Trends in Cognitive Sciences

CellPress REVIEWS

Review

Understanding Image Memorability

Nicole C. Rust[1,*] and Vahid Mehrpour[1]

Why are some images easier to remember than others? Here, we review recent developments in our understanding of 'image memorability', including its behavioral characteristics, its neural correlates, and the optimization principles from which it originates. We highlight work that has used large behavioral data sets to leverage memorability scores computed for individual images. These studies demonstrate that the mapping of image content to image memorability is not only predictable, but also non-intuitive and multifaceted. This work has also led to insights into the neural correlates of image memorability, by way of the discovery of a type of population response magnitude variation that emerges in high-level visual cortex as well as higher stages of deep neural networks trained to categorize objects.

Highlights

Individual image memorability scores are predictable from image pixel patterns (e.g., by MemNet, a deep artificial neural network trained for this purpose). However, what drives memorability variation is counterintuitive: naïvely, untrained subjects have misguided notions about what makes images memorable.

The principles by which image content determines image memorability are complex and multifaceted. While we understand many of these principles, a considerable proportion of explainable memorability variance (~25%) remains unexplained.

Pursuit of the neural correlates of image memorability have uncovered a new type of population response magnitude variation in high-level visual cortex. This magnitude variation is strongly predictive of how well images will be remembered.

Brain-analogous image memorability variation emerges at higher stages of deep artificial neural networks trained to categorize objects, suggesting that memorability variation is a consequence of the optimizations required for visual processing.

The Innovation: Quantifying Memorability for Individual Images

In a now classic paper, 'Learning 10,000 pictures', Lionel Standing [1] documented the remarkable ability that humans have to remember whether they've seen an image before, even after seeing thousands of images, each only once and only for a few seconds. Standing also determined that there is no indication that this form of 'image recognition memory' saturates as a function of the number of images viewed, at least up to 10 000 images. Nearly 50 years later, Standing's results remain robust [2] but yet we still understand little about how our brains manage to remember images so well. However, over the past handful of years, rapid advances have been made toward understanding an issue that brings us one step closer to understanding of image recognition memory: 'image memorability', or the systematic variation with which some images are better remembered than others. Progress in understanding image memorability has dovetailed with progresses in object and scene identification [3,4], and similar to that work, has leveraged advances in the training of deep neural networks [5,6], the acquisition of large human behavioral data sets [5,7,8], and population-based approaches for recording neural activity at single-unit resolution [6].

Many of the recent advances in our understanding of what drives image memorability variation can be traced back to the development that memorability can be reliably quantified for individual images [7–9] (Box 1). This development facilitated a better understanding of memorability in multiple ways. First, compared with foundational studies that focused on how one or a few factors impacted image memorability (e.g., image coloration [10] or object distinctiveness [11–13]), image memorability scores for a set of images provided a measure of total memorability variation that could be understood as the weighted combination of several factors (described in detail later) [5,7,8]. This in turn led to an appreciation that the factors that determine image memorability are, naïvely, not intuitive [7], and that a considerable amount of memorability variation remained (and still remains) unexplained [5,8]. Second, despite an incomplete understanding of how image content determines image memorability, the existence of image memorability scores allowed for the pursuit of the representational correlates of image memorability in the brain [6,14–16], as well as neural networks trained to categorize objects [5–7] (Box 2). These investigations led to the discovery of a population response magnitude coding scheme that emerges at higher stages of the visual hierarchy and is predictive of how well images will be remembered [6].

[1]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

*Correspondence: nrust@psych.upenn.edu (N.C. Rust).

Box 1. How Is Image Memorability Quantified?

Individual image memorability scores are typically measured from subjects engaged in a visual recognition memory task online, via Amazon Mechanical Turk or a similar platform, to facilitate measures of memory performance for a large number of subjects (typically ~80 subjects per image). In these memory tasks, subjects typically view one image per trial and report whether that image is novel or a repeat of an image presented in earlier in the sequence (Figure IA). Following on confirmatory analyses that memorability scores preserve their rank with time [7,53], most measures focus on delays of ~4 min and ~100 intervening trials between the first and repeated image presentations. Image memorability scores are computed as the subject average performance at remembering a particular image (the hit rate, HR), corrected for the rate of calling novel images familiar (the false alarm rate, FAR). The false alarm rate correction is most often computed per image [5,17,54] to capture the fact that images can differ in their tendency to appear familiar, even when viewed as novel. Specifically, the image memorability score (MB) for image i is computed using Equation I [5,54]:

$$MB(i) = HR(i) - FAR(i) \quad [I]$$

When required, image memorability scores are also adjusted to equalize for time delays between novel and repeated presentations [5]. The resulting memorability scores are normalized to range from 0 to 1, and they can be loosely interpreted as the fraction of subjects that will remember seeing an image after first seeing it minutes earlier and after seeing many images since. This approach has been applied to quantify memorability for large numbers of images with diverse content, including the publicly available data set LaMem, which contains image memorability scores for 60 000 images [5]. The effect size associated with image memorability variation is considerable: distributions of memorability scores for LaMem range from ~0.2 to 1, with a mean value of 0.76 (Figure IB). Memorability scores maintain their ranks across timescales ranging from minutes to weeks [53], and reliable image memorability scores do not require a subject to actively be engaged in a memory task [55]. The same task design has also been used in one animal model to measure image memorability behavior and its neural correlates: the rhesus monkey [6]. A related measure that has been used to quantify memorability is d' [17], which z-scores HR and FAR before taking their difference [56].

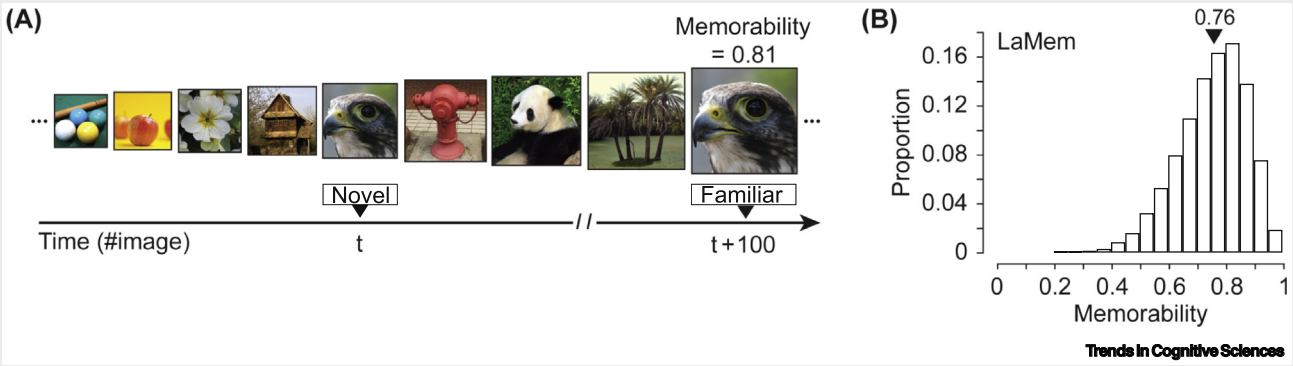


Figure I. Quantifying Image Memorability. (A) The visual recognition memory task used to compute image memorability scores. On each trial, subjects judge whether images are novel or familiar. Memorability scores are computed based on the subject-average performance for familiar images, corrected for false alarms. (B) Distribution of memorability scores for the LaMem data set, ~60 000 images pulled from a diversity of sources [5].

The Behavioral Characteristics of Image Memorability

The Factors that Drive Image Memorability Are Predictable but not Intuitive

Image memorability is typically investigated in the context of a recognition memory task ('Have you seen this image before?'), quantified as described in Box 1. In developing the techniques

Box 2. Image Memorability and Deep Artificial Neural Networks

Deep artificial neural networks have contributed in two conceptually different ways to recent advances in our understanding of image memorability. In the first instance, they have been applied as engineering tools to quantify and manipulate the memorability of images. For example, one deep neural network, MemNet, was developed to predict memorability scores for arbitrary images [5], and this tool has been used as a proxy for image memorability scores in investigations of the neural correlates of memorability [6]. Similarly, the multicomponent neural network GANalyze was developed to receive images as input and produce new images with similar content but manipulated image memorability [28]; this tool has been used to gain insight into the image properties that drive memorability.

In the second instance, deep neural networks have been used as models of how memorability emerges from neural processing in the brain. This application complements illustrations that deep neural networks trained to categorize objects (i.e., to accept images as input and produce object category labels as output) have a functional organization that bears considerable resemblance to the form processing pathway in the human and nonhuman primate brain [3,4]. Remarkably, image memorability variation also emerges in deep neural networks trained for object categorization [6].

to measure memorability, researchers were careful to consider the sources of memorability variation. If what makes an image more or less memorable is dominated by individual differences, measures of subject-averaged memorability performance will fail to capture it. However, if a large fraction of memorability variation is shared across different individuals, memorability can be meaningfully captured at the resolution of individual images through subject-averaged measures of memory behavior. Consistent with the latter notion, memorability scores (Box 1) are highly correlated across random splits of large subject pools (average Spearman rank correlations: $\rho = 0.68$ for faces [8,17] and $\rho = 0.68–0.78$ for images that contain faces as well as other content [5,7,18]), thereby establishing that image memorability scores do in fact capture reliable properties of individual images. Remarkably, memorability variation is not only consistent across different human subjects, but is also correlated between humans and rhesus monkeys [6], suggesting that memorability does not depend on extensive experience with the objects contained in images (such as cars and fire hydrants) and it does not require the capacity for language.

One striking finding is that, naïvely, subjects are bad at predicting how memorable images are: when untrained subjects were asked to predict memorability, their predictions and actual memorability scores were at best weakly correlated [7,8,17]. Rather, subjects' memorability predictions were strongly correlated with judgments of image esthetics and interestingness, which were each only weakly correlated with memorability (but were strongly correlated with one another) [7]. Examples of this are shown in Figure 1A, including one image that was predicted to be forgettable but had a high memorability score (Figure 1A, top) and another image that was predicted to be memorable but had a low memorability score (Figure 1A, bottom). This suggests that we have misguided notions about what makes images memorable. However, that is not to say that image memorability is not predictable: a convolutional neural network trained to predict image memorability ('MemNet' [5]), predicts memorability scores for untrained images near the noise ceiling present in the human behavioral data (Spearman's rank correlations between MemNet predictions and human behavior $\rho = 0.64$; Spearman's rank correlations between random splits of the human subject pool for the same image set $\rho = 0.68$ [5]). This confirms that, while image memorability is counterintuitive for humans, it is predictable from image pixel patterns (Box 2).

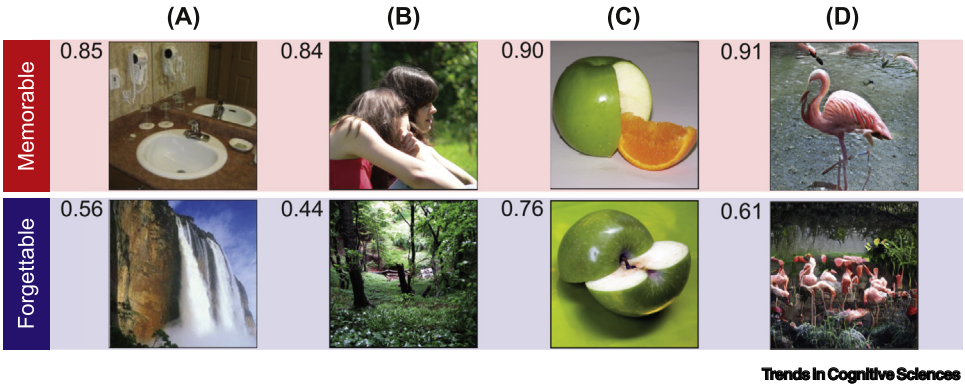


Figure 1. Examples of Memorable and Forgettable Images. Image memorability scores are labeled to the left of each image. (A) Examples illustrating that naïve untrained subjects have misguided notions about image memorability. Top: example of an image that was predicted to be forgettable by naïve subjects but that had a high memorability score. Bottom: example of an image that was predicted to be memorable by naïve subjects but that had a low memorability score. (B) Images containing people tend to be highly memorable, whereas nature scenes tend to have low memorability. (C) Atypical depictions of objects tend to be more memorable than typical depictions. (D) Two example images produced by GANalyze from the same seed image, with enhanced and reduced memorability. Images from [7] (A) and [28] (D).

The Mapping of Image Content to Image Memorability Is Multifaceted

Several foundational studies published before the acquisition of memorability scores for individual images documented a number of different types of image content that impact image memorability. These include lower-level image properties, such as the presentation of images in color as opposed to grayscale [10] and viewing images in 3D as opposed to 2D [19]. These also include higher-level image properties, such as distinctiveness and atypicality, particularly for faces [11–13]. In addition, these earlier studies established that conceptual distinctiveness (such as category membership) impacts memorability in a manner that cannot be accounted for by perceptual distinctiveness alone [20,21]. At the same time, this earlier work emphasized the rich details with which visual memories are stored, including the fact that we remember considerable detail about the configurations and contexts that we view objects in [2] and that we rely on specific detailed information to remember whether we have seen an image previously [22]. However, we are typically bad at remembering random patterns unless they take on object-like qualities [23], suggesting that visual memory is not driven entirely by visual details. These results have been summarized as 'meaningfulness' contributing to memorability, because images that are meaningful insofar as they contain recognizable content are better remembered than those that do not [24].

The quantification of memorability scores for individual images built on these foundational results to determine the relative importance of different known factors, discover new factors, and determine the fraction of total image memorability variation that all factors together could explain, once combined. Examples of the largest factors that drive image memorability variation include the fact that images containing people are on average highly memorable (average memorability score = 0.82), in contrast to images of nature scenes, which have lower average memorability (average memorability score = 0.61) [7] (Figure 1B). Images of atypical content, such as a chair shaped like a hand [25], are also typically highly memorable (average memorability score = 0.83) [5] (Figure 1C). The emotional valence of an image also impacts its memorability, where images that evoke disgust, amusement, and fear are on average more memorable, whereas images that evoke awe and contentment are, on average, less memorable [5]. As described earlier, memorability is only weakly correlated with subjective judgments of image esthetics and interestingness [5]. Similarly, while low-level image properties such as color and simple image features do contribute to memorability variation, they only account weakly for it [7,26].

When considered together, how much memorability variation do all known factors account for? To address this question, one study scored and regressed 127 semantic attributes (e.g., open/closed; static/dynamic; frightening/funny, person/not, etc.) against image memorability scores and found that together they accounted for ~75% of explainable variance (where explainable variance refers to the variance preserved across random splits of human subjects [7]). This suggests that, at the same time that we can describe many of the principles that dictate image memorability, a considerable fraction of this variation remains uncharacterized.

When restricted to images of faces, similar principles to those established for images at large generally hold. In line with earlier literature [11–13], studies targeted at measuring the memorability of individual face images found that atypical faces tended to be more memorable than typical ones [8]. At the same time, the perceived distinctiveness of faces, measured as the combined influence of several terms (e.g., atypical, uncommon), could not fully explain face memorability [8,27]. Regressions of 20 attributes (e.g., interesting/boring, calm/aggressive, etc.) against image memorability scores revealed several additional attributes that enhance the memorability of faces, including faces that are judged as intelligent, responsible, trustworthy, attractive, and kind [8,17]. When considered together, scorings for all 20 attributes only accounted for ~75%

of explainable variance in image memorability scores [8], indicating that, similar to images at large, a considerable amount of image memorability variation for faces remains unexplained by these attributes. In addition, these studies determined that across different views of a face, memorability tends to be preserved [17].

Further insight into the relationship between image content and image memorability has been gained from a technological advance that combined the power of Generative Adversarial Networks (GANs) with a deep neural network trained to predict image memorability (MemNet [5], described earlier). This network, GANalyze, receives images as input and produces new images with minimal modifications to image content but parametrically manipulated memorability [28] (Figure 1D). Human behavioral experiments confirmed that the network successfully manipulates image memorability, as intended [28]. What image properties does the network manipulate to enhance image memorability? As might be expected, the network tends to make images brighter and more colorful. In addition, the network adjusts a handful of previously unappreciated factors, including: increasing object size, centering the objects within images, and uncluttering the backgrounds of the objects. Intriguingly, the network also tends to make objects more square or circular.

Image Memorability for Recognition Memory versus Recollection

Image memorability has been investigated most extensively by probing 'recognition memory': asking subjects to report whether images are novel or familiar (Box 1). A distinct but complementary memory task, investigated extensively for lists of words [29], is one that requires subjects to view images and then recall what they have seen, absent any cue. Are the same images that are most difficult to remember in a recognition memory task also the most difficult to recall? One recent study took on the challenge of quantifying memorability variation for image recollection by asking subjects to view images and then later draw them [30]. They found no relationship between image memorability variation quantified for recollection versus recognition memory [30]. These results are consistent with notions that image memorability variation for these two memory tasks may be distinct.

Memorability and other Cognitive Phenomena

Image memorability is both correlated with, and distinguishable from, several other cognitive phenomena. One of these is visual salience, which refers to the fact that, when we look at images, certain regions tend to pop out and grab our attention. The visual salience of an image can be quantified based on the regularity with which patterns of fixations during free viewing are consistent across subjects [31], and several studies have determined that measures of visual salience are correlated with image memorability [5,26,32–34]. Similar to memorability, images tend to be more salient when they contain one or a few objects [26,31], including images that are presented more close-up and in an uncluttered context [5]. However, when images contain multiple objects and multiple points of fixation, the correlation between memorability and salience drops considerably [26], suggesting that memorability and salience are distinguishable. Similarly, differences in memorability exist across face images that are identical in saliency in terms of their shapes, parts, image features, and fixation patterns [8,17]. Together, these results suggest that, while the factors that determine visual salience and memorability are correlated, they are not one and the same. By a similar logic, a recent study investigated the relationship between memorability and several cognitive factors, including manipulations of bottom-up attention (through spatial cuing and visual search), manipulations of top-down attention (through cognitive control and depth of encoding), and priming [35]. None of these factors were able to account for modulations of memorability, suggesting that memorability is distinct from these other phenomena.

Image Memorability Depends on Image Set Context

Image memorability scores are highly replicable when images are viewed in the context of a sequence of other, randomly selected images. When the same images are viewed in a sequence of images selected from the same category (e.g., when a picture of a lighthouse is embedded in a sequence of other lighthouse images), image memorability scores remain equally reliable across subjects, but they take on new values [33]. Contextual changes in the magnitude and sign of image memorability scores relative to the random benchmark can be predicted by how distinct an image is from other images within the new image set (quantified based on the image activation patterns of a deep neural network trained for object and scene categorization) [33]. In the categorical context, images that produce the most similar activation patterns to the others in the set are the ones that undergo the largest decrements in memorability, whereas memorability can increase modestly for the images that are the most distinct. These results establish that a full account of image memorability requires a description not only of individual image identities, but also the context within which those images are embedded.

The Neural Correlates of Image Memorability

Image Memorability Is Reflected in the Magnitude of the Response to Novel Images

What neural mechanisms shape image memorability? In principle, image memorability variation could be a consequence of variation in how images are represented when they are viewed as novel, which is then carried over to variation in how well those images are remembered. Alternatively, image memorability variation could emerge for the first time when images are viewed as familiar, implying that memorability arises via the mechanisms involved in memory storage and/or memory signaling. As described later, existing evidence supports the former account (while not ruling out added contributions from the latter).

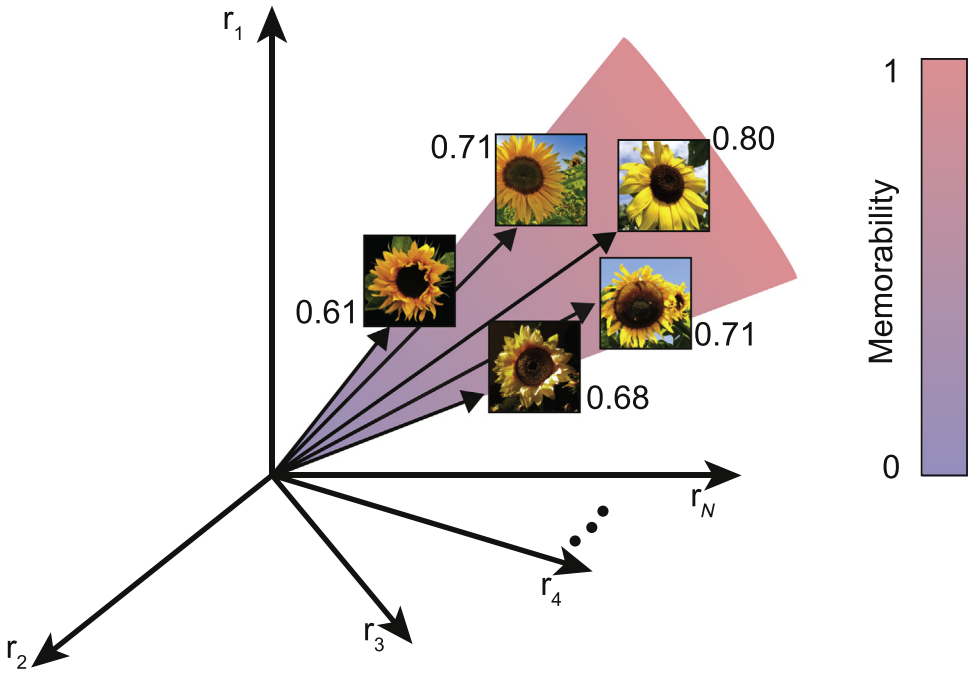
The first reports of the neural correlates of image memorability variation utilized human fMRI to pinpoint its locus [14,15]. These studies reported that image memorability within a category (e.g., faces or scenes) could be classified by increases in blood oxygen level-dependent (BOLD) activation in high-level visual cortex, as well by decoding the patterns of BOLD activation across voxels, as subjects viewed novel images. Similar results were determined with human electroencephalogram (EEG) responses to ambiguous pictures of faces, where images that evoked larger N170 activity during the novel viewing period were more likely to be remembered [24]. Similar results were also reported in a study whereby image identity was more robustly decoded from human magnetoencephalography (MEG) responses for more as opposed to less memorable novel images, even under conditions in which images were not remembered at all [i.e., in a rapid serial visual presentation (RSVP) sequence at short durations and with masking [16]]. Together, these results suggest that the neural correlates of image memorability are reflected in the visual representations of images when they are viewed for the first time. These results are complementary but distinct from earlier work on ‘subsequent memory effects’ where items are sorted for each subject into ‘remembered’ versus ‘not remembered’ and remembered items tend to evoke indicators of higher neural activity during memory encoding (reviewed by [36]). Subsequent memory effects have been recapitulated in fMRI image memorability investigations, both within high level visual cortex as well as other structures (e.g., the medial temporal lobe, and prefrontal and parietal cortex [14,15]). In comparison, the memorability activation patterns described earlier remain whether subjects remember images or not [15,16,24,27], and are limited to high-level visual cortex as well as the medial temporal lobe [14,15]. These results are consistent with the interpretation that image memorability effects follow from properties that are associated with images as opposed to other factors (such as stimulus-independent fluctuations in the attentional state of an observer).

Conceptualizing Image Memorability and Object Identity Representations

What exactly differs in the visual representations of novel images that are more as compared with less memorable? To address this question, one study recorded population activity at single-unit resolution from monkey inferotemporal cortex (ITC) as the monkeys performed a visual memory task similar to that described in Box 1 [6]. This study reported a strong correlation between image memorability scores and the overall magnitude of the ITC population response to novel images (Pearson correlation $\rho = 0.62$), where the most memorable images evoked ~20% larger magnitude responses versus images that were the least memorable. Notably, the existence of population magnitude variation in response to natural images had not previously been appreciated before investigations of image memorability in ITC, despite extensive investigation of visual representations in this structure (reviewed in [3,4,37]). Population response magnitude variation or equivalently 'magnitude coding' in ITC may have been overlooked due to its relatively subtle (albeit measurable) impact on perceptual behavior [38,39], coupled with assumptions that this type of variation is largely eliminated by neural mechanisms that work to maintain constant global firing rates across a cortical population (such as homeostatic plasticity [40] and divisive normalization [41]). By contrast, investigations of memorability demonstrate that population response magnitude variation can be considerable (up to 20%) and that this variation strongly covaries with at least one type of behavioral change: how well images will be remembered.

How can current accounts of object representations in high-level visual cortex be extended to incorporate image memorability variation? That is, how might brain areas such as ITC reflect representations in which different images of the same object are identified as the same and, simultaneously, some of those images are more memorable than others? The multiplexing of object identity and memorability in ITC has been proposed to happen through two complementary coding schemes: a spike pattern coding scheme for object identity, and spike magnitude coding scheme for image memorability [6]. In the case of object identity, spike pattern coding is a consequence of individual ITC neurons that are selectively responsive or 'tuned' for the high-level image properties that define objects. This translates into representations of different images that are reflected by different population spike patterns, and in the high-dimensional neural representational space that is typically used to conceptualize object representations (Figure 2), different angular directions for different images [37,42]. Given that ITC neurons tend to maintain their rank-order object selectivity across different transformations of an object (e.g., changes in position or background [37,42]), representations of different images containing the same object tend to cluster in this space (Figure 2) [37]. In this format, object identity can be easily decoded, for example, by determining the object cluster that a particular ITC population response pattern is most similar to (or by other variants of linear population decoding [37]). In comparison, memorability is reflected by the magnitude of the ITC population response, thus allowing for some images within an object cluster to be more memorable than others (Figure 2).

While Figure 2 provides a simple and intuitive account of ITC that is supported by considerable evidence, there are also results that it cannot account for [27]. Namely, more memorable images have more similar fMRI voxel response patterns compared with less memorable images, as assessed by a representational similarity analysis [14,15]. This result, replicated across two studies, suggests that, in addition to the reflection of memorability via a magnitude coding scheme in ITC (Figure 2), some aspect of memorability may be reflected via spike pattern coding. This result conflicts with proposals in which memorability follows from a multidimensional representational space where more typical (and less memorable) objects are represented more centrally, while more distinctive (and more memorable) items are represented more distantly [43,44]. One question going forward will be to understand how this finding integrates with the depiction presented in Figure 2.



Trends in Cognitive Sciences

Figure 2. Memorability Is Reflected in the Magnitude of the Population Response in High-Level Visual Cortex. In geometric depictions of how high-level visual cortex represents image identity, the population response to an image is depicted as a vector in an *N*-dimensional space, where *N* is determined by the number of neurons in the population. The direction that a population response vector points is determined by the amount of activation of each neuron. Identity is thought to be largely encoded by population vector direction. In comparison, image memorability is thought to be encoded by the magnitude (or equivalently length) of each population vector, where images that produce larger population responses are more memorable. Shown is a hypothetical depiction of a cluster of images all with the same content, but different image memorability.

Image Memorability Is Reflected in Deep Artificial Neural Networks

Where does the population response magnitude variation associated with image memorability come from? Intriguingly, image memorability variation and its brain-analogous correlates emerge at higher stages of deep artificial neural networks trained to categorize objects (Box 2) [6], and this result holds across different deep neural network architectures and optimization schemes (e.g., AlexNet [45], the Hybrid CNN [46], and VGG-16 [47]) [6]. Specifically, at earlier stages of these networks (e.g., the V1-analgous layers), population response magnitude only weakly correlates with image memorability, consistent with observations that low-level image properties only contribute weakly to image memorability variation [7] (Figure 3). Correlations between population response magnitude and image memorability grow in strength across the hierarchy of these networks up to the layers that are analogous to high-level visual cortex [6] (Figure 3). At the very highest layers of the network, correlations often begin to drop (e.g., the fully connected layers of VGG-16, Figure 3), consistent with representations that become categorically invariant (i.e., these layers increasingly respond in an identical manner to all images containing the same object). Stated differently, layers analogous to high-level visual cortex in deep neural networks trained to categorize objects respond more vigorously to some images than others, and the vigor of these responses is predictive of the images that we find most memorable. These findings complement illustrations that deep neural networks trained to categorize objects have a functional organization that bears considerable resemblance to the brain areas that comprise the form processing pathway in humans and nonhuman primates [3,4]. However, what makes the memorability result so compelling is its emergence: compared with optimizing a deep neural

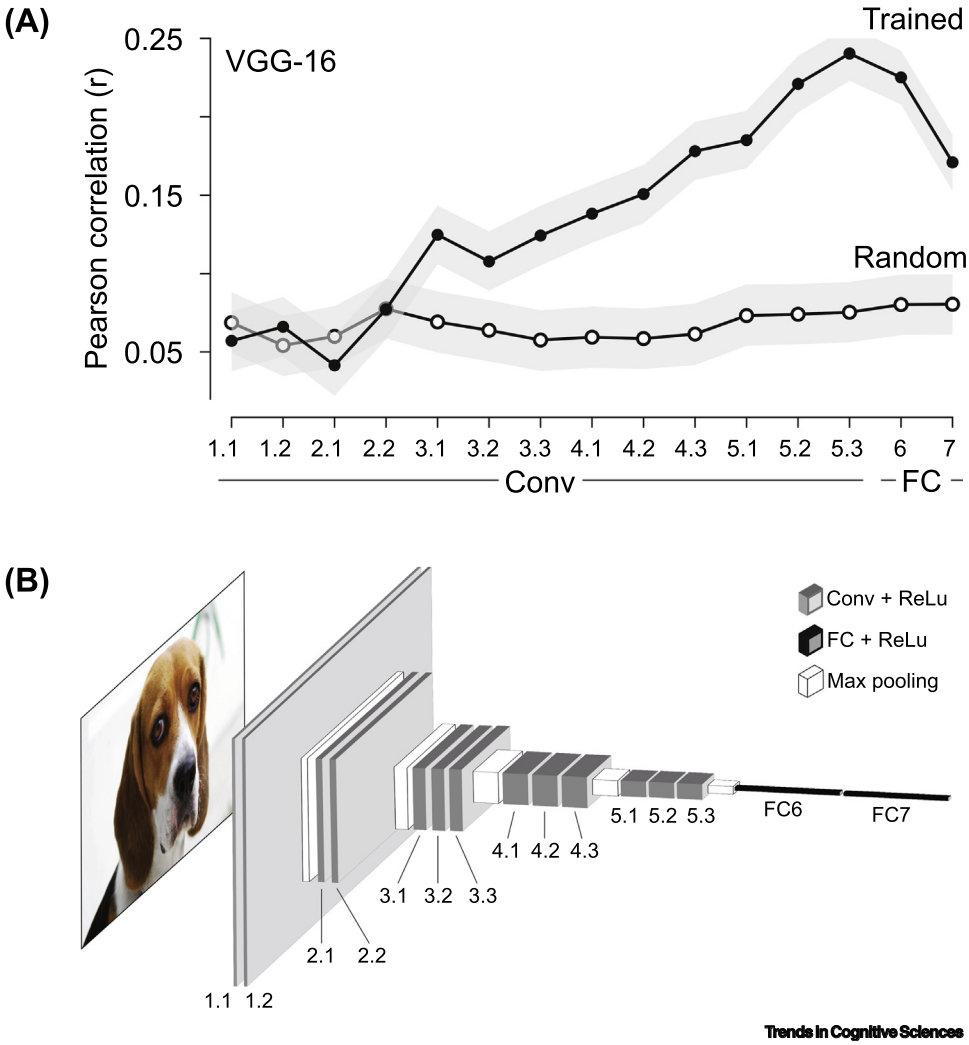


Figure 3. Representation of Image Memorability across Different Layers of a Deep Artificial Neural Network Trained for Object Categorization, VGG-16. (A) Pearson correlation between population response magnitude and image memorability scores computed for different layers of one neural network, VGG-16 [47]. The correlations for both a randomly connected version of the network and a fully trained network are shown, replotted from [6]. 'Conv', convolutional layer; 'FC', fully connected layer; 'ReLu', rectified linear unit. (B) Depiction of the VGG-16 network architecture. This architecture includes 13 convolutional layers (Conv), three fully connected layers (FC), and five Max pooling layers. The last layer of VGG-16 (FC8) has been omitted in both panels, because it reflects the output of a 1000-way classification. Adopted from [69] (B).

network for object categorization and finding that brain-like object representations emerge, memorability representations emerge in deep neural networks that are trained for object categorization but not explicitly to predict image memorability (nor do they, once trained, have a memory trace of anything that they have 'seen').

These results suggest that image memorability variation is shaped by the optimizations required for object-based (as opposed to memory-based) processing. However, intuitively, how do object-based optimizations shape network activity? For example, atypical images tend to be more memorable than typical ones; how is image atypicality reflected, as well as shaped, in a deep neural network trained for object categorization? Many questions remain.

Concluding Remarks and Future Directions

The fact that different human individuals [5,7,8,17,33], and even humans and rhesus monkeys [6], tend to find the same images memorable and forgettable is striking. However, arriving at a simple and parsimonious explanation of what drives this shared image memorability variation in terms of image content appears unlikely: not only is this mapping naïvely not intuitive [7], but it also depends on a diversity of factors that range from object atypicality to emotional valence [5,8,11–13], and even large sets of reasonably selected semantic attributes leave ~25% of explainable variance unaccounted for [7,8] (see Outstanding Questions). In comparison, first-order descriptions of the neural correlates of image memorability have proven to be relatively straightforward, where image memorability is strongly predicted by variation in population response magnitude, both in high-level visual cortex (Figure 2) and in deep neural networks trained to categorize objects (Figure 3) [6]. The apparent misalignment between the seeming complexity of image memorability behavior and the seeming simplicity of its neural correlates likely reflects the fact that many different factors combine to determine population response magnitude in high-level visual cortex. At the same time, whether the neural correlates of image memorability prove to be much more elaborate than what has been revealed thus far remains to be seen.

Going forward, image memorability can provide an important complement to object identification behavior for probing and constraining descriptions of how representations of different behaviorally relevant variables are transformed across the primate visual form processing pathway [3,4,48]. Analogous to the alignment of human and monkey object identification behavior [49,50], humans and monkeys tend to find the same images memorable and forgettable [6], thereby allowing for these investigations to be conducted with neural data collected in the brains of animals at high spatial and temporal resolution, as well as causal tests of existing hypotheses via perturbation approaches [51,52]. Moreover, the strong relationship between image memorability behavior and population response magnitude in high-level visual cortex (Figure 2) prompts a host of qualitatively new questions about how different types of memorability variation are reflected, as well as the relationship between neural representations of image memorability and visual salience.

At the behavioral level, image memorability cannot be isolated from memory, because image memorability is the systematic variation in the ability of subjects to report whether they have seen an image before. However, at the level of underlying mechanism, image memorability could be distinguished from memory: insofar as image memorability variation arises entirely from variation in the robustness of visual representations (which, in turn, has consequences for memory storage), one can regard the source of image memorability variation as 'visual processing'. That said, it is crucial to appreciate that current accounts of image memorability are fundamentally incomplete, because they lack descriptions of how visual representations are transformed into visual memories.

Box 3. Visual Memory Storage and Visual Familiarity Signaling

Visual memory storage and the signaling of whether an image is novel or familiar have been linked to processing within ITC and its primary projection area, perirhinal cortex. Within those structures, familiarity is reflected by repetition suppression: adaptation-like reductions in the population response to repeated images [57–61]. Potentially consistent with image memorability behavior, repetition suppression acts multiplicatively [62–64], and more repetition suppression is expected to follow when more memorable images are repeated (as a consequence of more vigorous responses). As an illustrative example, the same proportional reduction (e.g., 10%) applied to a larger compared with a smaller quantity (e.g., 100 versus 10) will result in a larger reduction (e.g., 10 versus one, respectively). However, whether repetition suppression in ITC and perirhinal cortex can fully account for image memorability behavior is still unclear. Other evidence suggests a role for the hippocampus in supporting visual recognition memory behavior [65,66], particularly in scenarios where subjects are asked to judge the familiarity of an image that is visually similar to one that has been viewed previously [67]. In those cases, the hippocampus is proposed to contribute via a process known as pattern separation [68]. Outside the medial temporal lobe, frontoparietal regions have been implicated in supporting visual memory behavior by way of differential responses elicited by these structures to images that are remembered versus those that are not [14,15].

Outstanding Questions

Is a simple and parsimonious account of image memorability attainable? The mapping of image content to image memorability described thus far is multifaceted. Can it be unified within a yet undetermined theoretical framework? The mapping of neural responses to behavior described so far is straightforward. Will this simplicity hold as we learn more?

How are different types of image memorability variation, as well as the neural correlates of related behaviors, reflected within and outside of visual cortex? Unknown is whether the neural correlates of some types of memorability variation, such as emotional valence, are better reflected in structures outside of visual cortex (e.g., the amygdala). Also unknown is how the neural variation associated with memorability overlaps with other related types of variation, including visual salience and priming.

What types of image memorability variation are and are not reflected in deep artificial neural networks trained for object categorization? It appears unlikely that some of the semantic properties that correlate with image memorability would emerge from deep neural networks trained to categorize objects; for example, correlations with emotional valence. Which factors emerge and which do not?

What neural correlates account for the contextual sensitivity of image memorability behavior? The memorability of an image depends on the set of images in which it is embedded. The magnitude and sign of these contextual effects depend on how distinct an image is from other images within the image set. What neural mechanisms drive these contextual effects?

How are the more robust visual responses that are associated with image memorability converted into more robust remembering? Image memorability provides a continuous way to manipulate visual memory behavior, and this manipulation could be exploited in future investigations of visual memory.

While several different brain areas have been implicated in visual memory storage (Box 3), we do not understand how processing in these different regions combines to support memory. What is clear is that image memorability variation can serve as an important way to continuously manipulate visual memory in future investigations of its neural correlates and thereby facilitate answers to a question that has largely eluded the field for 50 years: how do our brains manage to remember images so well [1]?

Acknowledgments

This work was supported by the National Eye Institute of the National Institutes of Health (award R01EY020851), the National Science Foundation (CAREER award 1265480), and the Simons Foundation (Simons Collaboration on the Global Brain award 543033).

Citation Diversity Statement

Recent work in neuroscience and related fields has identified citation biases whereby work from women and minorities are undercited relative to other papers in the field [70–72]. In crafting this manuscript, we sought to proactively consider citation bias. Following on [70], the gender balance of citations was quantified from the first names of the first and last authors using open source code [73]. Excluding self-citations, the references for this manuscript contain 50.0% man/man, 15.1% man/woman, 21.2% woman/man, and 13.6% woman/woman citations. Expected proportions estimated from five top neuroscience journals, as reported in [70], are 58.4% man/man, 9.4% man/woman, 25.5% woman/man, and 6.7% woman/woman.

References

1. Standing, L. (1973) Learning 10,000 pictures. *Q. J. Exp. Psychol.* 25, 207–222
2. Brady, T.F. *et al.* (2008) Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14325–14329
3. Kriegeskorte, N. (2015) Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* 1, 417–446
4. Yamins, D.L. and DiCarlo, J.J. (2016) Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365
5. Khosla, A. *et al.* (2015) Understanding and predicting image memorability at a large scale. In *International Conference on Computer Vision (ICCV)*, pp. 2390–2398, IEEE
6. Jaegle, A. *et al.* (2019) Population response magnitude variation in inferotemporal cortex predicts image memorability. *eLife* 8, e47596
7. Isola, P. *et al.* (2014) What makes a photograph memorable? *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 1469–1482
8. Bainbridge, W.A. *et al.* (2013) The intrinsic memorability of face photographs. *J. Exp. Psychol. Gen.* 142, 1323–1334
9. Isola, P. *et al.* (2011) What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 145–152, IEEE
10. Wichmann, F.A. *et al.* (2002) The contributions of color to recognition memory for natural scenes. *J. Exp. Psychol. Learn Mem. Cogn.* 28, 509–520
11. Bartlett, J.C. *et al.* (1984) Typicality and familiarity of faces. *Mem. Cogn.* 12, 219–228
12. Bruce, V. *et al.* (1994) What's distinctive about a distinctive face? *Q. J. Exp. Psychol. A* 47, 119–141
13. Vokey, J.R. and Read, J.D. (1992) Familiarity, memorability, and the effect of typicality on the recognition of faces. *Mem. Cogn.* 20, 291–302
14. Bainbridge, W.A. and Rissman, J. (2018) Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Sci. Rep.* 8, 8679
15. Bainbridge, W.A. *et al.* (2017) Memorability: a stimulus-driven perceptual neural signature distinctive from memory. *Neuroimage* 149, 141–152
16. Mohsenzadeh, Y. *et al.* (2019) The perceptual neural trace of memorable unseen scenes. *Sci. Rep.* 9, 6033
17. Bainbridge, W.A. (2017) The memorability of people: Intrinsic memorability across transformations of a person's face. *J. Exp. Psychol. Learn Mem. Cogn.* 43, 706–716
18. Goetschalckx, L. and Wagemans, J. (2019) MemCat: a new category-based image set quantified on memorability. *PeerJ* 7, e8169
19. Valsecchi, M. and Gegenfurtner, K.R. (2012) On the contribution of binocular disparity to the long-term memory for natural scenes. *PLoS One* 7, e49947
20. Huebner, G.M. and Gegenfurtner, K.R. (2012) Conceptual and visual features contribute to visual memory for natural images. *PLoS ONE* 7, e37575
21. Konkle, T. *et al.* (2010) Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J. Exp. Psychol. Gen.* 139, 558–578
22. Vogt, S. and Magnussen, S. (2007) Long-term memory for 400 pictures on a common theme. *Exp. Psychol.* 54, 298–303
23. Wiseman, S. and Neisser, U. (1974) Perceptual organization as a determinant of visual recognition memory. *Am. J. Psychol.* 87, 675–681
24. Brady, T.F. *et al.* (2019) The role of meaning in visual memory: face-selective brain activity predicts memory for ambiguous face stimuli. *J. Neurosci.* 39, 1100–1108
25. Saleh, B. *et al.* (2013) Object-centric anomaly detection by attribute-based reasoning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 787–794, IEEE
26. Dubey, R. *et al.* (2015) What makes an object memorable? In *IEEE International Conference on Computer Vision (ICCV)*, pp. 1089–1097, IEEE
27. Bainbridge, W.A. (2019) Memorability: how what we sees influences what we remember. In *Psychology of Learning and Motivation* (Federmeier, K. and Beck, D., eds), pp. 1–27, Elsevier
28. Goetschalckx, L. *et al.* (2019) GANalyze: toward visual definitions of cognitive image properties. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5744–5753, IEEE
29. Yonelinas, A.P. (2002) The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517
30. Bainbridge, W.A. *et al.* (2019) Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nat. Commun.* 10, 5
31. Judd, T. *et al.* (2009) Learning to predict where humans look. In *IEEE 12th International Conference on Computer Vision (ICCV)*, pp. 2106–2113, IEEE
32. Mancas, M. and Le Meur, O. (2013) Memorability of natural scenes: the role of attention. In *IEEE International Conference on Image Processing*, pp. 196–200, IEEE

33. Bylinskii, Z. *et al.* (2015) Intrinsic and extrinsic effects on image memorability. *Vis. Res.* 116, 165–178
34. Celikkale, B. *et al.* (2013) Visual attention-driven spatial pooling for image memorability. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 976–983, IEEE
35. Bainbridge, W.A. (2020) The resiliency of image memorability: a predictor of memory separate from attention and priming. *Neuropsychologia* 141, 107408
36. Paller, K.A. and Wagner, A.D. (2002) Observing the transformation of experience into memory. *Trends Cogn. Sci.* 6, 93–102
37. DiCarlo, J.J. *et al.* (2012) How does the brain solve visual object recognition? *Neuron* 73, 415–434
38. Potter, M.C. (2012) Recognition and memory for briefly presented scenes. *Front. Psychol.* 3, 32
39. Eberhardt, S. *et al.* (2016) How deep is the feature analysis underlying rapid visual categorization? In *29th Conference on Neural Information Processing Systems* (Lee, D.D. and von Luxberg, U., eds), pp. 1108–1116, Curran Associates Inc
40. Turrigiano, G. (2012) Homeostatic synaptic plasticity: local and global mechanisms for stabilizing neuronal function. *Cold Spring Harb. Perspect. Biol.* 4, a005736
41. Carandini, M. and Heeger, D.J. (2011) Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62
42. Kriegeskorte, N. *et al.* (2008) Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4
43. Valentine, T. (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A* 43, 161–204
44. Lukavsky, J. and Dechterenko, F. (2017) Visual properties and memorising scenes: effects of image-space sparseness and uniformity. *Atten. Percept. Psychophys.* 79, 2044–2054
45. Krizhevsky, A. *et al.* (2012) ImageNet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems, NeurIPS* (60), pp. 84–90
46. Zhou, B. *et al.* (2014) Learning deep features for scene recognition using places database. *Adv. Neural Inf. Proces. Syst.* 27, 5349
47. Simonyan, K. and Zisserman, A. (2015) Very deep convolutional networks for large-scale image recognition. *arXiv* 2015 1409.1556v6
48. Schrimpf, M. *et al.* (2020) Brain-Score: which artificial neural network for object recognition is most brain–like? *bioRxiv.* Published online January 2, 2020. https://doi.org/10.1101/407007
49. Rajalingham, R. *et al.* (2015) Comparison of object recognition behavior in human and monkey. *J. Neurosci.* 35, 12127–12136
50. Kriegeskorte, N. *et al.* (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141
51. Afraz, A. *et al.* (2015) Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. *Proc. Natl. Acad. Sci. U. S. A.* 112, 6730–6735
52. Rajalingham, R. and DiCarlo, J.J. (2019) Reversible inactivation of different millimeter-scale regions of primate IT results in different patterns of core object recognition deficits. *Neuron* 102, 493–505
53. Goetschalckx, L. *et al.* (2018) Image memorability across longer time intervals. *Memory* 26, 581–588
54. Khosla, A. *et al.* (2013) Modifying the memorability of face photographs. In *International Conference on Computer Vision (ICCV)*, pp. 3200–3207, IEEE
55. Goetschalckx, L. *et al.* (2019) Incidental image memorability. *Memory* 27, 1273–1282
56. Stanislaw, H. and Todorov, N. (1999) Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149
57. Desimone, R. (1996) Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13494–13499
58. Bogacz, R. and Brown, M.W. (2003) Comparison of computational models of familiarity discrimination in the perirhinal cortex. *Hippocampus* 13, 494–524
59. Xiang, J.Z. and Brown, M.W. (1998) Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology* 37, 657–676
60. Li, L. *et al.* (1993) The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* 69, 1918–1929
61. Meyer, T. and Rust, N.C. (2018) Single-exposure visual memory judgments are reflected in inferotemporal cortex. *eLife* 7, e32259
62. Grill-Spector, K. *et al.* (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23
63. McMahon, D.B. and Olson, C.R. (2007) Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. *J. Neurophysiol.* 97, 3532–3543
64. Alink, A. *et al.* (2018) Forward models demonstrate that repetition suppression is best modelled by local neural scaling. *Nat. Commun.* 9, 3854
65. Jutras, M.J. *et al.* (2013) Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. *Proc. Natl. Acad. Sci. U. S. A.* 110, 13144–13149
66. Jutras, M.J. and Buffalo, E.A. (2010) Recognition memory signals in the macaque hippocampus. *Proc. Natl. Acad. Sci. U. S. A.* 107, 401–406
67. Stark, S.M. *et al.* (2019) Mnemonic similarity task: a tool for assessing hippocampal integrity. *Trends Cogn. Sci.* 23, 938–951
68. Yassa, M.A. and Stark, C.E. (2011) Pattern separation in the hippocampus. *Trends Neurosci.* 34, 515–525
69. Ferguson, M. *et al.* (2017) Automatic localization of casting defects with convolutional neural networks. In *2017 IEEE International Conference on Big Data (Big Data)*, pp. 1726–1735, IEEE
70. Dworkin, J.D. *et al.* (2020) The extent and drivers of gender imbalance in neuroscience reference lists. *bioRxiv.* Published online January 11, 2020. https://doi.org/10.1101/2020.01.03.894378
71. Maliniak, D. *et al.* (2013) The gender citation gap in international relations. *Int. Organ.* 67, 889–922
72. Caplar, N. *et al.* (2017) Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nat. Astron.* 1, 0141
73. Zhou, D. *et al.* (2020) *Gender Diversity Statement and Code Note-Book v1.0*, Zenodo