AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2024 American Psychological Association
ISSN: 0096-3445

2024, Vol. 153, No. 4, 1131–1138
https://doi.org/10.1037/xge0001543

BRIEF REPORT

Deep Neural Network Decodes Aspects of Stimulus-Intrinsic Memorability Inaccessible to Humans

Chong Zhao[1, 2], Joie Kim[3], Tzu Hsuan Tang[3], Joseph M. Saito[4], and Keisuke Fukuda[3, 4]
[1] Department of Psychology, University of Chicago
[2] Institute for Mind and Biology, University of Chicago
[3] Department of Psychology, University of Toronto Mississauga
[4] Department of Psychology, University of Toronto

Some stimuli are more memorable than others. Humans have demonstrated partial access to the properties that make a given stimulus more or less memorable. Recently, a deep neural network named ResMem was shown to successfully decode the memorability of visual stimuli as well. However, it remains unknown whether ResMem's predictions of memorability reflect the influence of stimulus-intrinsic properties or other stimulus-extrinsic factors that are known to induce interindividual consistency in memory performance (e.g., interstimulus similarity). Additionally, it is not clear whether ResMem and humans share access to overlapping properties of memorability. Here, in three experiments, we show that ResMem predicts stimulus-intrinsic memorability independent of stimulus-extrinsic factors, and that it captures aspects of memorability that are inaccessible to human observers. Taken together, our results confirm the multifaceted nature of memorability and establish a method for isolating its aspects that are largely inaccessible to humans.

Public Significance Statement

Some images are easier to remember than others, making some images more memorable than others consistently across human observers. Previous research has shown that both humans and a pretrained neural network called ResMem can predict image memorability. However, whether humans and ResMem rely on the same aspects of the images in predicting their memorability remains unclear. Our study first demonstrated that ResMem predicted the memorability of images without relying on their interitem similarity among other images. More crucially, we found that humans and ResMem used largely nonoverlapping aspects of images for memorability prediction, suggesting that ResMem can be used to elucidate the aspects of memorability that are not explicitly accessible to humans.

Keywords: visual memory, deep neural network, memorability

Supplemental materials: https://doi.org/10.1037/xge0001543.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Despite our massive storage capacity in visual long-term memory (VLTM, Brady et al., 2008; Standing, 1973), not all information that we encounter is successfully encoded. Interestingly, variability in memory encoding success for visual information is surprisingly consistent across individuals; some information is consistently remembered across observers while other information is consistently forgotten (Bainbridge et al., 2013; Isola et al., 2014). This interindividual consistency in memory performance has been observed across a variety of visual stimuli (Bainbridge et al., 2013; Khosla et al., 2015; Mancas & Le Meur, 2013) and has been interpreted to reflect the existence of a stimulus-intrinsic factor that determines the likelihood of memory encoding success, namely, memorability.

This article was published Online First February 1, 2024.
Chong Zhao https://orcid.org/0000-0002-2356-5528
This research was supported by the Natural Sciences and Engineering Research Council (Grant 5009170). The data, stimuli, and experimental codes are posted publicly at the Open Science Framework (https://osf.io/x3258/). This study was not preregistered. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Chong Zhao served as lead for conceptualization, formal analysis, methodology, software, visualization, writing–original draft, and writing–review and editing and contributed equally to project administration. Joie Kim served in a supporting role for writing–review and editing. Tzu Hsuan Tang served in a supporting role for writing–review and editing. Joseph M. Saito served as lead for data curation and contributed equally to conceptualization, formal analysis, and writing–review and editing. Keisuke Fukuda served as lead for conceptualization, funding acquisition, methodology, project administration, supervision, validation, and writing–review and editing. Joie Kim and Tzu Hsuan Tang contributed equally to data curation.

Correspondence concerning this article should be addressed to Chong Zhao, Institute for Mind and Biology, University of Chicago, 940 East 57th Street, Chicago, IL 60637, United States. Email: chongzhao@uchicago.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

To provide further support for the existence of stimulus-intrinsic nature of memorability, a recent study demonstrated that a pretrained deep neural network named ResMem was able to predict interindividual consistency in memory encoding success purely based on perceptual features of a given image (Needell & Bainbridge, 2022). ResMem was constructed by appending extra learning layers to an existing neural network (i.e., ImageNet) trained to classify the category of a given image (e.g., balloon, strawberry, and bread) from its perceptual features (Krizhevsky et al., 2012).

While ResMem's ability to predict stimulus-specific memory performance solely based on perceptual features of a given image is consistent with the notion that memory encoding success is in part determined by stimulus-intrinsic properties, these properties are not the only factors that give rise to interindividual consistency in memory performance. Past studies have demonstrated that stimulus-extrinsic factors can also induce interindividual consistency in memory performance (Eysenck, 1979; Hunt & Worthen, 2006; Koch et al., 2020; Schmidt, 1985). For example, a stimulus (e.g., a picture of a cat) can be remembered better across individuals if it is presented with other stimuli that are homogeneously distinct (e.g., pictures of dogs) than those that are similar (e.g., pictures of other cats, von Restorff, 1933). Similarly, individuals can consistently and falsely "remember" seeing an unseen stimulus after encoding other stimuli that are similar (e.g., Roediger & McDermott, 1995). Indeed, Bylinskii et al. (2015) demonstrated the impact of stimulus-extrinsic factors on interindividual consistency in memory performance. More specifically, by manipulating the algorithmically derived interstimulus similarity among stimuli, they found that the interstimulus similarity heavily influenced memory performance such that stimuli that were more distinct from other stimuli were consistently better remembered across individuals than less distinct stimuli. ResMem, the deep neural network model for predicting visual memorability, was trained and tested on recognition memory performance collected from participants. Naturally, we may wonder that stimulus-extrinsic effects, such as stimulus similarity, that affect these ground truth memorability values may, therefore, bias weights of the artificial neurons in ResMem. Given that such stimulus-extrinsic effect can emerge in any stimulus set, it is possible that ResMem relied on extra-stimulus information uniquely available in the stimulus set used during its training, rather than the stimulus-intrinsic memorability of each image, to predict interindividual consistency in memory performance.

To provide direct evidence that memorability is in part stimulus-intrinsic, we had participants encode a set of pictures of real objects that were novel to ResMem while introducing a stimulus-extrinsic factor known to induce interindividual memory consistency, namely interstimulus similarity. In Experiment 1, we examined the influence of interstimulus similarity on interindividual consistency in memory performance and on ResMem's predictions about that consistency. Participants encoded 192 pictures of real-world objects comprising eight exemplars from 24 different object categories whose within-category perceived similarity has been empirically validated in a previous study (Hout et al., 2014). After encoding, participants completed a recognition test that contained all of the encoded stimuli and eight new exemplars from each of the 24 object categories. If there is memorability that is context-independent and stimulus-intrinsic and ResMem has access to it, then ResMem should be able to predict the interindividual consistency in memory performance for the novel stimuli even after controlling for the interindividual consistency induced by interstimulus similarity.

In Experiments 2A and 2B, we compared access to stimulus-intrinsic memorability between ResMem and human observers. In a recent study, Saito and colleagues demonstrated that human observers could predict the memorability of a given stimulus (Saito et al., 2023). This suggests that humans, like ResMem, can predict interindividual consistency in memory performance. However, both ResMem and humans are not perfect at predicting the intrinsic memorability of a stimulus. Given that memorability is composed of multiple factors spanning across perceptual and semantic features of a stimulus (Isola et al., 2014; Kramer et al., 2023; Rust & Mehrpour, 2020), it is possible that ResMem and humans utilized dissociable factors to make memorability predictions. To test this possibility, we examined whether ResMem could predict the subjective memorability judgments that were made by humans in Saito et al. (2023). If ResMem and humans rely on overlapping aspects of a stimulus to make their predictions about its memorability, ResMem should be able to predict humans' memorability judgments. If, on the other hand, they rely on unique aspects, ResMem should not be able to predict humans' memorability judgments even if it reliably predicts memorability. To further examine if our conclusions generalize to visual stimuli other than objects, we conducted Experiments 3A and 3B using a widely used scene database (SUN database, Xiao et al., 2010). Previous studies have shown that ResMem robustly predicted recognition memory performance with scenes (Wakeland-Hart et al., 2022). Here, we examined whether humans could predict the memorability of scenes, and if so, whether the variance explained by humans overlapped with those by ResMem.

Method

Transparency and Openness

This study was not preregistered before data analysis. We report how we determined all data exclusions, all manipulations, and all measures in the study. Analyses were performed in MATLAB 2020a and Python 3.7, with the package matplotlib, scipy, numpy, and seaborn for plotting. Analysis scripts are publicly accessible at https://osf.io/x3258/. Data and materials used in this study are accessible to the public in the Open Science Framework repository (see the online supplement materials for reliability of our measurements and model outputs).

Participants

For Experiment 1, 96 undergraduate students (62 female, 32 male, and two chose not to respond; 17–27 years old, $M_{\text{age}} = 19.39$, $SD = 2.43$) at the University of Toronto Mississauga participated in the experiment in fulfillment of a course requirement for an introductory psychology course. All participants reported normal or corrected-to-normal color vision.

Experiments 2A and 2B reanalyzed the data collected in Experiments 1A and 2A of Saito et al. (2023), respectively. For each experiment, 120 young residents of the United States and Canada (Experiment 2A: 18–31 years old, $M_{\text{age}} = 24.19$, $SD = 3.97$, $n_{\text{female}} = 61$; Experiment 2A: 18–30 years old, $M_{\text{age}} = 22.68$, $SD = 3.13$, $n_{\text{female}} = 91$) were recruited through Prolific and received monetary compensation (7.50 £/hr). All participants reported fluency in English, normal or corrected-to-normal vision, no

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

color blindness, no history of head injury, no history of mental illness/condition, and no cognitive impairment/Dementia. All participants had successfully completed 90% or more of the studies that they had participated in previously on Prolific (i.e., approval rate > 90%).

For Experiments 3A and 3B, 120 young residents of the United States and Canada were recruited through Prolific and received monetary compensation (7.50 £/hr). All participants satisfied the same criteria used for Experiments 2A and 2B.

Stimulus Selection and Assignment

For Experiment 1, we used an existing picture database (Hout et al., 2014). The database contained pictures of 16–17 exemplar objects from 24 object categories. Crucially, the multidimensional similarity rating for each exemplar within a corresponding category has been documented in the previous study (Hout et al., 2014). From this data set, we selected 24 object categories and sampled 16 exemplars from each category (384 objects total). To ensure that each stimulus served as old (i.e., presented during the encoding task) and new (i.e., not presented during the encoding task) for an equal number of participants, we first created 12 sets of old/new stimulus sets. For each set, a random half of the exemplars (i.e., eight exemplars) in each object category was selected as old stimuli, and the remaining half was assigned as new stimuli. An additional 12 sets were created by simply reversing the novelty status of the stimuli in the existing 12 sets. For example, the 13th set was created by swapping the old and new stimuli in the first set. As a result, we created 24 total sets of old and new stimuli, and each set was used to collect data from four participants.

For Experiments 2A and 2B, we selected a set of 600 images of random real-world objects from an existing database (Brady et al., 2008). The stimulus set was then randomly split into four smaller sets of 150 images so that one of the sets was used as old pictures and one of the remaining sets was used as new pictures. Each set was assigned as old and new pictures for 30 participants.

For Experiments 3A and 3B, the same procedure was followed using a set of 600 images of random real-world scenes from the SUN database (Xiao et al., 2010).

Apparatus for Online Testing

Due to the pandemic, experiments were conducted remotely using Inquisit 5 (2016). For Experiment 1, participants were given a private Zoom link and password to attend a live experiment with an experimenter. Given the remote nature of the experiment, screen resolution and size were not fixed. Each stimulus was presented to fit inside of an imaginary square whose side length was set to 20% of the height of the participants' computer monitor.

For the remaining experiments, we used the same setup as Experiment 1 except that (a) each stimulus was set to 12% of the height of the participants' computer monitor and (b) the participants were not monitored through Zoom during participation.

Procedure

All participants electronically signed the informed consent to the protocol approved by the Research Ethics Board of the University of Toronto. In Experiment 1, participants first completed the encoding task (Figure 1A). In this task, participants saw one picture at a time presented for 1,000 ms at the center of the screen, and they were asked to memorize the pictures as vividly as possible because their memory would be tested in the next phase. Each picture was presented once, and participants encoded 192 pictures (eight exemplars × 24 categories) in total.

After completing the picture encoding task, participants' memories were tested using a recognition memory task (Figure 1B). In the recognition memory task, participants were presented with one picture at a time at the center of the screen along with a 6-point Likert scale, and they judged whether they had seen the picture during the encoding task by clicking one of the six response buttons (1 = *definitely no*, 2 = *probably no*, 3 = *maybe no*, 4 = *maybe yes*, 5 = *probably yes*, 6 = *definitely yes*). The picture and the Likert scale remained on the computer screen until a response was made. Participants completed two blocks of 192 trials in which they were presented with 192 old (i.e., presented during the encoding task) and 192 new (i.e., not presented during the encoding task) pictures in a pseudo-random order.

In Experiments 2A and 3A, participants encoded 150 images of random objects (Experiment 2A) or scenes (Experiment 3A) while rating the perceived memorability (PM) of each image based on perceived encoding success (PES) (Figure 1C, PES rating). Each image was presented at the center of the screen for 1,000 ms. Five hundred milliseconds after the offset of the image, participants answered the question: "Are you going to remember the picture you just saw?" by choosing one of the six responses used in Experiment 1. The presentation order of the stimuli was randomized across participants. After the encoding task, participants performed the recognition task. The recognition task was identical to that in Experiment 1 except that participants saw 150 old and 150 new images in a random order.

In Experiments 2B and 3B, participants rated the PM of each image directly (Figure 1C, PM estimates). Our main goal here was to demonstrate humans' explicit access to memorability without the explicit memory encoding demand. The task was identical to the encoding task in Experiment 2A except that participants were not told to encode each image and they judged the memorability of each image by answering a question: "Would an average person remember the picture you just saw?" by choosing one of the six responses used in Experiment 2A. The experiment ended after the memorability judgment task without a recognition memory task.

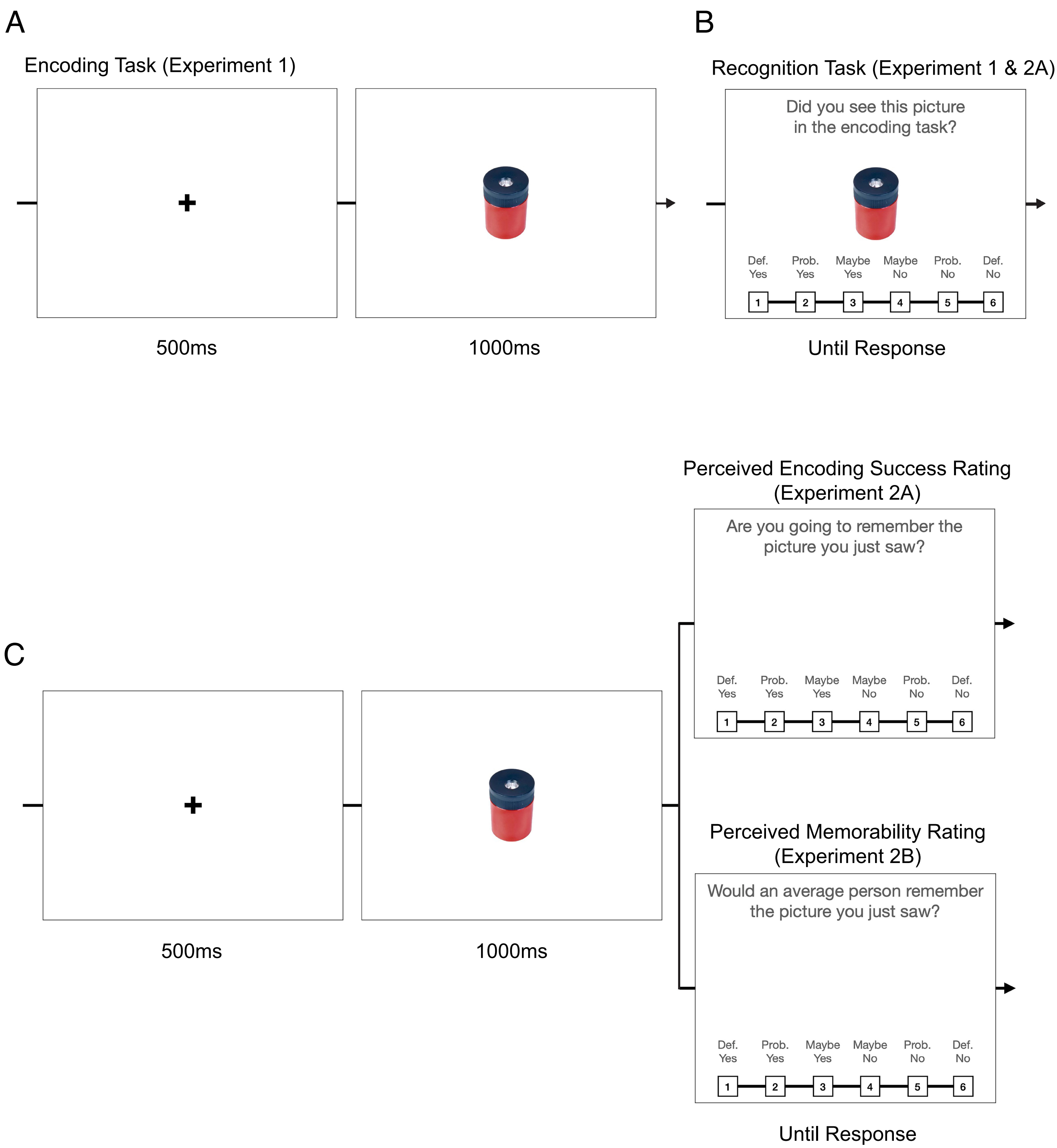
Data Analysis

Quantifying Memorability Scores

Memorability scores for each stimulus were computed by subtracting the average recognition response when the stimulus was new (average new response) from the average recognition response when it was old (average old response). Here, to increase the intuitiveness of the scores, we reverse-coded participants' responses such that higher values meant higher confidence in remembering (e.g., 6 = *definitely yes*) and lower values meant lower confidence in remembering (e.g., 1 = *definitely no*). Thus, more positive values indicate higher memorability (easier to remember), and less positive values indicate less memorability (harder to remember).

To measure humans' ability to predict memorability, PES (Experiments 2A and 3A) and PM (Experiments 2B and 3B) for each stimulus were computed by averaging the corresponding scores across participants. Similarly, to increase the intuitiveness of the scores, we reverse-coded participants' responses such that

Figure 1
Experimental Procedures



Note. Panel A depicts the encoding task in Experiment 1. Panel B depicts the recognition task in Experiments 1 and 2A. Panel C shows the memorability judgment task in Experiments 2A/3A (PES rating) and in Experiments 2B/3B (PM rating). PES = perceived encoding success; PM = perceived memorability; Def. = definitely; Prob. = probably. See the online article for the color version of this figure.

higher values meant higher encoding success/memorability (e.g., 6 = *definitely yes*) and lower values meant lower encoding success/memorability (e.g., 1 = *definitely no*).

To measure ResMem's ability to predict memorability, we used the ResMem network (Needell & Bainbridge, 2022) with no retraining features implemented.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Quantifying the Interstimulus Similarity Within Category (Similarity Score)

The similarity of exemplars within an object category was based on a two-dimensional similarity space quantified by Hout et al. (2014). The *x* and *y* coordinates for each exemplar within this similarity space were used to calculate interstimulus distances for each stimulus within a given object category. The interstimulus distance for each stimulus was quantified by computing its average Euclidean distance from all the other exemplars within the category. Low interstimulus distance indicates a high similarity of a given exemplar to the other exemplars in the category, whereas high interstimulus distance indicates a low similarity of a given exemplar to the other exemplars.

Results

Experiment 1

Interstimulus Similarity Predicts Memorability Scores

To confirm that interstimulus similarity can drive interindividual consistency in memory performance, we first examined the correlation between interstimulus distances and memorability scores. As predicted, we found a significant negative correlation between the two variables, $r(346) = -.29$, $p < .001$, Figure 2A, such that the more dissimilar a given item was to other items in the category, the more memorable it was. This is consistent with von Restorff effect (von Restorff, 1933), and, more importantly, confirms that interindividual consistency in memory performance can be induced by stimulus-extrinsic factors (Bylinskii et al., 2015). Therefore, interindividual consistency in memory performance alone cannot guarantee the existence of stimulus-intrinsic memorability.

ResMem Predicts Memorability Scores Independent of Interstimulus Similarity

Next, we examined whether ResMem would be able to predict observed memorability scores of this novel stimulus set. As can be seen in Figure 2B, despite a substantial difference between the training stimuli (i.e., pictures of objects and scenes with a diverse and rich background) and current stimuli (i.e., pictures of objects with a white background), ResMem's predictions reliably correlated with memorability scores, $r(346) = -.16$, $p = .002$, Figure 2B. To determine whether this relationship was attributable to ResMem's sensitivity to interstimulus distance, we regressed the effect of interstimulus distance out of observed memorability scores and remeasured the correlation. Upon doing so, we found that the relationship persisted, $r(346) = -.14$, $p = .005$, Figure 2C. Thus, we concluded that ResMem's ability to predict interindividual consistency in memory performance was not based on the within-category interitem similarity but likely reflected its sensitivity to stimulus-intrinsic properties that made a stimulus memorable or forgettable.

Experiments 2 and 3

ResMem and Humans Have Reliable, But Imperfect Access to Memorability

Similar to Experiment 1, we found in Experiments 2A and 2B that ResMem successfully predicted memorability scores for the novel object stimulus set, $r(598) = .11$, $p = .009$, Figure 3A. The prediction power of ResMem resembled the five data collections we found using the common object set (Zhao et al., 2022). As demonstrated in Saito et al. (2023), humans also predicted the memorability of stimuli when they explicitly encoded the stimuli, PES in Experiment 2A, $r(598) = .71$, $p < .001$, Figure 3B, and when they did not, PM in Experiment 2B, $r(598) = .60$, $p < .001$, Figure 3C.

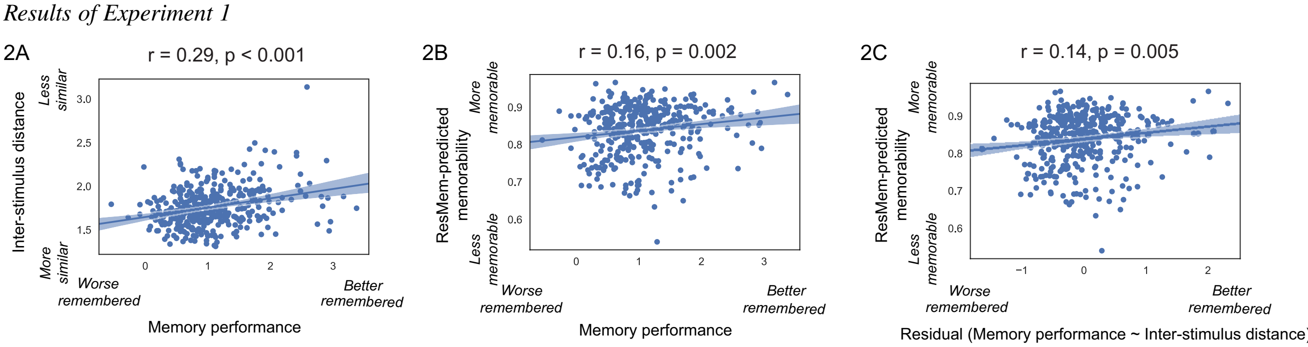
Experiments 3A and 3B replicated these findings with scene images. ResMem successfully predicted memorability scores for scene images, $r(598) = .43$, $p < .001$, Figure 4A. Additionally, we generalized the findings of Saito et al. (2023) to scene images. In other words, humans reliably predicted memorability scores when they explicitly encoded the stimuli, $r(598) = .55$, $p < .001$, Figure 4B, and when they did not, $r(598) = .41$, $p < .001$, Figure 4C. These results revealed that both ResMem and humans had reliable but imperfect access to memorability.

ResMem Predicts Aspects of Memorability Inaccessible to Humans

To test whether ResMem and human observers utilized the same aspects of memorability for their respective predictions, we tested whether ResMem could predict humans' PM ratings. As can be seen in Figure 3, ResMem did not predict the PM ratings made by humans during explicit encoding, i.e., PES ratings, PES, $r(598) = .03$, $p = .42$,

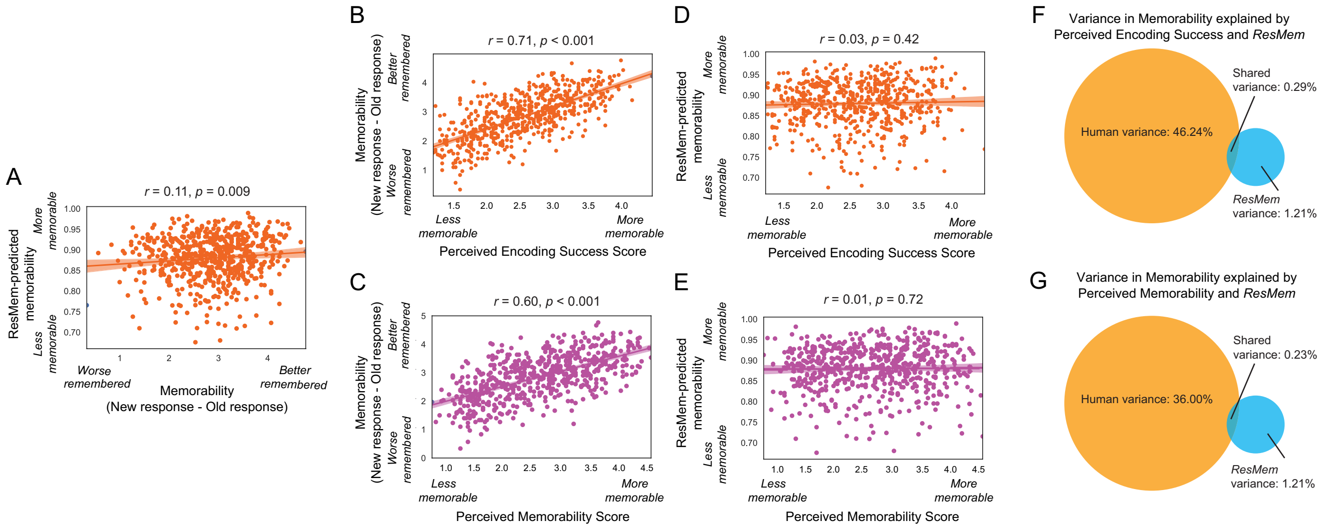
Figure 2

Results of Experiment 1



Note. Panel A shows the correlation between memorability scores and interstimulus distances. Panel B shows the correlation between observed memorability scores and the memorability scores that were predicted by ResMem. Panel C shows the correlation between observed memorability score residuals after regressing out interstimulus distance and the memorability scores that were predicted by ResMem. See the online article for the color version of this figure.

Figure 3
Results of Experiments 2A and 2B



Note. Panel A shows the correlation between observed memorability scores and ResMem predictions for object images. Panels B and C show the correlation between observed memorability scores and the PES ratings (Experiment 2A) and the PM ratings (Experiment 2B), respectively, that were made by human observers. Panels D and E show the correlation between ResMem predictions and PES (Experiment 2A) and PM (Experiment 2B), respectively. Panels F and G show Venn's diagrams for the variance in memorability explained by humans' predictions (PES in Experiment 2A and PM in Experiment 2B). The shared variance between humans' and ResMem's predictions was much smaller than the unique variance explained by each factor. PES = perceived encoding success; PM = perceived memorability. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

see Figure 3D, nor those made in the absence of explicit encoding, i.e., PM ratings, PM, $r(598) = .01$, $p = .72$, Figure 3E. When we compared ResMem's prediction strength for memorability and humans' memorability predictions, we found that ResMem performed significantly better in predicting the actual memorability than predicting its human predictions (Fisher's $Z = 2.577$, $p = .010$, for PES in Experiment 2A; Fisher's $Z = 2.743$, $p = .006$ for PM in Experiment 2B, cf. Meng et al., 1992 for the method of comparing the strength of correlated correlations). These findings suggest that ResMem and human observers accessed distinct properties of a given stimulus in order to successfully predict its memorability (Figure 3F and 3G actual memorability ∼ PES + ResMem in Experiment 2A, and actual memorability ∼ PM + ResMem in Experiment 2B).

While ResMem demonstrated more robust predictive power for scene memorability, we replicated that ResMem and humans accessed separable aspects of scene memorability. Although ResMem did predict the humans' memorability ratings made during explicit encoding, PES in Experiment 3A, $r(598) = .21$, $p < .001$, Figure 4D, and those made in its absence, PM in Experiment 3B, $r(598) = .08$, $p = .045$, Figure 4E, ResMem's prediction strength for the actual memorability was significantly better, PES in Experiment 3A: Fisher's $Z = 6.134$, $p < .001$; PM in Experiment 3B: Fisher's $Z = 8.44$, $p < .001$. These findings replicate that ResMem and human observers accessed distinct properties of scene images to successfully predict its memorability (Figure 4F and 4G, actual memorability ∼ PES + ResMem in Experiment 3A, and actual memorability ∼ PM + ResMem in Experiment 3B).

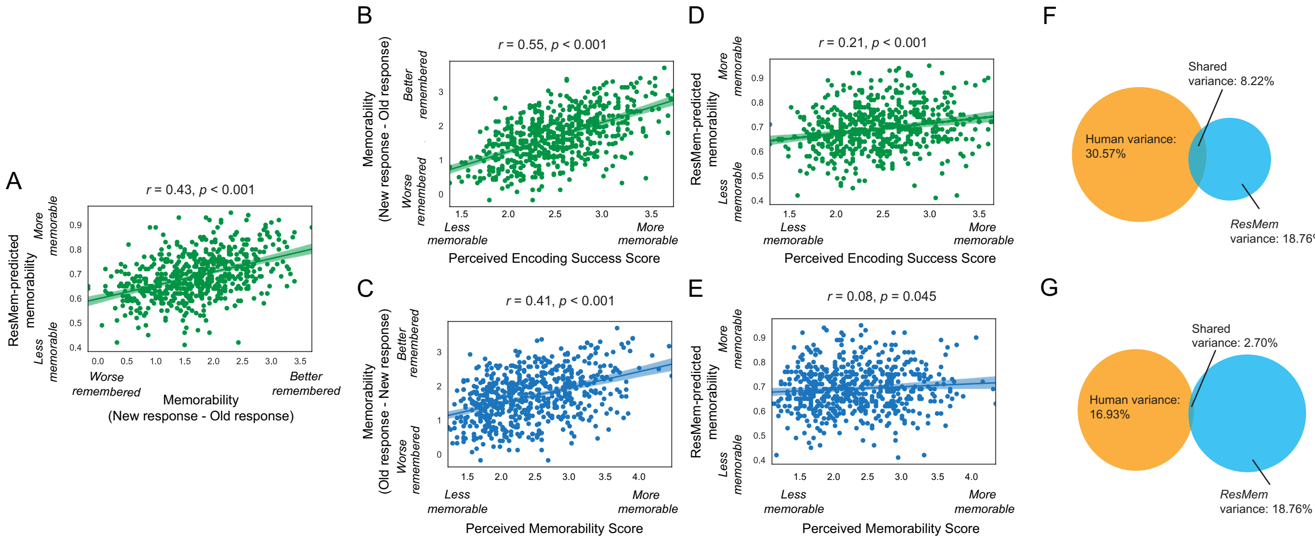
Discussion

Memory performance for a given visual stimulus is surprisingly consistent across individuals, suggesting the existence of memorability. A recent demonstration that a pretrained deep neural network called ResMem can predict the memorability based on the perceptual properties of a given images is consistent with this hypothesis. However, whether ResMem utilizes stimulus-extrinsic properties that emerge due to contextual relationship among a given set of stimuli (e.g., interstimulus similarity) used in its training or stimulus-intrinsic properties possessed by each item independent of other stimuli in the set has been unclear. To test this in Experiment 1, we had participants remember a novel stimulus set whose within-category interstimulus similarity was experimentally manipulated and had ResMem predict the memorability for each stimulus. Here, despite the novelty of the stimulus set to ResMem, it predicted the memorability independently of the interstimulus similarity. This suggests that ResMem captures interindividual consistency in memory performance that likely stem from stimulus-intrinsic properties of each stimulus that are immune to within-categorical, stimulus-extrinsic factors.

One crucial aspect of our finding was that more perceptually similar items were associated with lower memorability, though ResMem did not utilize this aspect in predicting memorability. This result might be explained by a high-dimensional account of memorability. According to this account, memorability is reflected in the magnitude of neural activations, whereas object identities are reflected in the direction of neuronal activations in the higher visual cortex (Rust & Mehrpour, 2020). If ResMem approximates the magnitudes of neural activations in the higher visual cortex, whereas the interstimulus similarity is characterized by the similarity in the directions of neural activations, it makes sense that ResMem's prediction is orthogonal to interstimulus similarity.

Our result, on the other hand, might seem contradictory to previous findings that associated higher similarities in neural activation patterns with memorable stimuli than with forgettable stimuli

Figure 4
Results of Experiments 3A and 3B



Note. Panel A shows the correlation between observed memorability scores and ResMem predictions for scene images. Panels B and C show the correlation between observed memorability scores and the PES ratings (Experiment 3A) and the PM ratings (Experiment 3B), respectively, that were made by humans. Panels D and E show the correlation between ResMem predictions and PES (Experiment 3A) and PM (Experiment 3B), respectively. Panels F and G show Venn's diagrams for the variance in memorability explained by human predictions (PES in Experiment 3A and PM in Experiment 3B). The shared variance between human's and ResMem's predictions was much smaller than the unique variance explained by each factor. PES = perceived encoding success; PM = perceived memorability. See the online article for the color version of this figure.

(Bainbridge & Rissman, 2018; Bainbridge et al., 2017). One possibility for this seeming discrepancy may be that the previous findings focused on within-category (i.e., faces) variabilities in memorability whereas our study examined cross-category (i.e., objects and scenes) variabilities in memorability. Future studies are necessary to elucidate the effect of category granularity on memorability.

Of equal importance, researchers have questioned whether and to what extent humans have explicit access to memorability (Isola et al., 2014; Saito et al., 2023). To answer this, Experiments 2A and 2B first showed that both ResMem and humans reliably predicted memorability. More critically, ResMem predicted the actual memorability significantly better than it predicted humans' predictions of memorability. We replicated these findings with scene images in Experiments 3A and 3B, despite much more robust memorability prediction by ResMem. These results reveal the utility of ResMem in isolating aspects of memorability that are not explicitly accessible to humans. Conversely, the observed insensitivity of ResMem to a stimulus-extrinsic contributor to memorability (i.e., within-category interitem similarity) might imply that humans uniquely utilize such factor to estimate memorability. Future studies should examine this possibility by measuring PM by directly manipulating such stimulus-extrinsic factors.

Although our findings extend the existing account that not all aspects of memorability are explicitly accessible to human observers (Bainbridge et al., 2013, 2017; Isola et al., 2014; Saito et al., 2023), the present study does not inform us about what these aspects are or what underlying processes differentiate the accessible aspects from the inaccessible aspects. Future studies should explore this as the findings can inform us as to whether and how we can train human observers to be fully aware of stimulus-intrinsic memorability. Additionally, ResMem's inability to predict PM does not mean that a neural network with the same architecture cannot be trained to predict them. For instance, a recent paper has presented a theoretical account using artificial networks in explaining the emergence of metamemory (Yamato et al., 2022). Additional studies in training either human or artificial network in predicting memorability would be informative in determining how the divergence of PM from true memorability emerge through the hierarchy of visual information processing. Lastly, our demonstration of the dissociability for humans' and ResMem's memorability predictions was limited to static images of individuated objects and scenes. Given that a recent study demonstrated memorability for dynamic stimuli such as dance moves (Ongchoco et al., 2023), future studies should also explore whether and how humans and deep neural networks can predict memorability for such dynamic stimuli. Going beyond, future studies should also explore memorability in other stimulus domains, such as verbal stimuli (Aka et al., 2023; Kramer et al., 2023; Madan, 2021). With the recent development in large language models, more insights could be gained from comparing machine learning and human observers in predicting memorability using verbal and even syntactic materials.

Constraints of Generality

Our experiments used a large number of isolated objects (Experiments 1 and 2) as well as scenes (Experiment 3) as our stimuli. We replicated our key findings, that ResMem predicts stimulus-intrinsic memorability inaccessible to human observers, with both objects and scenes as the stimulus set. Therefore, we expect our key findings to generalize to a wide range of meaningful visual stimuli.

Our experiments contained both in-lab and online participants. Specifically, we recruited students from the University of Toronto

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

population in Experiment 1, and we recruited online participants from Prolific in Experiments 2 and 3. Therefore, we believe that our results are generalizable across samples of human populations both inside and outside of lab settings.

References

Aka, A., Bhatia, S., & McCoy, J. (2023). Semantic determinants of memorability. *Cognition*, *239*, Article 105497. https://doi.org/10.1016/j.cognition.2023.105497

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

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

Bainbridge, W. A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Scientific Reports*, *8*(1), Article 8679. https://doi.org/10.1038/s41598-018-26467-5

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*(Pt B), 165–178. https://doi.org/10.1016/j.visres.2015.03.005

Eysenck, M. W. (1979). Depth, elaboration, and distinctiveness. In L. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Lawrence Erlbaum Associates.

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the Massive Memory picture database. *PLoS ONE*, *9*(11), Article e112644. https://doi.org/10.1371/journal.pone.0112644

Hunt, R. R., & Worthen, J. B. (2006). *Distinctiveness and memory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169669.001.0001

Inquisit 5 [Computer software]. (2016). https://www.millisecond.com

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

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015, December 7–13). *Understanding and predicting image memorability at a large scale* [Paper presentation]. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile.

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

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. https://doi.org/10.1126/sciadv.add2981

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012, December 3–6). *ImageNet classification with deep convolutional neural networks* [Paper presentation]. NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, United States.

Madan, C. R. (2021). 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

Mancas, M., & Le Meur, O. (2013, September 15–18). *Memorability of natural scenes: The role of attention* [Paper presentation]. 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*(1), 172–175. https://doi.org/10.1037/0033-2909.111.1.172

Needell, C. D., & Bainbridge, W. A. (2022). Embracing new techniques in deep learning for estimating image memorability. *Computational Brain & Behavior*, *5*(2), 168–184. https://doi.org/10.1007/s42113-022-00126-5

Ongchoco, J. D. K., Chun, M. M., & Bainbridge, W. A. (2023). What moves us? The intrinsic memorability of dance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(6), 889–899. https://doi.org/10.1037/xlm0001168

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rust, N. C., & Mehrpour, V. (2020). Understanding image memorability. *Trends in Cognitive Sciences*, *24*(7), 557–568. https://doi.org/10.1016/j.tics.2020.04.001

Saito, J. M., Kolisnyk, M., & Fukuda, K. (2023). Judgments of learning reveal conscious access to stimulus memorability. *Psychonomic Bulletin & Review*, *30*(1), 317–330. https://doi.org/10.3758/s13423-022-02166-1

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 565–578. https://doi.org/10.1037/0278-7393.11.3.565

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. https://doi.org/10.1080/14640747308400340

von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [The effects of field formation in the trace field]. *Psychologische Forschung [Psychological Research]*, *18*(1), 299–342. https://doi.org/10.1007/BF02409636

Wakeland-Hart, C. D., Cao, S. A., deBettencourt, M. T., Bainbridge, W. A., & Rosenberg, M. D. (2022). Predicting visual memory across images and within individuals. *Cognition*, *227*, Article 105201. https://doi.org/10.1016/j.cognition.2022.105201

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010, June 13–18). *Sun database: Large-scale scene recognition from abbey to zoo*. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, United States (pp. 3485–3492). IEEE.

Yamato, Y., Suzuki, R., & Arita, T. (2022). Evolution of metamemory based on self-reference to own memory in artificial neural network with neuromodulation. *Scientific Reports*, *12*(1), Article 6233. https://doi.org/10.1038/s41598-022-10173-4

Zhao, C., Fukuda, K., Park, S., & Woodman, G. F. (2022). Even affective changes induced by the global health crisis are insufficient to perturb the hyper-stability of visual long-term memory. *Cognitive Research: Principles and Implications*, *7*(1), Article 62. https://doi.org/10.1186/s41235-022-00417-2

Received February 17, 2023
Revision received November 19, 2023
Accepted December 19, 2023 ■