Memorability: Reconceptualizing Memory as a Visual Attribute

Wilma A. Bainbridge¹ (ORCID: 0000-0002-7554-0736)

1 - Department of Psychology, University of Chicago, Chicago, IL

Abstract

In the world of visual memory, we often focus our study on the process of memory, but equally important are the inputs to the process—the images we are remembering. A growing body of work has shown that images have a strong power over what we will remember or forget; images have an intrinsic memorability that causes them to be remembered or forgotten across people. In this chapter, I describe our current understanding of memorability as a stimulus property, and its relationship to various aspects of vision and memory. The memorability of an image remains consistent across people, tasks, images, and timing, and shows specific stereotyped patterns in the brain that are separate from those of perception and memory. Recent evidence has proposed that memorability could represent how perceived inputs are prioritized for memory encoding. Although there is currently no comprehensive model of what makes something memorable, deep learning has shown strides in being able to predict and manipulate the memorability of an image. Armed with the memorability scores of an image, one can then create high-powered memory experiments, or develop tests that can more efficiently identify cognitive decline. There are still many open questions about memorability, but a deeper understanding will promise to give us agency over our memories and the images that create them.

Introduction

Memory exists as a word with multiple, entangled definitions. In psychology and neuroscience, "memory" often implies the cognitive process that serves to write and read information stored in the brain. This process is attributed as an abstract entity to individuals, where one can have a "poor memory" or a "good memory", implying they differ in their abilities to perform this process. Meanwhile, a "memory" also denotes the specific event being remembered, an episode with its rich sensory content, emotions, and top-down interpretations. In the study of the "memory", we often focus on the former process-based definition rather than the latter content-based definition. However, understanding the content of our memories themselves—the information represented by the memory system—can serve as a powerful force by which we can understand the greater process. To understand how we remember, we must understand what we are remembering, and why. In other words, we must understand how different contents of memory influence the process of memory—are certain types of information remembered better than others, and why?

In fact, a converging body of work has revealed that certain visual events are more likely to successfully become memories than others (Figure 1: Isola, Xiao, Torralba, & Oliva, 2011a; Bainbridge, Isola, & Oliva, 2013). In other words, visual events can be conceptualized has having an intrinsic memorability, and this has been observed to be consistent across a diverse set of observers, and a wide range of tasks and images (e.g., Bainbridge, 2020). Further, viewing intrinsically memorable images results in stereotyped patterns in the brain that may reveal the computations that occur between perception and memory encoding (Bainbridge, Dilks, & Oliva, 2017; Xie, Bainbridge, Inati, Baker, & Zaghloul, 2020). In this chapter, I will present our current understanding of visual memorability, and how that specifically relates to the processes of vision and memory. First, I will delve into what we know about memorability as an image property, and how that relates to other commonly studied image properties such as brightness or aesthetics. Second, I will discuss how memorability effects interact with other cognitive processes and tasks known to influence memory, such as attention, reward, and depth of processing. With these bases of understanding, I will describe the current hypotheses on the mechanisms behind these memorability effects, and how they relate to vision and memory. Finally, I will present the big new questions about visual memorability that will guide future research.

<< Insert Figure 1 (10_Ch10_fig1.tif) about here >>

Figure 1. A collection of highly memorable (left) and highly forgettable (right) visual stimuli from a range of stimulus categories including artwork, faces, outdoor / indoor scenes, animate / inanimate objects, and dance moves within a video (danced by a person or a stick figure). In spite of many visual and semantic similarities between these two sets of images, observers consistently remember the items on the left, and forget the items on the right.

Memorability as an intrinsic image attribute

One prevailing view of memory is that it is highly idiosyncratic, determined by the sum of our individual experiences and interpretations of those experiences. However, a collection of studies in the last decade have pushed back on this idea—instead demonstrating that a surprising degree of memory behavior is consistent across people. In spite of our different experiences, we tend to remember and forget the same scenes (Isola et al., 2011a), the same faces (Bainbridge et al., 2013), and even the same abstract infographics (Borkin, Vo, Bylinskii, Isola, Sunkavailli, Oliva, et al., 2013) and noise images (Lin, Yousif, Chun, & Scholl, 2021). This implies that images have an intrinsic *memorability*, or a likelihood for successful memory driven by the image itself. While memory is a process performed by the observer across stimuli (what we often call "subject memory"), stimulus memorability is an attribute of the stimulus across observers.

Memorability effects have been repeatedly replicated across a wide range of stimuli and tasks. Most often, memorability is measured using continuous recognition tasks, where participants view a stream of images and identify image repeats that occur long after an image's first presentation. To see whether people are similar in their memory behavior, we conduct consistency analyses (Figure 2), where the participants are split into random halves and then their memory performance for each image is correlated across these halves. These correlations can be incredibly high; for example, the first work on memorability measured a Spearman rank correlation (ρ) of 0.75 for scene images (Isola et al., 2011a), and 0.68 for face images (Bainbridge et al., 2013). This indicates that the rankings for memory performance are similar across individuals, and people tend to remember and forget the same images. These analyses have observed significant consistency regardless of memory performance measure, including hit rate (HR), false alarm rate (FA), d-prime, and corrected recognition (CR) (Isola et al., 2011a; Bainbridge et al., 2013; Bainbridge & Rissman, 2018). This consistency across individuals has also been repeatedly observed in diverse samples taken from the internet, where participants are likely to have had largely non-overlapping experiences with specific face and scene images (Isola et al., 2011a; Bainbridge et al., 2013). Given this high consistency across diverse participants, memorability can thus be considered a relatively stable image attribute (measured by any of the aforementioned memory performance measures). This also implies that memorability can be easily measured by online experiments, and in fact there are open resources that generate the code for experimenters to obtain memorability scores for their own sets of images¹.

<< Insert Figure 2 (10_Ch10_fig2.tif) about here >>

Figure 2. A depiction of the consistency analysis with toy data. Across 1000+ iterations, participants are split into two halves (Group 1 and Group 2), and their memorability

¹ This online tool takes in image URLs and some basic attributes as inputs, and creates the webpage code to run your own memorability experiment (whether on your own server, or using a crowd-sourcing experimental platform like Amazon Mechanical Turk): http://wilmabainbridge.com/makeexperiments.html (Bainbridge, 2017). For additional advice on how to design a memorability experiment, refer to Bainbridge (2021).

scores (HR, FA, d-prime, or CR) are Spearman rank correlated between groups to see how consistent memory performance is between them. This correlation is compared to a correlation with a shuffled Group 2, which approximates chance (no relationship between groups). Plotting a curve of the memory performance for each group averaged across iterations, if the Group 2 line shows a similar trend to Group 1 (versus the Shuffled Group 2 line), then this also implies consistent performance between groups.

The "memorability score" derived from such a study (whether it is HR, FA, d-prime, or CR) is impressively translatable across tasks and other experimental image contexts. Memorability effects emerge if a stimulus is shown for as little as 13ms (Broers, Potter, & Nieuwenstein, 2018), or as long as 10s (Borkin, Bylinskii, Kim, Bainbridge, Yen, Borkin, et al., 2015). Memorability is also highly consistent regardless of the time between image repeats—whether it is 36s (Isola, Xiao, Parikh, Torralba, & Oliva, 2013), or a week (Goetschalckx, Moors, & Wagemans, 2017) between the study and test of an image, memorable images will be remembered better than their forgettable counterparts. While memorability scores are often recorded using a continuous recognition task, they remain stable when measured with other memory paradigms. such as separated study and test phases (Goetschalckx et al., 2017), or when participants have to simultaneously perform perceptual judgments on the images (Broers & Busch, 2020). Even when participants are not aware of a memory task (an incidental memory task), but instead are performing an alternate task such as judging the gender of a face or qualities of a scene, memorability effects still robustly emerge both in behavior (i.e., when tested with a surprise recognition test; Goetschalckx, Moors, & Wagemans, 2019b; Bainbridge, 2020), and in the brain (Bainbridge et al., 2017). Memorability scores also translate seamlessly across participants; for example, scores measured in a large, diverse online sample significantly correlate with those measured in a smaller sample of epileptic patients with intracranial electrodes (Xie et al., 2020). Memorability also seems relatively stable across image contexts (Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015), implying that a memorability score can be recorded in one experiment, but then used in another with other images. It has not yet been explored how extreme manipulations of image context may influence the impact of memorability (e.g., if all images in an experiment are memorable or forgettable). Automatic computer vision algorithms have also accomplished success at quantifying image memorability (Khosla, Raji, Torralba, & Oliva, 2015; Needell & Bainbridge, 2021) as well as manipulating image memorability (Khosla, Bainbridge, Torralba, & Oliva, 2013; Goetschalckx, Andonian, Oliva, & Isola, 2019a), suggesting that it is an imagespecific property that can be measured and changed. In fact, when face images were automatically manipulated to be memorable or forgettable and then retested in a new face memory experiment, participants significantly remembered a memorable version over the forgettable version of that same face (Khosla et al., 2013).

With this combined evidence across studies, it is evident that memorability is a stimulus property that is relatively immutable across contexts, tasks, and observers. This makes memorability incredibly easy to work with, as it can be treated, measured, and manipulated like any image attribute. However, this also opens up big questions about

what memorability may represent, and how it relates to the space of all other image attributes.

Memorability as compared to other image attributes

In thinking about memorability, it is easy to get stuck on how it is measured. If memorability is defined by memory performance from a group of people, how is it a stimulus property rather than a subjective process? To understand memorability as an image attribute, one can compare its concept to a similar other image attribute: facial attractiveness. While a person can be attracted to another individual as a process, that face can also be ascribed an intrinsic "attractiveness" rating defined by a group. Attractiveness ratings for a given face are generally consistent across observers, even across cultures (with correlations across racial groups as high as r = 0.93; Cunningham, Roberts, Barbee, Druen, & Wu, 1995) and in infants (Langlois, Ritter, Roggman, & Vaughn, 1991), but any given individual may deviate in how attractive they find a given face based on individual experience (e.g., if a face resembles a favorite actor). Thus, the final perception of a face is likely to be determined by a combination of the intrinsic, group-defined attractiveness of the face, and idiosyncratic influences from the observer. While "attractiveness" can be observed, quantified, and manipulated as a singular factor, it is defined by a combination of various factors at different levels of abstraction, including low-level visual features (e.g., symmetry, Perrett, Burt, Penton-Voak, Lee, Rowland & Edwards, 1999; skin uniformity, Fink, Grammer, & Thornhill, 2001), and high-level attributes (e.g., emotional expression, Goelle, Mast, Lobmaier, 2014; cultural differences, Cunningham et al., 1995). Memorability resembles attractiveness with respect to many of these characteristics as a complex image attribute. In the same way, we can remember a face (as a process), but faces (and other images) also have an intrinsic memorability across observers. While any individual may deviate in what they remember (e.g., if a face resembles a family member, they may remember it better), this intrinsic memorability accounts for as much as 50% of the variance in what one will ultimately remember (Bainbridge et al., 2013). Memorability also may be driven by both low-level and high-level features, but many of them are still relatively unclear; thus far a comprehensive, fully predictive model of memorability has yet to be discovered.

What we do know so far is what memorability *is not*. A handful of studies have shown that memorability cannot be accounted for by another singular attribute. For scene images, photographs of indoor scenes containing people tend to be more memorable than expansive, landscape images (Isola et al., 2015). Therefore, images with less blue and green tend to be more memorable (Isola et al., 2013), as well as images that are brighter and have higher contrast (Dubey, Peterson, Khosla, Yang, & Ghanem, 2015). However, these low-level features show weak correlations with memorability, and memorability effects emerge when these low-level features are controlled (Bainbridge et al., 2017). Similarly, visual saliency of an image shows correlations with memorability when looking at photographs of individual objects, but not more complex scenes (Dubey et al., 2015). In the realm of faces, low-level features show almost no relation to face memorability, including color, edges, and the geometric positions of individual face features that are often used to measure distance from an average face (Bainbridge,

2019). Thus, it is generally agreed that memorability does not just reflect heightened visual saliency, nor a difference from other images in low-level visual features.

At the same time, memorability is also not synonymous with other high-level features, like aesthetics, emotionality, or object content (Isola et al., 2013). In fact, memorability is so surprisingly opaque that we as the creators of our memories cannot predict what we will remember: when participants rated which images they thought they would remember or forget, this rating showed no correlation with what was actually memorable or forgettable (Isola et al., 2011b; Bainbridge, 2017). Unlike attractiveness, memorability is also not a straightforward combination of several attributes. A study modeling face memorability using a regularized multiple regression model (Lasso) of attributes from across the face and memory literature (including confidence, trustworthiness, attractiveness, typicality, and familiarity) found this model could only explain 46.6% of the variance of memorability. The other half of the variance was highly consistent across people for predicting memory behavior, but could not be explained by these attributes. Another study using a comprehensive set of 49 dimensions to characterize our perceptual space for objects (Hebart, Dickter, Kidder, Kwok, Corriveau, Van Wicklin, & Baker, 2019) found that these dimensions captured less than 35% of the variance of memorability (Kramer, Hebart, Baker, & Bainbridge, 2021). Similarly, no parsimonious combination of attributes for predicting scene memorability has been discovered.

In recent years, deep learning neural networks (DNNs) have been proposed as impressively predictive models of the human visual system (Yamins, Hong, Cadieu, Solomon, Seibert, & DiCarlo, 2014). These models comprise of a series of layers with "neurons" that transform inputs from a prior layer's neurons into a thresholded output. After training a model with thousands of images, it learns weights to attribute to each neuronal unit, and then can make predictions on images outside of its training set. This method has been successful with a range of tasks such as prediction of object and scene categories from an image (Krizhevsky Sutskever, & Hinton, 2012; Zhou, Lapedriza, Xiao, Torralba, & Oliva, 2014), and impressive correspondences have been observed between representations in human early visual cortex (V1, V2, V3) and early layers in these networks (Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016). Some work has created models that can predict image memorability at the level of human consistency, with correlations of $\rho = 0.68$ between predictions and human memory behavior (Khosla, Raji, Torralba, & Oliva, 2015; Needell & Bainbridge, 2021), suggesting that memorability may be image-computable. Recent developments in deep learning such as recurrent connections, residual networks (Needell & Bainbridge, 2021), and adversarial networks to improve learning (Goetschalckx et al., 2019a) may also improve the biological plausibility of these models. However, recent work has put into question the similarity of DNNs to human late visual cortex (e.g., the inferotemporal cortex) more broadly (Xu & Vaziri-Pashkam, 2021), and so the use of deep learning for understanding the human brain is still an ongoing debate. If memorability is indeed predictable from an image's pixels, this may suggest that memorability is reflecting a sensitivity to the visual statistics of the world that doesn't rely on specific memories or experiences. A next step would be to examine how predictive ability differs based on image input set—to what degree does one need to experience the natural image statistics present on Earth to develop these consistent memory biases?

Memorability and other cognitive processes

Attention is another cognitive process that is known to influence the successful encoding of specific images. For example, when one is in a receptive attentional state, they tend to remember an item better than when they are inattentive (deBettencourt, Norman, & Turk-Browne, 2018). However, when combined into a comprehensive model, one's current attentional state and memorability have independent influences on memory encoding, where both memorable items and highly attentive trials are separately predictive of successful memory (with memorability having a slightly stronger influence; Wakeland-Hart, deBettencourt, Bainbridge, & Rosenberg, 2021). Memorable images also do not cause heightened attention; memorable images do not cause more spatial orienting than forgettable images, and memorable distractors do not capture attention during visual search (Bainbridge, 2020). Similarly, tasks manipulating topdown attention on images do not seem to influence effects of memorability. When participants are asked to deeply encode images (judging the honesty of a face image), in a surprise memory test they remember memorable faces better than forgettable faces. However, this difference also emerges when participants are made to encode images shallowly, by just judging the color of a fixation cross shown on top of a face (so the faces did not even need to be encoded). Thus, regardless of how deeply one encodes an image, memorable images will always be remembered better than forgettable ones. These effects are replicated even when participants are motivated with a reward – participants remember memorable items they are incentivized to forget better than forgettable images they are incentivized to remember (Bainbridge, 2020). Thus, reward also does not appear to exhibit an influence on these memorability effects. At the same time, a growing body of work has shown that fixation patterns on an image are predictive of the memorability of that image (Bylinskii et al., 2015; Lyu, Choe, Kardan, Kotabe, Henderson, & Berman, 2020), demonstrating that the parts of an image that drive eye movement may be those that help that image be remembered.

Given that memorability is a perceptual process of an image reflecting memory, it is also important to consider how memorability may relate to perception and memory. A handful of neuroimaging experiments utilizing spatially resolved functional magnetic resonance imaging (fMRI) and temporally resolved magnetoencephalography (MEG) have examined the patterns of activation for viewing memorable versus forgettable images. Importantly, these studies have examined differences between face and scene image sets that only differ in memorability, but are matched for both low-level visual features (e.g., color, spatial frequency, number of objects), and high-level features (e.g., emotionality, aesthetics, category). These experiments observe that sensitivity to memorability emerges in the inferotemporal cortex (IT) implicated for late perceptual processing as well as the medial temporal lobe (MTL) and hippocampus implicated for memory (Bainbridge et al., 2017). However, reflecting the lack of low-level visual differences in the stimuli, no patterns emerge in the primary visual cortex. The timing of memorability sensitivity occurs at the same time scale as other high-level visual features, with differences emerging at 150-400ms after stimulus onset (Mohsenzadeh, Mullin, Oliva, & Pantazis, 2019).

These patterns of memorability observed in the brain are surprisingly distinct from other memory processes previously observed in the brain. Voxels sensitive to memory encoding are usually identified using a subsequent memory task, where encoding trials are back-sorted based on whether they were ultimately remembered versus forgotten (Brewer et al., 1998; Wagner et al., 1998). Usually, parietal cortex and frontal areas emerge with heightened activation for the images that will later be remembered (Kim, 2011). Thus, it is possible that patterns of memorability could reflect the patterns for items that are to be later encoded. However, when memorability scores were collected (from a separate set of participants) for images from an fMRI subsequent memory task, topographically separate patterns of activation emerged for sensitivity to memorability versus memory encoding (Bainbridge & Rissman, 2018), with ventral regions implicated in memorability sensitivity, and parietal regions implicated in memory encoding. Furthermore, memorability activation patterns remain strong and unaffected when only analyzing images that were later forgotten by an individual participant (Bainbridge et al., 2017). Thus, regardless of one's memory, their brain is still sensitive to the intrinsic memorability of that image. This sensitivity to memorability also emerges earlier than the time course of memory encoding, which is generally thought to occur after 300 ms (Friedman, 1990; Friedman & Johnson, 2000; Khaligh-Razavi, Bainbridge, Pantazis, & Oliva, 2016).

These results suggest that a sensitivity to image memorability may reflect the existence of a specific process or computation occurring in the brain that supports the transition from perception to successful memory coding. Yet at the same time, this computation is separate from early visual processing of the features in the image, and separate from memory encoding both in terms of the regions involved and the timing of these processes. What might this "memorability computation" reflect, and what are the mechanisms that underlie this shared memory behavior across individuals for images?

Mechanisms of memorability

The current leading idea is that memorability may be an emergent property of an image that reflects how images are prioritized by the brain. Essentially, as we are unable to encode all visual inputs into memory, our brains must select information to encode, and memorability may be quantifying the underlying sorting process by the brain. In a study recording neuronal firing in the anterior temporal lobes (ATL) of epileptic patients, Xie and colleagues (2020) observed that memorability of a word was predicted by its semantic connectedness with other words. If one conceptualizes our representations of words as an interconnected network where network edges represent semantic relatedness, then words with more connections (roots of the network) tended to be more memorable. In contrast, words with few connections (the leaves of this network) tended to be more forgettable. These results were not accounted for by the concreteness of the word or its frequency in language. These memorable, highly connected words were reinstated faster in the ATL during a cued recall task, and when participants made errors, they falsely recalled more memorable words. These results suggest that during retrieval, our brain begins its memory search at these high-priority words positioned at

roots in a semantic network, and then traverses this network to find the item to be recalled (thus why it takes longer to reinstate the forgettable leaves of the network).

Thus, memorability may reflect how we represent perceptual information and sort and search it for encoding and retrieval. Evidence for these patterns has also been observed for face and scene images; when observing the representational geometry (i.e., neuronal pattern similarity) in IT and MTL for images, memorable items exist at the centers of this representational space, while forgettable item representations are on the outskirts (Bainbridge et al., 2017; Bainbridge and Rissman, 2018). Specifically, memorable images show highly similar neuronal patterns, while forgettable images show wider pattern variation. However, other studies have also observed the opposite, where memorable image representations are more distributed during MEG recording across the whole brain (Mohsenzadeh et al., 2019), and during recording from rhesus macaque IT (Jaegle, Mehrpour, Mohsenzadeh, Meyer, Oliva, & Rust, 2019). These results have been used to support the hypothesis that IT neurons may code for memorability through magnitude variation (Rust & Mehrpour, 2020). These different patterns could in fact be measuring different aspects of this process - where IT and visual areas may first capture the distinctiveness of memorable stimuli, and then downstream, the MTL and ATL prioritize these memorable items and thus represent them in a similar way to each other. This transformation across regions and times is still an open question that will require well-controlled stimuli and tasks, and simultaneously time-resolved and spatially-resolved methods to answer.

There is also the question of how this neural sensitivity to memorability occurs in the first place, which may lend insight into this question of shifting representations. In the behavioral realm, memorability has traditionally been related to stimulus distinctiveness or atypicality (e.g., Light, Kayra-Stuart & Hollander, 1979; Bartlett, Hurry, & Thorley, 1984), and images that were more sparsely distributed and distinctive in a DNN feature space tended to be more memorable (Lukavský & Děchtěrenko, 2017). However, recent work has found that the relationship between memorability and typicality may be subtler; in fact, scene images that are dissimilar from others in early DNN features (akin to lowlevel visual features) but similar in late DNN features (conceptual, semantic information) may be the most memorable (Chapter 12; Koch, Akpan & Coutanche, 2020). As mentioned previously, words that are more semantically connected (rather than distinctive) are also more memorable (Xie et al., 2020). To elaborate on the question of memorability as it relates to typicality, an extensive study was run quantifying memorability with close to 14,000 participants, for the 26,107 images in the THINGS database (Hebart et al., 2019), capturing the spread of 1,854 diverse object concepts in this world (Kramer et al., 2021). Surprisingly, it was found that the relationship of memorability to typicality and sparseness varied by object category – for some categories, images of the most typical objects were the most memorable (e.g., weapons), while for others, the most atypical objects were most memorable (e.g., kitchen appliances). This implies that memorability may not be merely stimulus atypicality or distinctiveness. Zooming into the different dimensions underlying the object category space, however, conceptual dimensions showed a stronger relationship to memorability than perceptual dimensions (Kramer et al., 2021). This finding is echoed in work within deep learning showing that memorability predictions improve when

deeper conceptual information is incorporated versus shallow perceptual information (Needell & Bainbridge, 2021). However, memorability effects also occur for stimuli stripped of semantic information (Lin et al., 2021), and in monkeys who presumably lack conceptual representations for novel object photographs (Jaegle et al., 2019). Thus, memorability does not appear to be solidly predicted by stimulus typicality nor atypicality, and the role of perceptual versus conceptual dimensions is still unclear. The computations that determine this highly consistent attribute of memorability are still an open question of active interest.

Implications of memorability

While a full mechanistic explanation of memorability is still under exploration, our current understanding of memorability already motivates several applications to other topics in the field, as well as to the real world.

Just as thinking about memorability is important for the understanding of the general process of memory, memorability can also help us understand disruptions of memory. A stimulus-based perspective has revealed that certain images are better able to predict whether an individual is suffering from signs of early dementia (Bainbridge, Berron, Schütze, Cardenas-Blanco, Metzger, Dobisch, et al., 2019), and these images show stronger correlations to biomarkers measured in the brain (Bai, Schütze, Jessen, Spottke, Nestor, Bürger, et al., 2021). Specifically, images that are highly memorable to healthy controls but highly forgettable to those with subjective cognitive decline appear to be highly consistent and predictive of cognitive impairment across cross-validated samples. This has led to a push in the field of Alzheimer's Disease (AD) research to consider not only the biomarkers and anatomical shifts that lead to development of AD, but also how transformations in representational content and stimulus-based behaviors may reveal impairments early on (Grande, Berron, Maass, Bainbridge, & Düzel, 2021). Considering memorability may also lend insight into the impairments and underlying neural differences in those with disrupted perceptual representations (e.g., prosopagnosia for faces) as well as those with other disruptions in memory (e.g., aphantasia – an inability to perform visual imagery or recall; Bainbridge, Pounder, Eardley, & Baker, 2021).

Memorability is also essential to consider in the design of experiments of vision and memory. Memorability effects occur in the brain even during a purely perceptual task (Bainbridge et al., 2017; Bainbridge, 2020), and can be pulled out from old brain data testing other questions (Bainbridge & Rissman, 2018). So as long as participants are perceiving a stimulus, that stimulus always has a memorability that may influence participant performance. Memorability as a potential stimulus confound is particularly important to consider for tasks that use different image categories to study memory (such as celebrities versus non-celebrities in familiarity tests, who may differ in their memorability), or subsequent memory studies that post-hoc contrast images that were remembered versus forgotten (as it could be stimulus memorability differentiating these two image sets, rather than differences in encoding processes). This is relatively simple to avoid, by either controlling for memorability of stimuli when designing a study, or testing to see whether memorability can account for variance in the final results. One

can even leverage the power of memorability by selecting stimuli in advance meant to drive strong differences in memory, such as by only testing highly memorable and highly forgettable images. Memorability can be measured in a relatively straightforward manner by collecting ground truth scores online (Bainbridge, 2017), or at estimating them using a DNN (Needell & Bainbridge, 2021)². There are also many pre-existing datasets that have memorability scores for thousands of face images (Bainbridge et al., 2013), scene images (Isola et al., 2011a), and object images (Goetschalckx & Wagemans, 2019c) that could serve as effective stimulus sets.

One important note is that this chapter has focused on memorability for images, with some discussion of memorability for words. However, there is also work to suggest that these memorability effects occur across a range of stimuli. Memorability has been shown to be intrinsic to face identities, across images of different expressions and viewpoints of the same person (Bainbridge, 2017). Memorability is also intrinsic to movement, where a specific dance movement is memorable across different dancers, including a simple stick-figure dancer (Ongchoco, Chun, & Bainbridge, submitted). Similarly, intrinsic memorability has also been observed in short video clips (Newman, Fosco, Casser, Lee, McNamara, & Oliva, 2020) and popular movie scenes (Cohendet, Yadati, Duong, & Demarty, 2018). Memorability effects have also been observed for concrete nouns, which are highly familiar, conceptual stimuli with limited specific visual information (Xie et al., 2020). Important next questions will be the degree to which memorability also operates in other sensory domains, such as audition and olfaction.

While the current chapter focuses on accurate memories, we have also repeatedly observed that false alarms are consistent across observers (Bainbridge et al., 2013; Bainbridge, 2017), implying that images may not only have a "memorability", but also a "false-memorability". In other words, some images cause strong false memories, while others cause very accurate memories. This effect has also been observed in pop culture, with certain images that elicit pervasive false memories across the majority of the population (dubbed the Visual Mandela Effect, VME; Prasad & Bainbridge, 2021); for example, Pikachu, the famous character from the Pokémon franchise, does not have any black on his tail, even though people will falsely report a black tip on his tail, and consistently select such an image in a lineup. These effects are particularly interesting because participants act consistently for an image they have never seen (they have likely never observed an incorrect Pikachu image with a black tip on his tail, but report this specific false memory). Further research will be needed to understand the underlying causes of the VME and compare its mechanisms to memorability for accurate information.

Finally, along with the automatic estimation of image memorability with DNNs, an emerging avenue is also the manipulation of memorability using computational models (Khosla et al., 2013; Bylinskii, Goetschalckx, Newman, & Oliva, 2021), such as generative adversarial networks, which are able to modify images in a highly naturalistic

_

² To make DNN-based predictions of memorability easily accessible to the field, we have made a publicly available web app and python package that utilizes ResMem—the current state-of-the-art network in memorability prediction—to give back a memorability score for any image you upload (https://brainbridgelab.uchicago.edu/resmem/; Needell & Bainbridge, 2021).

way (Goetschalckx et al., 2019a). These advances can allow for memorability to enter the real world and the hands of the average person, serving as camera filters to boost the memorability of one's photographs. This can have incredibly positive implications for society: by allowing us to create more memorable educational materials, memorable therapeutic environments for those with memory loss, or to help design memory aids (and even forgetting aids to those with post-traumatic stress disorder). However, there are also looming ethical questions to consider if we are able to easily exert control over people's memories through the images we use. For example, companies could make their advertisements or mascots unforgettable, or political or social influencers could also make themselves stick out in our memories. These ethical issues will be important to consider as we elucidate these mechanisms of memorability and develop increasingly sophisticated algorithms for its manipulation.

Conclusion

While the field has long had a focus on memory as a process, a growing body of work is now shifting focus to also examine the items we remember, and how certain items are particularly successful at being encoded into memory. In this chapter, I covered the stimulus-focused world of memorability, where visual events have an intrinsic likelihood for memory across people. Memorability has proven to be a particularly interesting image attribute, separate from other well-studied attributes (e.g., color, aesthetics, subjective memorability), and sensitivity to it in the brain has been separated from processes such as attention, early vision, and memory encoding. A deeper dive into this sensitivity has shown a potential for memorability to reveal the computations occurring between perception and memory encoding—perhaps reflecting how we prioritize perceptual information, based on its location in our larger representational spaces of the visual and semantic world. A greater understanding of memorability promises to cause a cascade of impacts on our understanding of memory disorders, false memories, and the relationship of other sensory modalities to memory. It also presents some looming ethical considerations as computer vision research able to measure and manipulate image memorability becomes more advanced. In sum, it is essential to consider not only how we are remembering the world, but also what we are remembering, and why.

References

Bai, Y., Schütze, H., Jessen, F., Spottke, A., Nestor, P., Bürger, K., et al. (2021). Memorability analysis for diagnostic photographs in cognitive assessment: Linking behavioral performance with biomarker status. *Alzheimer's Association International Conference 2021*, Denver, CO USA.

Bainbridge, W.A. (2017). The memorability of people: Intrinsic memorability across transformations of a person's face. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 706-716.

Bainbridge, W.A. (2019). Memorability: How what we see influences what we remember. *Psychology of Learning and Motivation*. (Eds: K. Federmeier & D. Beck). Elsevier, Cambridge, MA. Pgs. 1-27.

Bainbridge, W.A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia*, *141*, 107408.

Bainbridge, W.A. (2021). Shared memories driven by the intrinsic memorability of items. *Human Perception of Visual Information: Psychological and Computational Perspectives.* (Eds: B. Ionescu, W.A. Bainbridge, N. Murray, P. Le Callet). Springer, New York City, NY.

Bainbridge, W.A., Berron, D., Schütze, H., Cardenas-Blanco, A., Metzger, C., Dobisch, L., et al. (2019). Memorability of photographs in subjective cognitive decline and mild cognitive impairment for cognitive assessment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 11, 610-618.

Bainbridge, W.A., Dilks, D.D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage, 149*, 141-152.

Bainbridge, W.A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142*, 1323-1334.

Bainbridge, W.A., Pounder, Z., Eardley, A.F., & Baker, C.I. (2021). Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. *Cortex, 135,* 159-172.

Bainbridge, W.A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Scientific Reports*, *8*, 8679.

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12*, 219-228.

Borkin, M. A., Bylinskii, Z., Kim, N. W., Bainbridge, C. M., Yen C. S., Borkin, D., Pfister, H., & Oliva, A. (2015). Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, *22*, 519-528.

Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavailli, S., Oliva, A., et al. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, *19*, 2306-2315.

- Brewer, J.B., Zhao, Z., Desmond, J.E., Glover, G.H., & Gabrieli, J.D.E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, 281, 1185-1187
- Broers, N., & Busch, N.A. (2020). The effect of intrinsic image memorability on recollection and familiarity. *Memory & Cognition*, 1-21.
- Broers, N., Potter, M.C., & Nieuwenstein, M. R. (2018). Enhanced recognition of memorable pictures in ultra-fast RSVP. *Psychonomic Bulletin & Review, 25*, 1080-1086.
- Bylinskii, Z., Goetschalckx, L., Newman, A., & Oliva, A. (2021). Memorability: An image-computable measure of information utility. *Human Perception of Visual Information: Psychological and Computational Perspectives.* (Eds: B. Ionescu, W.A. Bainbridge, N. Murray, P. Le Callet). Springer, New York City, NY.
- Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*, 165-178.
- Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*, 1-13.
- Cohendet, R., Yadati, K., Duong, N.Q., & Demarty, C.H. (2018). Annotating, understanding, and predicting long-term video memorability. *Proceedings of the 2018 ACM International Conference on Multimedia Retrieval*, 178-186.
- Cunningham, M.R., Roberts, A.R., Barbee, A.P., Druen, P.B., & Wu, C.-H. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, 68, 261-279.
- deBettencourt, M.T., Norman, K.A., & Turk-Browne, N.B. (2018). Forgetting from lapses of sustained attention. *Psychonomic Bulletin & Review, 25,* 605-611.
- Dubey, R., Peterson, J., Khosla, A., Yang, M.H., & Ghanem, B. (2015). What makes an object memorable? *Proceedings of the IEEE International Conference on Computer vision*, 1089-1097.
- Fink, B., Grammer, K., & Thornhill, R. (2001). Human (Homo sapiens) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology*, 115, 92.
- Friedman, D. (1990). ERPs during continuous recognition memory for words. *Biological Psychology*, *30*, 61-87.
- Friedman, D., & Johnson, Jr., R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research & Technique*, *51*, 6-28.
- Goetschalckx, L., Andonian, A., Oliva, A., & Isola, P. (2019a). Ganalyze: Towards visual definitions of cognitive image properties. *Proceedings of the IEEE International Conference on Computer Vision*, 5744-5753.

- Goetschalckx, L., Moors, P., & Wagemans, J. (2017). Image memorability across longer time intervals. *Memory*, 26, 581-588.
- Goetschalckx, L., Moors, P., Wagemans, J. (2019b). Incidental image memorability. *Memory, 27,* 1273-1282.
- Goetschalckx, L., & Wagemans, J. (2019c). MemCat: A new category-based image set quantified on memorability. *PeerJ*, 7, e8169.
- Golle, J., Mast, F.W., & Lobmaier, J.S. (2014). Something to smile about: The interrelationship between attractiveness and emotional expression. *Cognition and Emotion*, *28*, 298-310.
- Grande, X., Berron, D., Maass, A., Bainbridge, W.A., & Düzel, E. (2021). Content-specific vulnerability of recent episodic memories in Alzheimer's disease. *Neuropsychologia*, *160*, 107976.
- Hebart, M.N., Dickter, A.H., Kidder, A., Kwok, W.Y., Corriveau, A., Van Wicklin, C., & Baker, C.I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLoS ONE*, *14*, e0223792.
- Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2013). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Learning*, *36*, 1469-1482.
- Isola, P., Xiao, J. X., Torralba, A., & Oliva, A. (2011a). What makes an image memorable? *24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 145-152.
- Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011b). Understanding the intrinsic memorability of images. *Advances in Neural Information Processing Systems*, 1-9.
- Jaegle, A., Mehrpour, V., Mohsenzadeh, Y., Meyer, T., Oliva, A. & Rust, A. (2019). Population response magnitude variation in inferotemporal cortex predicts image memorability. *eLife*, *9*, e47596.
- Khaligh-Razavi, S-M., Bainbridge, W.A., Pantazis, D., & Oliva, A. (2016). From what we perceive to what we remember: Characterizing representational dynamics of visual memorability. *bioRxiv*: https://doi.org/10.1101/049700.
- Khosla, A., Bainbridge, W.A., Torralba, A., & Oliva, A. (2013). Modifying the memorability of face photographs. *Proceedings of the International Conference on Computer Vision (ICCV)*.
- Khosla, A., Raji, A.S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. *International Conference on Computer Vision (ICCV)*, 2390-2398.
- Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: a metaanalysis of 74 fMRI studies. *Neurolmage*, 54, 2446-2461.

- Koch, G. H., Akpan, E., & Coutanche, M. N. (2020). Image memorability is predicted by discriminability and similarity in different stages of a convolutional neural network. *Learning & Memory, 27,* 503-509.
- Kramer, M.A., Hebart, M.N., Baker, C.I., & Bainbridge, W.A. (2021). Revealing the relative contributions of conceptual and perceptual information to visual memorability. *Vision Sciences Society Annual Meeting*, Online.
- Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*.
- Langlois, J.H., Ritter, J.M., Roggman, L.A., & Vaughn, L.S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology*, 27, 79.
- Light, L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 212-228.
- Lin, Q., Yousif, S., Chun, M.M., & Scholl, B.J. (2021). Visual memorability in the absence of semantic content. *Cognition*, *212*, 104714.
- Lukavský, J., & Děchtěrenko, F. (2017). Visual properties and memorizing scenes: Effects of image-space sparseness and uniformity. *Attention, Perception, & Psychophysics, 79*, 2044-2054.
- Lyu, M., Choe, K. W., Kardan, O., Kotabe, H. P., Henderson, J.M., & Berman, M.C. (2020). Overt attentional correlations of memorability of scene images and their relationships to scene semantics *Journal of Vision*, *20*, 1-17.
- Mohsenzadeh, Y., Mullin, C., Oliva, A., & Pantazis, D. (2019). The perceptual neural trace of memorable unseen scenes. *Scientific Reports*, *9*, 1-10.
- Needell, C.D., & Bainbridge, W.A. (2021). Embracing new techniques in deep learning for predicting image memorability. *arXiv*, 2105.10598.
- Newman, A., Fosco, C., Casser, V., Lee, A., McNamara, B., & Oliva, A. (2020). Multimodal memorability: Modeling effects of semantics and decay on video memorability. *16th European Conference on Computer Vision*, *16*, 223-240.
- Ongchoco, J. D. K., Chun, M., & Bainbridge, W. A. (submitted). What moves us? The intrinsic memorability of dance.
- Perrett, D.I., Burt, D.M., Penton-Voak, I.S., Lee, K.J., Rowland, D.A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior, 20,* 295-307.
- Prasad, D., & Bainbridge, W.A. (2021). The Visual Mandela Effect: Evidence for specific shared false memories in popular iconography. *psyArXiv*, 10.31234/osf.io/nzh3s.
- Rust, N.C., & Mehrpour, V. (2020). Understanding image memorability. *Trends in Cognitive Sciences*, *24*, 557-568.

- Wagner, A.D., Schacter, D.L., Rotte, M., Koutstaal, W., Maril, A., Dale, et al. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. Science, 281, 1188-1191.
- Wakeland-Hart, C.D., Cao, S. deBettencourt, M.T., Bainbridge, W.A., & Rosenberg, M.D. (2021). Predicting visual memory across images and within individuals. *PsyArXiv*, 10.31234/osf.io/zbu3k.
- Xie, W., Bainbridge, W.A., Inati, S.K., Baker, C.I., & Zaghloul, K. (2020). Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. *Nature Human Behaviour, 4*, 937-948.
- Xu, Y., & Vaziri-Pashkam, M. (2021). Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature Communications*, *12*, 2065.
- Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., & DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*, 8619-8624.
- Zhou, B., Lapedriza, A., Xiao, J., Torralba, & Oliva, A. (2014). Learning deep features for scene recognition using Places database. *Advances in Neural Information Processing Systems*, 27.