



Memory for artwork is predictable

Trent M. Davis^a and Wilma A. Bainbridge^{a,1}

Edited by Henry Roediger III, Washington University in St. Louis, St. Louis, MO; received February 10, 2023; accepted May 17, 2023

Viewing art is often seen as a highly personal and subjective experience. However, are there universal factors that make a work of art memorable? We conducted three experiments, where we recorded online memory performance for 4,021 paintings from the Art Institute of Chicago, tested in-person memory after an unconstrained visit to the Art Institute, and obtained abstract attribute measures such as beauty and emotional valence for these pieces. Participants showed significant agreement in their memories both online and in-person, suggesting that pieces have an intrinsic “memorability” based solely on their visual properties that is predictive of memory in a naturalistic museum setting. Importantly, ResMem, a deep learning neural network designed to estimate image memorability, could significantly predict memory both online and in-person based on the images alone, and these predictions could not be explained by other low- or high-level attributes like color, content type, aesthetics, and emotion. A regression comprising ResMem and other stimulus factors could predict as much as half of the variance of in-person memory performance. Further, ResMem could predict the fame of a piece, despite having no cultural or historical knowledge. These results suggest that perceptual features of a painting play a major role in influencing its success, both in memory for a museum visit and in cultural memory over generations.

memorability | naturalistic memory | art

Why do we remember seeing some works of art but not others? We may be inclined to believe that this experience is reflective of style preference, personal judgments, or even how one emotionally connects to a piece. As you navigate an art museum and store this experience in memory, many other factors may influence what you will remember, such as your level of attention or fatigue, your social interactions with the people around you, your previous experiences with artwork, and how you navigate around the exhibit. However, one’s memory of a work of art may have less to do with the observer than is expected, and rather the properties of the artwork itself.

Artwork could potentially have an intrinsic *memorability*, which reflects the likelihood that a given piece will be remembered across viewers, based purely on its visual properties. (1) People are highly consistent in what they remember and forget for a wide variety of stimuli such as face images (2), natural scene photographs (3), words (4), and even dance moves (5), suggesting that there is an intrinsic property to the stimulus that generalizes across observers. Image memorability has been shown to be independent of other viewer-oriented features such as attentional state (6), motivation (7), and personal preference (8). Further, while the surrounding context of an image has clear influences on memory (9), there is a separable memorability intrinsic to an image itself that influences memory regardless of the surrounding images (10, 11). This is exciting, because it suggests that memorability can be calculated from individual images and used to make predictions about people’s memories across tasks. Indeed, memorability as measured in an explicit intentional memory task can successfully predict memory in an implicit incidental memory task (12, 13), and memorability effects persist across long delays [1 week (14), possibly even years (15)].

Artwork presents an interesting test case to examine the generalizability of people’s memory. Viewing artwork is generally understood as a subjective experience, and it is often the goal of an artist or museum curator to leave a lasting and powerful impact on the viewer’s memory. A visit to an art museum may be one of the few times outside of an experiment that we view a series of static images with the intention of remembering them. Importantly, if there are intrinsic perceptual aspects to an image that makes it memorable, then it may be possible to make honed predictions of human memory in the real world. Here, we determine the factors that influence memory for paintings in both an online experiment and an in-person museum visit. We show that we can make targeted predictions about which pieces people will remember, to the degree that we can predict the pieces that become culturally famous.

Significance

Despite art’s connotation as a subjective field, we find that there is a quantifiable intrinsic memorability to artwork by studying 4,021 paintings from the Art Institute of Chicago. After examining visual properties of paintings and influences of the gallery setting, we reveal a model that can predict naturalistic memory for artwork based solely on its perceptual features. Additionally, we found through using ResMem, a neural network that predicts an image’s memorability, that memorability can predict what works of art become famous.

Author affiliations: ^aDepartment of Psychology, University of Chicago, Chicago, IL 60637

Author contributions: T.M.D. and W.A.B. designed research; T.M.D. performed research; W.A.B. contributed new reagents/analytic tools; T.M.D. and W.A.B. analyzed data; and T.M.D. and W.A.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

¹To whom correspondence may be addressed. Email: wilma@uchicago.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2302389120/-/DCSupplemental>.

Published July 3, 2023.

Results

We conducted three experiments to design and test a model for predicting the art pieces that leave a lasting impact on one's memory. In *Experiment 1*, we collected memory performance measures on all usable 4,021 paintings from the Art Institute of Chicago's online database. 3,216 unique participants engaged in a continuous recognition task on online experiment platform Amazon Mechanical Turk, where they indicated their memory (old/new) for a random subset of 50 target paintings intermixed with foils from the same image set (Fig. 1A). With counterbalancing, each painting was seen as a target image by at least 40 participants. We calculated the "memorability score" of each piece as its corrected recognition (CR) score across those participants, measured by the hit rate (HR) minus the false alarm (FA) rate.

Our first key question was whether participants tended to remember and forget the same pieces, or if their memory was idiosyncratic. A split-half consistency analysis across 1,000 iterations showed that random participant halves had significant agreement in the paintings that they remembered and forgot (Spearman–Brown corrected $\rho = 0.53$, $P < 0.001$). This suggests that works of art do have an intrinsic memorability that spans across participants. There was a wide spread in people's memory performance for the pieces (Mean $CR = 0.64$, $SD = 0.11$, $min = 0.23$, $max = 0.95$); examples at each extreme of memorability are

shown in Fig. 2. We make all memory data publicly available, forming the largest dataset of art pieces with memory performance measures (<https://osf.io/vhp5d/>) (16).

We next assessed our ability to predict memory in this online experiment. We utilized ResMem, a deep learning neural network trained to predict the memorability of images based on a training set consisting primarily of real-world photographs (17). Prior work has shown that ResMem predictions significantly correlate with adult memory performance in a similar continuous recognition task for a wide range of photographs (Spearman $\rho = 0.67$) (17), but it has not been tested for predictions of memory for artwork, which vary highly in visual content and higher-level factors like meaning and cultural context. ResMem predictions for the 4,021 art pieces ($M = 0.73$, $SD = 0.12$, $min = 0.41$, $max = 0.98$) showed a significant correlation with online human memory performance (Spearman $\rho = 0.45$, $P < 0.0001$), demonstrating that ResMem can predict the art pieces people will remember in an online experiment.

More importantly, we sought to test how well we could predict memory for art in the real world. In *Experiment 2*, separate participants ($N = 19$) engaged in an in-person experiment in the American Art wing of the Art Institute. Participants were instructed to freely explore each floor of the wing at their own pace, look at all of the paintings, and then exit to the hallway where they took a mobile experiment on their phones (Fig. 1B). The mobile experiment showed

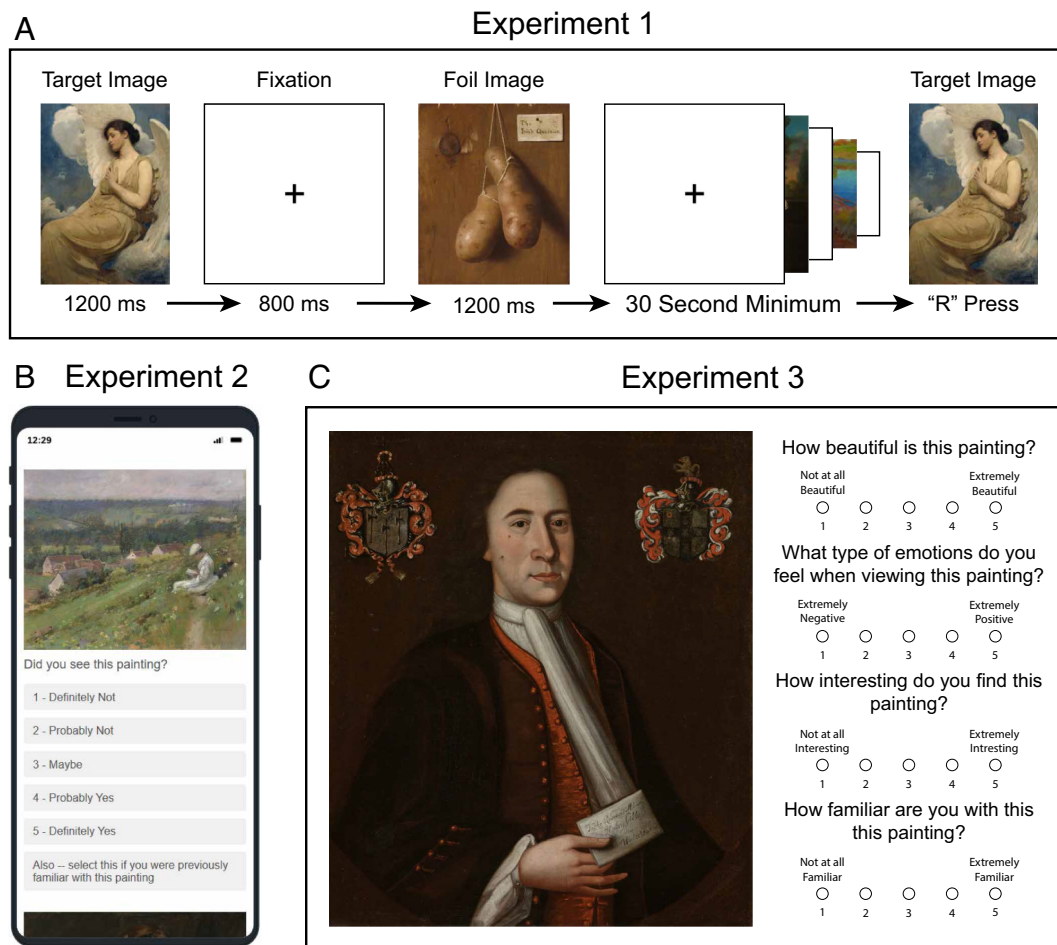


Fig. 1. Methods for Experiments 1 to 3. (A) In an online experiment (*Experiment 1*), 3,216 participants performed a continuous recognition memory task, where they viewed images and pressed a key when they recognized an image from earlier in the sequence. (B) In an in-person experiment (*Experiment 2*), 19 new participants freely explored the Art Institute of Chicago, and then afterwards rated their memory for a set of target and foil paintings on their mobile phone. (C) In an online experiment (*Experiment 3*), 40 new participants provided ratings of beauty, emotion, interest, and familiarity for each painting.

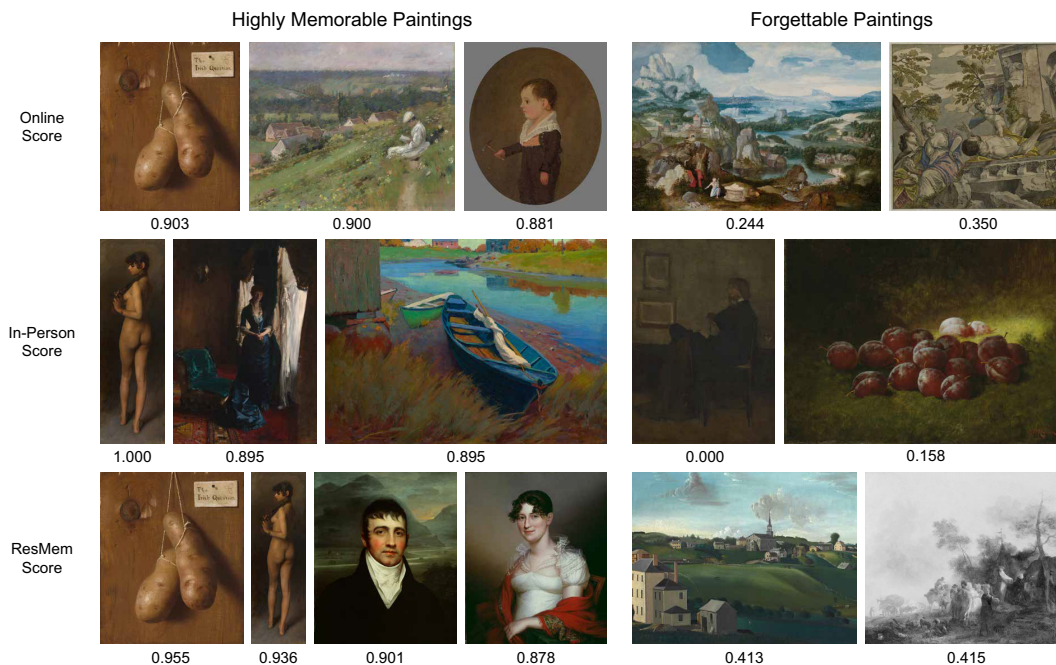


Fig. 2. Example pieces from the Art Institute of Chicago and their memorability scores. These examples show some of the most memorable and forgettable pieces, from *Experiment 1* (online memory experiment, *Top row*), *Experiment 2* (in-person museum memory experiment, *Middle row*), and neural network ResMem's predictions (*Bottom row*). The score under each image is the memorability score from that set (ranging from 0 as most forgettable up to 1 as most memorable).

participants all 162 paintings in the wing, and asked participants to indicate their memory for each one, intermixed with foil paintings matched for content, region, and time period.

In-person participants were significantly consistent across 1,000 random split-halves in the pieces they remembered and forgot (Spearman-Brown corrected $\rho = 0.67$, $P < 0.001$; descriptive statistics: $M = 0.61$, $SD = 0.19$, $min = 0$, $max = 1$). We next created a multiple regression model to test how well our neural network ResMem could predict in-person memory. By nature of these pieces existing alongside each other in the real world (Fig. 3), we also included predictors for real-world properties of these pieces that would be inaccessible to ResMem. Specifically, we included the physical size of each painting (area in cm^2), its floor (1st or 2nd), the average memorability of the other pieces in the same room, and the average size of the other pieces in the same room (see *Materials and Methods* for model specification). This model significantly predicted people's memory in the museum ($R^2_{\text{adjusted}} = 0.35$, $P < 0.0001$, Cohen's $f^2 = 0.61$, $AIC = 390.41$; see *SI Appendix, Table S1* for detailed statistics). Among its predictors, ResMem was significantly able to predict in-person memory for the pieces ($\beta = 0.18$, $t = 2.34$, $P = 0.020$), although the memorability of surrounding pieces showed no influence ($P = 0.091$). Larger paintings were also better remembered ($\beta = 0.45$, $t = 3.21$, $P = 0.002$), with a significant interaction ($\beta = -0.24$, $t = 4.89$, $P < 0.0001$) whereby smaller paintings gained a boost in memory when surrounded by larger pieces ($t = 2.41$, $P = 0.018$, Cohen's $d = 0.60$). Paintings on the second floor were also more memorable ($\beta = -0.44$, $t = 6.11$, $P < 0.0001$), possibly due to differences in the painting styles and content on each floor.

Next, we delved into visual and subjective properties that may influence memorability of a piece or account for ResMem's successful predictions. In *Experiment 3*, we asked online participants ($N = 40$) to rate all 162 in-person target paintings ($N = 40$) on a five-point scale based on their beauty, emotional valence, interest, and familiarity (Fig. 1C). These features were chosen because they would most intuitively seem to predict memory for art pieces (e.g., more

aesthetic pieces being more memorable). We also collected measures of low-level visual features for each piece, specifically testing color (mean and SD of hue, saturation, and luminance), spectral energy, and clutter. We also categorized each piece based on content type (abstract, indoor scene, outdoor scene, object-oriented, or portrait) and the presence of people.

No low-level visual properties showed a correlation with in-person memory performance, with the exception that less cluttered pieces were more memorable (*SI Appendix, Table S2*). Memorability did not vary by painting content or the presence of people (both $P > 0.05$, *SI Appendix, Table S2*). In-person memory also showed no correlation with ratings of beauty ($P = 0.792$), familiarity ($P = 0.375$), or emotion ($P = 0.304$, nor emotion intensity: See *SI Appendix*). However, memory performance was significantly correlated with how interesting people found the piece (Spearman $\rho = 0.28$, $P = 0.0003$). We included the subjective attributes into our multiple regression model, and observed an improvement in prediction, with a lower Akaike Information Criterion ($R^2_{\text{adjusted}} = 0.44$, $P < 0.0001$, Cohen's $f^2 = 0.93$, $AIC = 370.00$). All prior significant regressors remained as significant predictors of in-person memory (*SI Appendix, Table S3*, see *SI Appendix, Table S4* for unique variance of each factor), including ResMem-predicted memorability ($\beta = 0.35$, $t = 4.28$, $P < 0.0001$), painting size ($\beta = 0.39$, $t = 2.93$, $P = 0.004$), the interaction with the sizes of surrounding pieces ($\beta = -0.18$, $t = 3.92$, $P = 0.0001$), and floor ($\beta = -0.43$, $t = 5.59$, $P < 0.0001$). Beauty, familiarity, and emotion were nonsignificant predictors ($P > 0.15$), although interest was a significant predictor ($\beta = 0.46$, $t = 4.78$, $P < 0.0001$). A data-driven stepwise regression (*SI Appendix, Table S5*) including all properties measured in this study as possible predictors ($R^2_{\text{adjusted}} = 0.53$, $P < 0.0001$, Cohen's $f^2 = 1.27$, $AIC = 344.77$) also maintained a significant unique contribution of ResMem-predicted memorability ($\beta = 0.28$, $t = 4.14$, $P = 0.0001$). The high R^2 of these models suggests that half of the variance in memory can be manipulated through these factors external to the observer (e.g., image memorability, painting size, surrounding

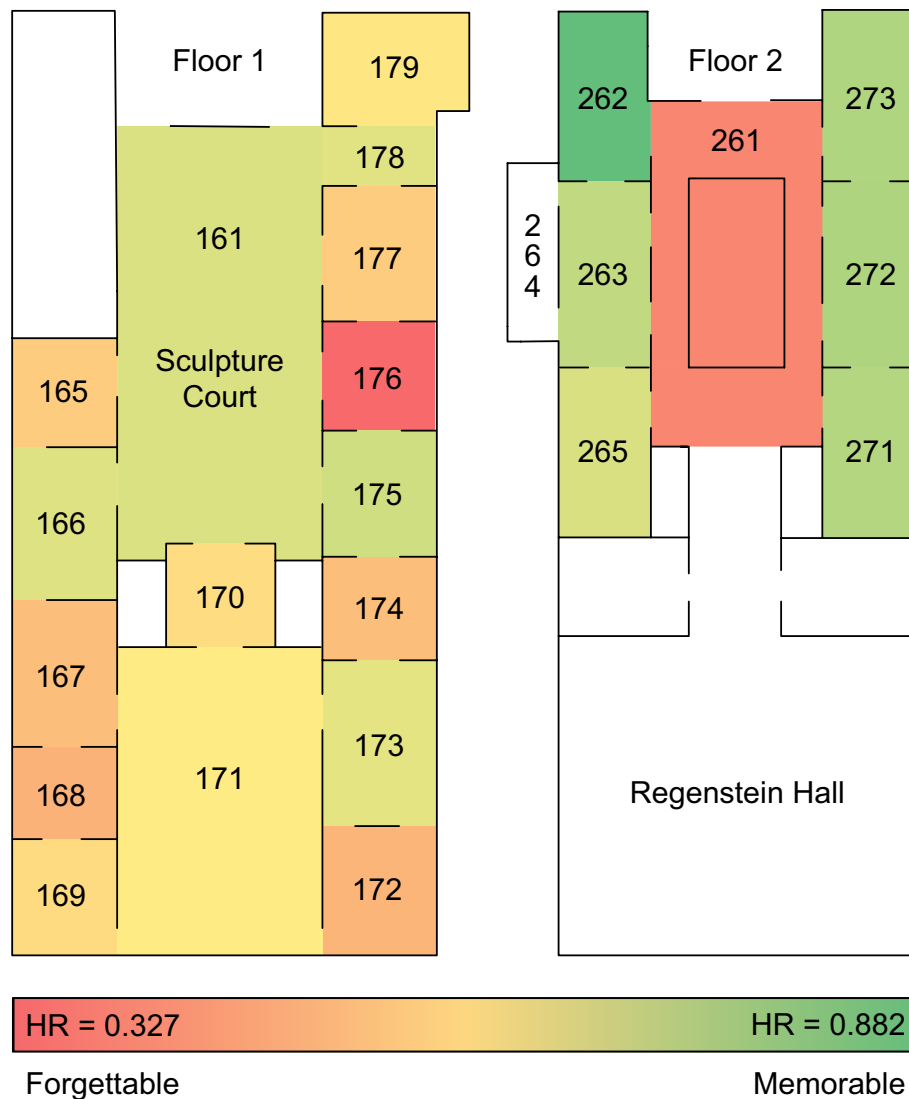


Fig. 3. Map of the American Art wing at the Art Institute of Chicago. Numbers indicate the gallery room number. Each room is colored by the average in-person memorability (hit rate, HR) of the paintings within that room, where green rooms are more memorable and red rooms are more forgettable.

pieces). These results also suggest that ResMem’s successful predictions of memory occur beyond any contribution of subjective properties like how interesting a piece is and cannot be explained by low-level features like color or general content of the piece. Also, surprisingly, features such as the beauty or emotion of a piece do not predict your ability to remember it.

We have shown that the neural network ResMem can make successful predictions of human memory, both in an online memory experiment (*Experiment 1*) and in an in-person visit to a museum (*Experiment 2*). Further, ResMem’s success cannot be fully accounted for by other low-level and subjective characteristics of a painting (*Experiment 3*).

The strongest test of the ability to predict human memory is to assess what pieces last in cultural memory across generations and society: the fame of a piece. 216 paintings in the Art Institute’s database are labeled as “boosted” by museum curators in the Art Institute online database, as pieces that are particularly famous and deserve extra publicity (such as *Self-Portrait* by Vincent Van Gogh, *Cow’s Skull with Calico Roses* by Georgia O’Keeffe, and *American Gothic* by Grant Wood). These pieces were intentionally excluded from all experiments reported here due to their extremely high familiarity. However, we collected ResMem memorability score

predictions for each of these paintings. We found that these famous paintings had significantly higher ResMem-predicted memorability than non-boosted paintings ($t = 2.17$, $P = 0.030$, Cohen’s $d = 0.16$), despite ResMem having no knowledge of art history, cultural significance, or cultural context. These results suggest that part of what causes a painting to be successful could be intrinsic visual properties of the piece that make it last in memory across people.

Discussion

Here, we have demonstrated the remarkable ability of a neural network, ResMem, to predict memory for art in a gallery, as well as the fame of a piece. This model can be used as a powerful tool to help design pieces and galleries that ensure the greatest impact on a viewer’s memory. We have also determined other factors that influence the memory for a piece in the real world: its size and the level of interest it conjures. A model combining all factors is able to predict as much as half of the variance in memory performance, suggesting that much of memory for images is external to the observer and can be manipulated. This has important implications for artists and curators, who can have an impressive amount of control over the memory of the viewer through changing these

factors. Surprisingly, many other factors such as beauty, emotion, or color appear to make little impact on a piece's ability to be remembered. These results highlight a surprising consistency in people's memory for art—while art appreciation feels like an entirely subjective experience, there are universals in the types of pieces that will last in our memories.

One important question we are left with is *what* is ResMem using to determine the intrinsic memorability of a piece? Several studies have shown that the distinctiveness of an image in relation to its surrounding images influences people's memories (9, 10, 18). However, here we isolate a property of memorability that is intrinsic to individual images and separable from image context—as shown by ResMem's ability to predict memorability without any contextual information. Indeed, recent work has shown that item memorability and experimental context are separable influences on memory, and that item memorability may relate more to the prototypicality of a stimulus than its distinctiveness (11). Future work can expand upon the current model by incorporating measures of set similarity or confusability in addition to image memorability to make even better honed predictions of memory. In terms of the features that make images memorable, prior work has shown that image memorability is more reliant on high-level semantic properties of an image, such as the function of what is depicted, rather than low-level visual properties like color, contrast, or edges (8, 11, 13). Given our limited memory capacity, memorability effects are thought to reflect how our brain prioritizes incoming visual information for what should be encoded, based on the visual statistics of the world (4, 12, 19). Indeed, recent work has shown that images that are memorable tend to be more quickly reinstated (4), may cause a more efficient processing state (9, 20), and may be the items that most successfully make it through the visual working memory bottleneck into long-term memory (21). The fact that less cluttered pieces are more memorable (*SI Appendix*) may support the idea that more easily processed items tend to be more memorable. However, based on prior work, it does not appear to be the case that memorable images just drive higher attention (6, 7, 22). Research is still ongoing to understand the precise factors that make an image memorable. In fact, while the average person has mixed insight into what is memorable that underperforms ResMem's predictions (8, 23, 24), it is possible that some artists may have particularly good intuitions about memorability that cause their pieces to be successful. Thus, further study with artists could provide fruitful discoveries on the factors that make a piece memorable. For example, one could create novel pieces designed to intentionally vary in memorability, with no chance of prior familiarity. In the meantime, this work has resounding implications for fields such as advertising, education, museum curation, art-making, and has already shown potential for improving diagnostic tests of early Alzheimer's disease (25). In sum, despite the highly personal nature of both art-making and viewing, our memories of art are surprisingly predictable and universal.

Materials and Methods

Experiment 1.

Participants. A total of 3,216 adults participated from the online experiment platform Amazon Mechanical Turk (AMT; 1,653 female, 1,383 male, 180 other; $M_{\text{age}} = 39.0$, $SD = 12.4$). In order to be recruited, participants had to be located in the United States, with at least a 95% approval rating on at least 50 prior tasks. Participants also had to correctly answer three basic English common-sense questions and provide at least one correct key press on the memory task in order to pass the task. The number of participants to recruit was determined as the number of participants required to get at least 40 ratings on each image in the database,

as 40 ratings has been shown in prior work to be sufficient to estimate memorability scores for a single image (3). Participants in all Experiments consented to participation and were compensated for their time. The full study protocol was approved by the University of Chicago Institutional Review Board (IRB19-1395). **Stimuli.** Stimuli consisted of 4,021 paintings in the online collection of the Art Institute of Chicago (<https://www.artic.edu/collection>), of which 725 are physically on display at the museum. These paintings span a diversity of media and content, composed of acrylic, oil, tempera, gouache, and watercolor on two-dimensional substrates such as canvas and wooden panels. We excluded pieces labeled in the online database that: 1) had low image quality (e.g., pixelated), 2) were fully text pieces (e.g., book pages), 3) included the picture frame in their online image, 4) were photographs of three-dimensional pieces, or 5) had an extreme aspect ratio (i.e., extremely wide or tall) that would prevent viewing the piece in a single page online.

For each stimulus image, we also saved information about that image from the Art Institute's online database. This database includes information regarding a painting's title, medium, substrate, size (height and width in cm), color (hue, saturation, and luminance), and style. The database also marks a select subset of 216 pieces as "boosted", as pieces that are particularly famous and deserve extra publicity (e.g., *Self-Portrait* by Vincent Van Gogh, *Cow's Skull with Calico Roses* by Georgia O'Keeffe, *American Gothic* by Grant Wood). We excluded these boosted pieces from the 4,021 set used by all experiments and experiment analyses, to avoid effects of high familiarity on memory. However, we do analyze the memorability of these pieces with the neural network ResMem.

Procedures. Participants engaged in a continuous recognition task, as has been used in prior work to obtain memorability scores for photographs (2, 3). Participants viewed a stream of 256 images and had to press the "R" key whenever they identified a repeated image from earlier in the sequence. Each image was presented for 1,200 ms, with an 800 ms interstimulus interval. 50 of these images were target images, for which their repetition occurred at least 30 s later. The remaining images were filler foil images, for which some repeated after a brief interval (1 to 5 images) to maintain participant vigilance. From the participant perspective, nothing differentiated targets from foil images. Images were pseudorandomly selected across individual participants' experiments so that each of the 4,021 stimuli were viewed as targets by at least 40 participants (with the foils also coming from the pool of 4,021 stimuli). For each image, we then calculated its HR as the proportion of participants who correctly identified that image as a repeat, amongst those who saw it as a repeated target. We also calculated its false alarm rate (FA) as the proportion of participants who pressed the "R" key on the first presentation of the image, amongst those who saw it as a target. We calculated the memorability score for each image as the CR rate, or $HR - FA$. Prior work has shown that memorability effects hold regardless of whether HR, FA, CR, or d' are used as the measure of memorability (2). The experiment took approximately 9 min and participants were compensated \$1.00 each for their participation.

Analyses. To test whether participants show agreement in what images they remember and forget, we conducted a split-half consistency analysis across 1,000 iterations. For each iteration, the participant pool was split into two random halves, and CR score was computed for each image within each participant half. We then calculated the Spearman rank correlation of the two sets of CR scores (from each participant half) and corrected it with Spearman-Brown split-half reliability correction. The final consistency was taken as the mean correlation coefficient across the 1,000 iterations. A P value was estimated with a permutation test, by shuffling the image order for one of the two participant halves, and then correlating it with the other half. This provides an estimate of the null hypothesis (that there is no relationship between participant halves). The final consistency was compared against 1,000 shuffled null correlations to provide a final P value.

Experiment 2.

Participants. Twenty adults in the Chicago area participated in this in-person experiment. Participants were only recruited if they had not visited the Art Institute of Chicago in the last 6 mo. One participant withdrew partway through due to mobility issues, resulting in a final $N = 19$ (11 females, 8 males; $M_{\text{age}} = 24.1$, $SD = 7.2$). After the participants completed the experiment, the Art Institute closed the wing for renovations and has since changed the pieces in the wing.

Stimuli. Target images for this experiment were the 162 nonboosted pieces from the American Art wing at the Art Institute of Chicago. We also selected 166 foil images for use in the mobile memory experiment. Foil images were selected

from the larger database of 4,021 paintings and were selected to be made in the same region, time period, and media as the target paintings. When possible, foil paintings were also selected from the same artists as the target paintings. This selection process was done to ensure that participants were relying on their memory for specific pieces rather than more general features like the painting's style or subject. Both target and foil images are also part of the larger set of 4,021 paintings used in *Experiment 1*.

We also collected spatial features about each of the 162 target paintings. We mapped out the American Art wing in order to find the exact location of each painting (wall, gallery, and floor). This allowed us to identify which paintings were in the same room (gallery) as each other. We also retrieved measures of physical size (area in cm²) for each painting from the Art Institute's online database.

Procedures. Each participant was instructed to visit both floors of the American Art wing and view all paintings in this wing (exact instructions in the *SI Appendix*). Participants first completed a demographics survey on their phone. Then, participants were instructed to visit one floor (N = 11 began with Floor 1, N = 8 began with Floor 2), complete a survey (a memory test for the pieces on that floor), visit the other floor, and then complete another survey (a memory test for the pieces on that floor). Participants were not explicitly told that they would complete a memory task, however they likely could guess for the second floor that they visited. However, prior work has shown that memorability effects can seamlessly translate between intentional and incidental memory paradigms (12). Participants were allowed to explore the wing freely, with no time or viewing order constraints. If the participants came with friends, they were also able to interact freely with their friends but were instructed not to discuss anything related to the experiment until it was completed. The experimenter was not always present for all participant trials (as instructions were sent over e-mail, see *SI Appendix*), but when the experimenter was present, they stayed separate from participants during their free exploration of the museum. The experiment was conducted when the Art Institute was at a limited 25% capacity due to COVID-19, allowing participants to freely view the pieces without large crowds (the Art Institute typically has 1.5 million visitors annually).

When participants were done viewing all paintings on each floor, they were told to exit to a seat in a hallway where none of the paintings were in view. They then completed self-paced surveys on their mobile phone, in which they viewed each of the 162 paintings randomly intermixed with 166 foil paintings. Each survey contained the target paintings for one of the given floors, as well as an appropriate subset of foil images. Participants had to mark if they did or did not remember seeing each painting, from a scale of 1 (definitely not) to 5 (definitely yes). In our analyses, ratings of 1, 2, or 3 were collapsed to indicate no memory, and ratings of 4 and 5 were collapsed to indicate presence of a memory. Participants were compensated \$10 for their time, in addition to free admission to the Art Institute.

Analysis. In this Experiment, in-person memory for a given painting was measured by its HR, or the proportion of participants who successfully identified seeing that image in the museum. We did not have FA rates for these images (only for the matched foil images), as all paintings in the wing served as targets. To examine the factors contributing to in-person memory, we tested a multiple linear regression model. In this model, we aimed to make predictions at the image-level. Below is how the model was specified:

$$\begin{aligned} (\text{in-person hit rate for a painting}) \sim & \beta_1 * (\text{ResMem score}) \\ & + \beta_2 * (\text{painting size}) + \beta_3 * (\text{floor}) \\ & + \beta_4 * (\text{surrounding ResMem score}) + \beta_5 * (\text{surrounding size}) \\ & + \beta_6 * (\text{ResMem score}): (\text{surrounding ResMem score}) \\ & + \beta_7 * (\text{painting size}): (\text{surrounding size}) + \beta_0. \end{aligned}$$

ResMem score was the predicted memorability for a given image (see *The ResMem Deep Learning Neural Network*). *Painting size* was the real-world area of the painting (cm²). *Floor* was the floor of the wing (1 or 2), where the second floor contained fewer pieces (68 paintings, versus 94 paintings on the first floor) with more varied content. *Surrounding ResMem score* was the average

ResMem-predicted memorability of the other pieces in the same gallery room. *Surrounding size* was the average area of the other pieces in the same gallery room. We also looked at whether there was a statistical interaction between the memorability of the current piece and the surrounding pieces, as well as the size of the current piece and the surrounding pieces. The last term represents the intercept of the model. All predictors were standardized before they were entered into the model.

Experiment 3.

Participants. Forty participants were recruited on the online experimental platform Prolific to provide subjective ratings for the paintings (25 females, 15 males; $M_{\text{age}} = 37.4$, $SD = 12.5$). In order to be recruited, participants had to be located in the United States, have at least a 90% approval rating on at least 50 prior tasks, as well as English as their first language.

Stimuli. The 162 target paintings from *Experiment 2* were used for this experiment.

Procedures. Participants saw a subset of 40 or 41 of the paintings, presented in a random order. For each painting, they were instructed to rate each piece for four attributes on a Likert scale of 1 to 5. Participants rated 1) its beauty ("how beautiful do you find this painting?": *not at all beautiful* to *extremely beautiful*), 2) its emotionality ("what type of emotions do you feel when viewing this painting?": *extremely negative* to *extremely positive*), 3) its interest ("how interesting do you find this painting?": *not at all interesting* to *extremely interesting*), and 4) its familiarity ("how familiar are you with this painting?": *not at all familiar* to *extremely familiar*). The paintings were counterbalanced across experiments so that each painting was rated on the four traits by 10 participants. The task took approximately 12 min, and participants were compensated \$1.60.

Analysis. Average ratings for the paintings were included as additional regressors in the multiple regression model (bolded below):

$$\begin{aligned} (\text{in-person hit rate for a painting}) \sim & \beta_1 * (\text{ResMem score}) \\ & + \beta_2 * (\text{painting size}) + \beta_3 * (\text{floor}) \\ & + \beta_4 * (\text{surrounding ResMem score}) + \beta_5 * (\text{surrounding size}) \\ & + \beta_6 * (\text{ResMem score}): (\text{surrounding ResMem score}) + \beta_7 * (\text{paintingsize}): \\ & (\text{surrounding size}) + \beta_8 * (\mathbf{beauty}) + \beta_9 * (\mathbf{emotion}) \\ & + \beta_{10} * (\mathbf{interest}) + \beta_{11} * (\mathbf{familiarity}) + \beta_0. \end{aligned}$$

All regressors were standardized before they were entered into the model. Model performance was assessed with adjusted R^2 and Akaike Information Criterion (AIC).

We also ran a data-driven stepwise regression with forward selection and backward elimination ($P < 0.05$ to be added, $P > 0.10$ to be removed) to determine the best possible model with all the factors we measured. We included as possible predictors: ResMem score, painting size, floor, surrounding ResMem score, surrounding size, beauty, emotion, interest, familiarity, mean hue, mean saturation, mean luminance, hue contrast, saturation contrast, luminance contrast, visual clutter, spectral energy, number of pieces in the same room, painting category, and the presence of people. The results of this model are reported in *SI Appendix, Table S5*.

The ResMem Deep Learning Neural Network. ResMem is a deep learning neural network designed to predict the intrinsic memorability of an image (17). Several other neural networks have been developed for this purpose (26–30), but ResMem is one of the more recent models and has shown high generalizability to other tasks and image sets (11, 31, 32). Its architecture consists of two branches of processing. The first branch uses the architecture of AlexNet (33), a shallow neural network of 8 layers that has been shown to make successful predictions of object categories from images and shows similarities to the human visual processing stream (34). The second branch uses the architecture of ResNet-152, a deeper neural network of 152 layers that is thought to be able to extract deeper, semantic information from an image due to its depth and ability to skip layers through residual connections (35). It is currently thought that this deeper processing has allowed ResMem to improve performance in comparison to prior networks that used much shallower

architectures (26). The ResMem architecture was trained and validated with close to 70,000 images from a combined image set (26, 36) consisting mostly of photographs of various objects and scenes. These images were tested in a continuous recognition task (as in *Experiment 1*) and memorability was estimated by HR, which was used to train the network. ResMem has been shown to predict human memory performance for photographs with a Spearman rank correlation of $\rho = 0.67$ (17) and has been validated on novel image sets such as black and white scene photographs (25) and posed food photographs (32). However, ResMem had not yet been tested with artwork or predicting memory in the real world.

ResMem is a publicly available tool, where anyone can upload an image and see its predicted memorability score for that image (<https://brainbridgelab.uchicago.edu/resmem/>). We collected ResMem predictions for each painting using this publicly available platform.

Data, Materials, and Software Availability. Data (csv files) have been deposited in Memorability of Art, on the Open Science Framework (<https://osf.io/vhp5d/>) (16).

ACKNOWLEDGMENTS. Thank you to Coen D. Needell for his help with assembling the art pieces for this experiment from the Art Institute of Chicago Online Database, and Max A. Kramer for his help with data analysis and code. Thank you also to Jennifer Hu for piloting the online experiment. Finally, thanks to the Art Institute of Chicago for allowing us to conduct our study with their museum and online database.

1. W. A. Bainbridge, Memorability: How what we see influences what we remember. *Psychol. Learn. Motiv.* **70**, 1–27 (2019).
2. W. A. Bainbridge, P. Isola, A. Oliva, The intrinsic memorability of face photographs. *J. Exp. Psychol. Gen.* **142**, 1323–1334 (2013).
3. P. Isola, J. Xiao, A. Torralba, A. Oliva, "What makes an image memorable?" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Institute of Electrical and Electronics Engineers, Colorado Springs, CO, 2011), pp. 145–152.
4. W. Xie, W. A. Bainbridge, S. K. Inati, C. I. Baker, K. A. Zaghoul, Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. *Nat. Hum. Behav.* **4**, 937–948 (2020).
5. J. D. K. Ongchoco, M. M. Chun, W. A. Bainbridge, What moves us? The intrinsic memorability of dance. *J. Exp. Psychol. Learn. Mem. Cogn.* **49**, 889–899 (2022).
6. C. D. Wakeland-Hart, S. A. Cao, M. T. deBettencourt, W. A. Bainbridge, M. D. Rosenberg, Predicting visual memory across images and within individuals. *Cognition* **227**, 105201 (2022).
7. W. A. Bainbridge, The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia* **141**, 107408 (2020).
8. P. Isola, J. Xiao, D. Parikh, A. Torralba, A. Oliva, What makes a photograph memorable? *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 1469–1482 (2014).
9. L. Goetschalckx, P. Moors, S. Vanmarcke, J. Wagemans, Get the picture? Goodness of image organization contributes to image memorability. *J. Cogn.* **2**, 22 (2019).
10. Z. Bylinskii, P. Isola, C. Bainbridge, A. Torralba, A. Oliva, Intrinsic and extrinsic effects on image memorability. *Vision Res.* **116**, 165–178 (2015).
11. M. A. Kramer, M. N. Hebart, C. I. Baker, W. A. Bainbridge, The features underlying the memorability of objects. *Sci. Adv.*, **9**, eadd2981 (2023).
12. L. Goetschalckx, J. Moors, J. Wagemans, Incidental image memorability. *Memory* **27**, 1273–1282 (2019).
13. W. A. Bainbridge, D. D. Dilks, A. Oliva, Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *Neuroimage* **149**, 141–152 (2017).
14. L. Goetschalckx, P. Moors, J. Wagemans, Image memorability across longer time intervals. *Memory* **26**, 581–588 (2017).
15. R. Cohendet, K. Yadati, N. Q. K. Duong, C.-H. Demarty, "Annotating, understanding, and predicting long-term video memorability" in *ICMR 18: 2018 International Conference on Multimedia Retrieval* (Association for Computing Machinery, Yokohama, Japan, 2018).
16. T. M. Davis, W. A. Bainbridge, Memorability of Art. Open Science Framework. <https://osf.io/vhp5d/>. Deposited 14 July 2022.
17. C. D. Needell, W. A. Bainbridge, Embracing new techniques in deep learning for estimating image memorability. *Comput. Brain Behav.* **5**, 168–184 (2022).
18. T. Valentine, A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A* **43**, 161–204 (1991).
19. N. C. Rust, V. Mehrpour, Understanding image memorability. *Trends Cogn. Sci.* **24**, 557–568 (2020).
20. M. Gedvila, J. D. K. Ongchoco, W. A. Bainbridge, Memorable beginnings, but forgettable endings: Intrinsic scene memorability alters our subjective experience of time. *PsyArXiv [Preprint]* (2023). <https://psyarxiv.com/g4shr> (Accessed 12 April 2023).
21. G. Gillies *et al.*, Tracing the emergence of the memorability benefit. *Cognition*, **238**, 105489 (2023).
22. R. Dubej, J. Peterson, A. Khosla, M.-H. Yang, B. Ghanem, "What makes an object memorable?" in *Proceedings of the IEEE International Conference on Computer Vision* (Institute of Electrical and Electronics Engineers, Santiago, Chile, 2015), pp. 1089–1097.
23. W. A. Bainbridge, The memorability of people: Intrinsic memorability across transformations of a person's face. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 706–716 (2017).
24. J. M. Saito, M. Kolisnyk, K. Fukuda, Judgments of learning reveal conscious access to stimulus memorability. *Psychon. Bull. Rev.* **30**, 317–330 (2023).
25. W. A. Bainbridge *et al.*, Memorability of photographs in subjective cognitive decline and mild cognitive impairment: Implications for cognitive assessment. *Alzheimer's Dement. (Amst.)* **11**, 610–618 (2019).
26. A. Khosla, A. S. Raju, A. Torralba, A. Oliva, "Understanding and predicting image memorability at a large scale" in *Proceedings of the IEEE International Conference on Computer Vision* (Institute of Electrical and Electronics Engineers, Santiago, Chile, 2015), pp. 2390–2398.
27. J. Lu, M. Xu, Z. Wang, "Predicting the memorability of natural-scene images" in *Visual Communications and Image Processing* (Institute of Electrical and Electronics Engineers, Chengdu, China, 2016).
28. L. Goetschalckx, A. Andonian, A. Oliva, P. Isola, "GANalyze: Toward visual definitions of cognitive image properties" in *Proceedings of the IEEE International Conference on Computer Vision* (Institute of Electrical and Electronics Engineers, Seoul, South Korea, 2019), pp. 5744–5753.
29. H. Squalli-Houssaini, N. Q. K. Duong, M. Gwenaelle, C.-H. Demarty, "Deep learning for predicting image memorability" in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (Institute of Electrical and Electronics Engineers, Calgary, CA, 2018), pp. 2371–2375.
30. E. Cetinic, T. Lipic, S. Grgic, A deep learning perspective on beauty, sentiment, and remembrance of art. *IEEE Access* **7**, 73694–73710 (2019).
31. X. Guo, W. A. Bainbridge, Children develop adult-like visual sensitivity to image memorability by the age of four. *bioRxiv [Preprint]* (2022). <https://www.biorxiv.org/content/10.1101/2022.12.20.520853v1> (Accessed 12 April 2023).
32. X. Li, W. A. Bainbridge, A. Bakkour, Item memorability has no influence on value-based decisions. *Sci. Rep.* **12**, 22056 (2022).
33. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2012).
34. R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, A. Oliva, Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* **6**, 27755 (2016).
35. K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Institute of Electrical and Electronics Engineers, Las Vegas, NV, 2016), pp. 770–778.
36. L. Goetschalckx, J. Wagemans, MemCat: A new category-based image set quantified on memorability. *PeerJ* **7**, 8169 (2019).