provide a brief presentation of the formal hybrid-similarity exemplar model itself, because the experimental results are best understood within the framework of the model. Finally, we report the experimental and modeling results along with a discussion of future directions for potentially improving the ability of the model to predict the context-dependent nature of individual-item memorability.

Experiment Method

Subjects

The subjects were 203 undergraduates from Indiana University who received credit towards an introductory psychology requirement.

Stimuli

The stimulus materials were a set of 240 rock images organized into 24 categories of 10 samples each. Most of the categories were ones defined in the geologic sciences. However, to strengthen the intended category-size

manipulation, we thought it important to use category structures in which within-category similarities tended to clearly exceed between-category similarities. This structural constraint is not always satisfied for geologically-defined categories. Hence, in constructing our stimulus set, we did not include outlier samples that we judged to be more similar to members of contrast categories than to members of their own category. In addition, in several cases we created our own "psychological" categories. For example, without use of physical and chemical tests, samples of shale and slate cannot be discriminated. Hence, we defined a "gray slate" category composed of gray samples of both slate and shale and a "colored shale" category composed of colored samples of slate and shale (see Procedure section for further details). To take a second example, it is extremely difficult to discriminate between amphibolite and gabbro based solely on visual inspection. Hence, we used light-shopping techniques to make dark gray in color all samples of gabbro and added a blue tint to all samples of amphibolite. A listing of the set of 24 rock categories along with a set of descriptors for the categories used in the experiment is provided at osf.io/mc6ys.

Procedure

Each subject saw a single list of 75 study items. For each subject, each of the 24 categories was randomly assigned to a different category-size condition. There were five size-1 categories, 5 size-2, 5 size-4 and 5 size-8 categories. In addition, there were 4 size-0 categories that didn't appear on the study list. For each individual subject, the rock-image samples chosen for study in each size condition were chosen randomly from each of the 24 categories. On each trial of the study phase, a rock image was presented on the computer screen together with a table listing the 24 category names along with a descriptor of each category (e.g., "Amphibolite. Blue, coarse-grained, rough."). Subjects chose the category whose descriptor they believed best characterized the rock image. Following the response, corrective feedback was then provided. The purpose of this manipulation was to keep subjects engaged in the task and to potentially strengthen the intended category-size manipulation by having subjects encode the images in terms of the category descriptors. The order of presentation of the 75 study images was fully randomized for each subject. Following the study phase there was an 82-trial test phase in which we presented two randomly sampled old targets and two new lures from each of the categories (1 target and 1 lure from the size-1 categories; and 2 lures from the size-0 categories) and subjects judged whether each test probe was old or new. The order of presentation of the test items was fully randomized for each subject and no corrective feedback was provided.

Fitting the hybrid-similarity exemplar model to the data requires that it be provided with an input space that specifies the coordinates of each object along a set of psychological dimensions. As described extensively in previous articles (e.g. Nosofsky et al., 2018), we used multidimensional scaling (MDS) techniques based on the modeling of similarity-judgment data to derive this space. As in the previous work, based on a combination of overall fit and interpretability of the derived dimensions, we settled on use of an 8-dimensional scaling solution for the present project (for a detailed statement and illustration of the underlying dimensions of the space, see, e.g., Nosofsky et al., 2018 and Meagher & Nosofsky, 2023). In addition, following Meagher and Nosofsky (2023), an independent group of participants provided ratings of the extent to which each rock image contained distinctive features that made it stand out from other objects in the set. The top row of Figure 1 provides examples of rock images that received high distinctive-feature ratings, and the bottom row provides examples of rock images that received low distinctive-feature ratings.

Figure 1. Examples of rock images that received high vs. low distinctive-feature ratings.

Hybrid-Similarity Exemplar Model

In this article we focus on a baseline version of the model that uses a minimum of parameter estimation. According to the model, each test item i is presumed to give rise to a familiarity signal F_i. The probability that the test item i is judged to be "old" is then given by $P(\text{Old}|i) = F_i / (F_i + k)$, where k is a response-criterion parameter. Familiarity F_i is computed by summing the hybrid-similarity of test item i to each of the individual study exemplars j. These hybrid-similarities (hs_{ij}) are computed as follows. First, the continuous-distance of i to j is computed from the MDS solution derived for the exemplars by using a Euclidean metric, $d_{ij}=[\sum|x_{im} - x_{jm}|^2]^{1/2}$, where x_{im} is the value of item i on dimension m. Following Shepard (1987), the continuous-dimension similarity of item i to exemplar j is then assumed to be an exponential decay function of the distance between i and j, $s_{ij} = \exp(-c \cdot d_{ij})$, where c is an overall sensitivity parameter that describes the rate at which similarity declines with distance. The hybrid similarities (hs_{ij}) are computed by multiplying the continuous-dimension similarities by factors associated with matching or mismatching distinctive features. The hybrid-similarity ideas are motivated by Tversky's (1977) classic feature-contrast model of similarity, which assumes that common features boost similarity (including self-similarity); and that features that differ between two items reduce

similarity. Thus, the self-similarity of item *i* to itself is assumed to be *boosted* by the factor $\exp(\beta \cdot \delta_i)$, where δ_i is the mean distinctive-feature rating for item *i*. The intuition, for example, is that a face with a highly distinctive feature such as a scar is likely to produce a strong match to its trace in memory. By contrast, the hybrid-similarity of item *i* to a mismatching exemplar *j* is assumed to be *reduced* by the factor $\exp(-\alpha \cdot \delta_i)$. Our assumption is that to the extent that item *i* is judged to have a highly distinctive feature, it is unlikely that the feature would be shared by some mismatching exemplar in the stimulus set, so the natural assumption is that the distinctive feature would tend to reduce similarity between item *i* and a mismatching exemplar *j*. The factors β and α in the above hybrid-similarity equations are freely estimated scaling parameters.

Experimental and Formal Modeling Results

We fitted the model to the *individual-trials* data of the individual participants by using a maximum-likelihood criterion. Note that the baseline model makes use of only 4 free parameters: the response-criterion parameter *k*, the sensitivity parameter *c*, and the distinctive-feature scaling parameters β and α. For simplicity, here we report results in which the parameters were held fixed across all subjects. The purpose is to demonstrate that even a low-parameter baseline version of the model is able to capture many of the fundamental phenomena that we observed in the experiment. We also conducted hierarchical Bayesian modeling to account for individual differences, but the predictions for the aggregate data were essentially identical to the ones we report here. We emphasize that in the ensuing presentation of summary results and predictions, the 4 parameters are held fixed across all summary data sets.

Figure 2 shows that false-alarm rates for new lures increased dramatically with increases in category size, a trend that is predicted well by the model ($r = 0.997$) and is consistent with

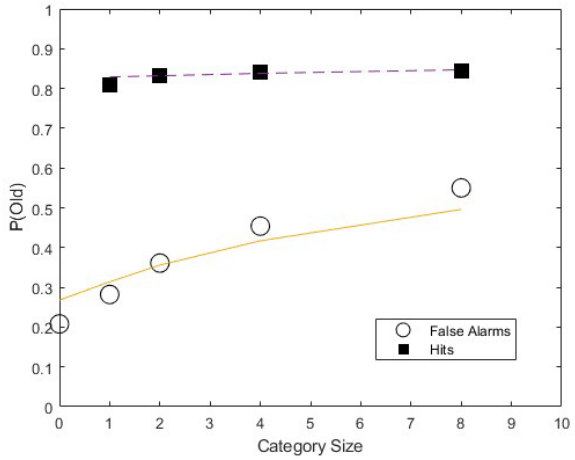

Figure 2. Mean observed and predicted hit and false-alarm rates as a function of category size.

prior studies manipulating category size. As category size increases, summed similarity of novel test probes to the study items increases, so false-alarm rates increase. The figure also shows that there was a much smaller increase in hit rates with increases in category size, a trend that is again well captured by the model ($r = 0.83$). Clearly, the model also captures the overall magnitude of the hit and false-alarm rates for the different category-size conditions.

Figure 3 plots the observed against predicted false-alarm and hit rates for each of the 24 categories themselves, averaged across the different size conditions. The model does a good job of capturing the false-alarm rates associated with the 24 categories ($r = 0.84$). Apparently, some categories have

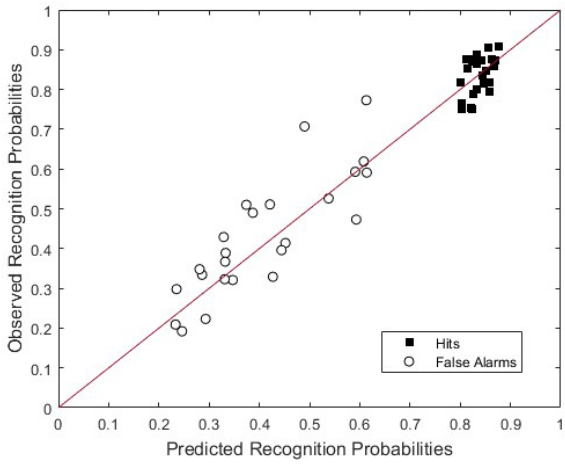

Figure 3. Mean observed against predicted hit and false-alarm rates for each of the 24 categories.

greater degrees of within-category similarity than do others, and the greater the degree of within-category similarity, the greater is the summed similarity. There is not nearly as much variation in the hit rates as for the false-alarm rates in these data. Nevertheless, the hybrid model captures reasonably well the variability in the overall hit rates across the 24 categories ($r = 0.50$). It is crucial to point out that without making use of the distinctive-feature scaling parameter β, the summed-similarity exemplar model failed to account for any of the variance in the category-level hit rates. Apparently, some categories contain members with more distinctive features than do other categories, and those categories tend to have higher overall hit rates.

In Figure 4 we plot the observed-against-predicted hit-minus-false-alarm rates for each of the 24 categories, which provide a rough measure of the extent to which observers could discriminate the old members of each category from the new members. The model does a good job of capturing this measure of performance as well ($r = 0.89$). Note that to capture cases in which the hit-minus-false-alarm rate is high, the model needs to simultaneously predict a relatively high hit rate and a relatively low false-alarm rate for the category.

According to the model, this tends to occur when the items in the category contain highly distinctive features: the self-matches of target items on their distinctive features will tend to boost their hit rates, but the mismatch of lure items on their distinctive features will tend to reduce their false-alarm rates.

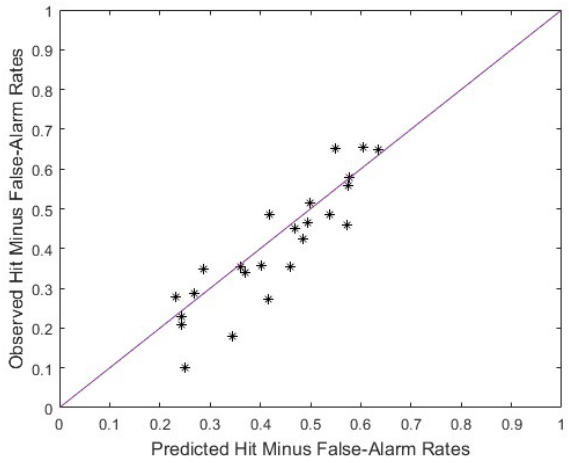

Figure 4. Mean observed against predicted hit-minus-false-alarm rates for the 24 categories.

In Figure 5 we plot the observed-against-predicted false-alarm rates for lures from the size-0 categories. There is a significant correlation between the observed and predicted false-alarm rates (r = 0.57), but the model systematically under-predicts the overall magnitude of these false-alarm rates. One reason why the correlation arises is because there are differing degrees of between-category similarity across the 24 categories: Items from size-0 categories that are similar to members of *other* studied categories will tend to have higher false-alarm rates. We discuss below possible reasons why the current version of the model tends to under-predict the overall magnitude of the false-alarm rates for the size-0 categories.

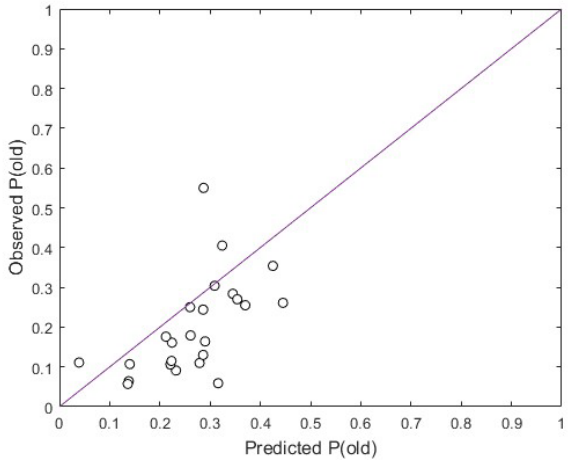

Figure 5. Mean observed against predicted false-alarm rates for the size-0 categories.

In our next analysis, we divided the 240 individual rock images into four equally sized distinctiveness-rating bins and computed the observed and predicted mean hit rates and false alarm rates for each bin. The results are shown in Figure 6. As can be seen, the mean hit rates increase with increases in rated distinctiveness, whereas the false-alarm rates decrease. The hybrid-similarity model does an excellent job of capturing both effects (r_H = 0.95, r_{FA} = 0.92). Recall that the hybrid-similarity model computes a β scaling parameter for estimating the boost in self-similarity that arises due to matching distinctive features. Crucially, if β is held fixed at zero, then the model fails to predict any increase in hit rates as a function of rated distinctiveness. The model predicts decreasing false-alarm rates with increases in rated distinctiveness for two reasons. First, those items with high values of rated distinctiveness tend to lie in isolated regions of the derived MDS space for the rocks, resulting in low values of summed similarity. Second, the presence of a highly distinctive feature reduces even more the similarity of a lure item to the studied exemplars.

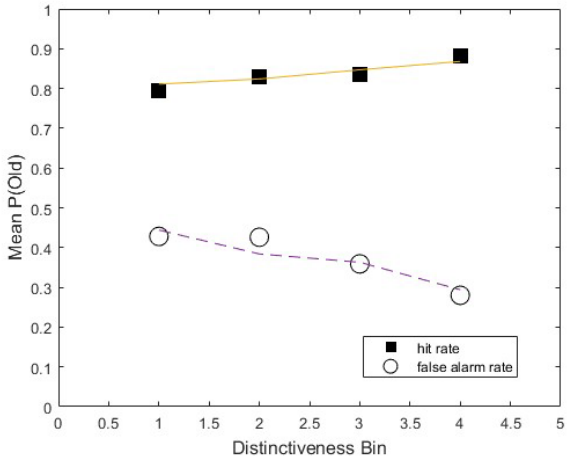

Figure 6. Mean observed and predicted hit and false-alarm rates as a function of distinctive-feature rating bin.

Finally, in a rather ambitious analysis, we computed the observed and predicted false-alarm and hit rates associated with the individual 240 items, averaged across the different category-size conditions of the experiment. Sample size for each individual item is small because each participant was tested with only a small subset of items from each category; therefore, these results need to be interpreted with caution. However, as shown in Figure 7, the model captures much of the variation in the individual-item false-alarm rates (r = 0.63). Although the model also captures some of the individual-item variation in the hit rates (r = 0.32), accounting for this aspect of the data remains a big challenge. In part, the lower correlation for the hit rates may reflect that there is less overall variability in the hit-rate than in the false-

alarm-rate data (so less total variability to account for). However, we discuss below various routes of future research that may yield improved accounts of the individual-item hit-rate data as well.

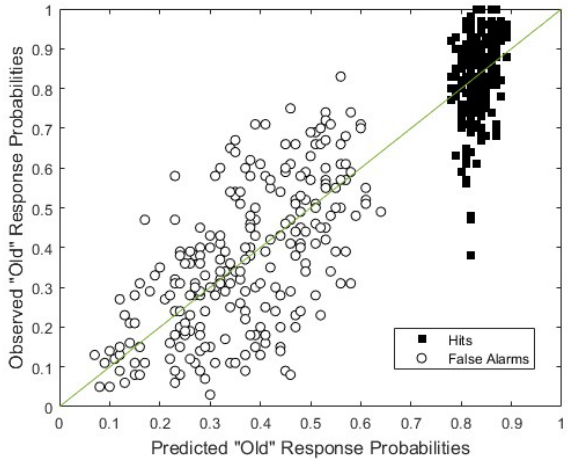

Figure 7. Observed against predicted hit and false-alarm rates for the individual 240 items.

Discussion

Using a minimum of parameter estimation, a baseline version of a global-familiarity hybrid-similarity exemplar model was able to provide reasonably good accounts of a wide variety of phenomena involving the context-dependence of memorability in a complex, high-dimensional category domain. The phenomena included effects of category size, within- and between-category similarity, and the presence of distinctive features on both false-alarm rates for lures and hit rates for old study items. A core idea in the present modeling is that the presence of unique distinctive features is assumed to boost self-similarity but to decrease interitem similarity. This results in an interesting dynamic with respect to the prediction of how hit rates may vary with the presence of distinctive features. Whether or not hit rates are predicted to increase depends on whether the boost to self-similarity exceeds the reduction in inter-item similarity in the overall summed-similarity computation. Here we tended to see boosts in hit rates with presence of distinctive features but it is conceivable that alternative contexts could lead to different patterns of results.

There are various directions that are likely to yield still improved quantitative accounts of the present data. First, more elaborate versions of the model make allowance for nonlinear relations between psychological familiarity and summed similarity and between psychological and rated distinctiveness (for details, see Meagher & Nosofsky, 2023). Second, we relied here on use of an 8-dimensional MDS solution for the objects derived from an independent set of similarity-ratings data. Past work has suggested, however, that for purposes of learning categories, participants make use of a variety of supplementary diagnostic dimensions of the rock images not revealed by the MDS techniques (Nosofsky et al., 2020). Incorporating these supplementary dimensions is likely to significantly improve the detailed quantitative predictions of old-new recognition as well. Third, past research provides evidence that old-new recognition judgments are influenced by the structure and sequence of items presented during the test phase itself (e.g., Criss et al., 2011; Fox & Osth, 2023; Osth, Jansson, Dennis, & Heathcote, 2018). Such test-phase information has not yet been incorporated in the present version of the model. Fourth, participants bring with them to the experiment a past history of memories of certain categories of rocks. Expanding the model with pre-loaded forms of rock familiarity is also likely to yield improved accounts of these old-new recognition data. Fifth, the present modeling did not take account of the well-established idea that observers may give differential attention weight to the dimensions that compose the objects in making their categorization and recognition judgments (Nosofsky, 1991).

Perhaps most important, another limitation of the present research approach is that we relied on use of only generic ratings of presence of distinctive features as a source of input to the model. The conception advanced by Nosofsky and Zaki (2003) in advancing the hybrid-similarity exemplar model was that the distinctive features of objects lie "outside" the continuous-dimension space in which the objects are embedded, causing them to "stand out" from the other objects in the set. Using current psychological-scaling methods, the presence of such features is reflected only indirectly in the continuous-dimension MDS solutions for objects by locating such objects in more isolated parts of the continuous-dimension space. A crucial goal of future work is to instead develop techniques that can identify the specific distinctive features that play a role in influencing memorability and incorporate them as part of the feature space itself. A potentially fruitful route to achieving this goal may be to merge the impressive deep-learning approaches to accounting for memorability with the present type of cognitive-modeling approach. These future lines of research will also need to account for the idea that the extent to which an object is judged to have a "distinctive" feature is itself a highly context-dependent phenomenon. For example, a feature is likely to be judged as highly distinctive in study contexts in which it rarely occurs but to be lacking in distinctiveness in study contexts in which it is common. In addition, it will be interesting to explore in future research the precise manner in which the variables of category size and feature distinctiveness may interact.

Acknowledgments

This work was supported in part by Australian Research Council Discovery Project Grant DP240101264.

References

Bainbridge, W. A. (2019). Memorability: How what we see influences what we remember. In *Psychology of Learning and Motivation—Advances in Research and Theory* (Vol. 70, pp. 1–27). Academic Press Inc. https://doi.org/10.1016/bs.plm.2019.02.001

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1210.

Bylinskii, Z., Goetschalckx, L., Newman, A., & Oliva, A. (2022). Memorability: An image-computable measure of information utility. *Human Perception of Visual Information: Psychological and Computational Perspectives*, 207-239.

Criss, A.H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*, 316-326

Dubey, R., Peterson, J., Khosla, A., Yang, M. H., & Ghanem, B. (2015). What makes an object memorable?. In *Proceedings of the ieee international conference on computer vision* (pp. 1089-1097).

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological review*, *91*(1), 1.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological review*, *95*(4), 528.

Fox, J., & Osth, A. F. (2023). Modeling the continuous recognition paradigm to determine how retrieval can impact subsequent retrievals. *Cognitive Psychology*, 147, 101605.

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. In *Proceedings of the IEEE international conference on computer vision* (pp. 2390-2398). IEEE, Santiago, Chile. https://doi.org/10.1109/ICCV.2015.275

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of experimental Psychology: General*, *139*(3), 558.

Kramer, M. A., Hebart, M. N., Baker, C. I., & Bainbridge, W. A. (2023). The features underlying the memorability of objects. *Science advances*, *9*(17), eadd2981.

Meagher, B. J., & Nosofsky, R. M. (2023). Testing formal cognitive models of classification and old-new recognition in a real-world high-dimensional category domain. *Cognitive Psychology,* 145, 101596.

Needell, C. D., & Bainbridge, W. A. (2022). Embracing new techniques in deep learning for estimating image memorability. *Computational Brain and Behavior*, 1–17.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 3–27.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280–315.

Nosofsky, R. M., & Meagher, B. J. (2022). Retention of exemplar-specific information in learning of real-world high-dimensional categories: Evidence from modeling of old-new item recognition. *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, *50*(2), 530–556. https://doi.org/10.3758/s13428-017-0884-8.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2020). Search for the missing dimensions: Building a feature-space representation for a natural-science category domain. *Computational Brain and Behavior*, *3*, 13–33.

Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(6), 1194–1209. https://doi.org/10.1037/0278-7393.29.6.1194

Osth, A. F., & Dennis, S. (in press). Global matching models of recognition memory. *The Oxford Handbook of Human Memory*.

Osth, A. F., Jansson, A., Dennis, S., & Heathcote, A. (2018). Modeling the dynamics of recognition memory testing with an integrated model of retrieval and decision making. *Cognitive Psychology*, 104, 106-142.

Robinson, K. J., & Roediger III, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*(3), 231-237.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317. https://doi.org/10.1126/science.3629243

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 267.

Tversky, A. (1977). Features of similarity. *Psychological review*, *84*(4), 327.

Valentine, T., & Ferrara, A. (1991). Typicality in categorization, recognition and identification: Evidence from face recognition. *British Journal of Psychology*, *82*(1), 87-102.