The Intrinsic Memorability of Face Photographs

Wilma A. Bainbridge (1), Phillip Isola (1), & Aude Oliva (2)

1 – Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

2 – Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology

Corresponding authors:

Wilma Bainbridge (wilma@mit.edu) or Aude Oliva (oliva@mit.edu)

Massachusetts Institute of Technology, 32-D430

77 Massachusetts Avenue

Cambridge, MA 02139

Keywords: perception, memory, memorability, face photograph database, facial attributes, social traits, personality traits.

Abstract

The faces we encounter throughout our lives make different impressions on us: some are remembered at first glance, while others are forgotten. Previous work has found that the distinctiveness of a face influences its *memorability* – the degree to which faces images are remembered or forgotten. Here, we generalize the concept of "face memorability" in a large-scale memory study. First, we find that memorability is an intrinsic feature of a face photograph – across observers some faces are consistently more remembered or forgotten than others – indicating that memorability can be used for measuring, predicting, and manipulating subsequent memories. Second, we determine the role that twenty personality, social, and memory-related traits play in face memorability. Whereas we find that certain traits (such as kindness, atypicality, and trustworthiness) contribute to face memorability, they do not suffice to explain the variance in memorability scores, even when accounting for noise and differences in subjective experience. This suggests that memorability itself is a consistent, singular measure of a face that cannot be reduced to a simple combination of personality and social facial attributes. We outline modern neuroscience questions that can be explored through the lens of memorability, and we release with this paper the 10k US Adult Faces Database with memorability and attribute scores.

Introduction

Every day, we encounter new faces – on social networks, in the media, and in person. While we may only glance at them once, some faces will stick in our minds, while others will fade. These faces are differentially memorable or forgettable – not all will be equally remembered after only a single exposure. Whereas previous research has shown individual variability in human memory and the importance of subjective experience on face memory (Chiroro & Valentine, 1995; Duchaine, Yovel, Butterworth, & Nakayama, 2006), little work has characterized whether faces have intrinsic and systematic features that would make some more memorable or forgettable to everyone.

Recent large-scale visual memory studies have shown that people have a remarkable ability at remembering specific details of images (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle, Brady, Alvarez, Oliva, 2010a, 2010b; Vogt & Magnussen, 2007), with some images being consistently more memorable or forgettable than others (Isola, Xiao, Torralba, & Oliva, 2011a). This work shows that images have an intrinsic *memorability* level that is independent of the observers' past experiences and reproducible across a population. Thus, the degree of memorability of an image allows one to predict, *from encoding*, if an individual image is more likely to later be remembered or forgotten. However, it is not intuitive how intrinsic image memorability may apply to faces, which have little perceptual variation and depend largely on personal experience.

The Intrinsic Memorability of Face Images

Is memorability an intrinsic feature of face photographs? Is there a component of face memorability independent of individuals' personal context, familiarity, and experience? If there is, then memorability could be used as a singular attribute with which to analyze, predict, and manipulate pictures of faces. Memorability lends itself to several useful applications including examining different learning methods for forgettable versus memorable faces, or creating new metrics to assess memory performance using a graded scale based on item memorability.

Can memorability be explained by other known attributes from personality, social, and memory research? It is known that a *distinctive* or *atypical* face (i.e., a face distant from an average or prototypical face) is more likely to be remembered, while a face that looks *familiar* is more likely to create false memories (Bartlett, Hurry, & Thorley, 1984; Bruce, Burton, & Dench, 1994; Busey, 2001; Valentine, 1991; Vokey & Read, 1992). However, it is not known whether these traits of distinctiveness and familiarity make up the whole story of memorability, and what additional facial attributes might contribute to making a face memorable or forgettable after a single glance. As with images of objects and places (Isola, Parikh, Torralba & Oliva, 2011b), is there a basis of facial traits attributed to an individual face that will make it more memorable or forgettable? Several works have proposed dimensions along which to evaluate faces, such as trustworthiness and dominance (Oosterhof & Todorov, 2008; Todorov, 2011), warmth and competence (Fiske, Cuddy, & Glick, 2007), and goodness, hardness, and activeness (Rosenberg, Nelson, & Vivekananthan, 1968). However, the relationship of a multidimensional trait space to memorability has not been evaluated thoroughly and at a large scale. So far, work linking higher-level traits to memory have only examined a few preselected traits with memorability, such as untrustworthiness (Rule, Slepian, & Ambady, 2012), or own-age and own-race biases (Anastasi & Rhodes, 2005; Chiroro & Valentine, 1995; Meissner & Brigham, 2001).

The Neuroscience of Memorability

Memorability as an intrinsic attribute of faces would open new endeavors in the neuroscience of memory. The pioneering work of Scoville and Milner (1957) pointed to the hippocampus in the medial temporal lobe (MTL) as a key region implicated in long-term memory. Since then, other regions such as the amygdala (Cahill, Babinsky, Markowitsch, & McGaugh, 1995; Kleinhans, Johnson, Mahurin, Richards, Stegbauer, Greenson, Dawson, & Aylward, 2007) and perirhinal cortex (Wan, Aggleton, & Brown, 1999) have been identified as integral parts of the MTL memory system. Memorability as an intrinsic feature of an image would contribute to three outstanding questions in the neuroscience of memory: 1) What is the relationship between perception and memory?, 2) How do the relationships between high-level facial attributes and memorability play out in the brain?, and 3) To what extent are MTL sub-structures selective for particular stimuli (e.g., faces, scenes, objects)?

If face images do indeed have an intrinsic memorability, is face memorability *perceptual* in nature and how is it related to the encoding stage of memory (Shrager, Kirwan, & Squire, 2008)? In the neuroscience literature, memory and perception have often been studied separately, with each domain linked to different experimental paradigms (Bussey & Saksida, 2007), and associated with separate key cortical regions, such as the ventral visual stream for perception (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999) and the medial temporal lobe (MTL) for memory (Squire & Zola-Morgan, 1991). Only recently has the MTL been studied outside of an encapsulated box of memory, with studies looking at these regions in complex perceptual tasks, including the hippocampus for spatial processing and the perirhinal cortex for object processing (Bonnici, Kumaran, Chadwick, Weiskopf, Hassabis, & Maguire, 2012; Buckley & Gaffan, 2006; Lee, Yeung, & Barense, 2012; Murray & Richmond, 2001). Similarly, other work has looked at how activity in the MTL during the encoding phase affects later memory task performance

(Knight, 1996; LaRocque, Smith, Carr, Witthoft, Grill-Spector, & Wagner, 2013; Shrager et al., 2008). Knowing the memorability score of an image allows predictions on the encoding strength of a specific image when it is first perceived. This encoding strength can be generalized and reproduced in different people (Qin, van Marle, Hermans, & Fernández, 2011), and as such could allow one to compare observer-effects versus item-effects generalizable across observers.

With knowledge on which social, personality, and emotional traits are correlated with memorability, one can explore relationships between regions implicated in traits processing and MTL structures. For example, the own-race effect on face memory (Chiroro & Valentine, 1995; Meissner & Brigham, 2001) appears to cause differential activity in not only the MTL regions, but also fusiform regions selective to faces (Golby, Gabrieli, Chiao, & Eberhardt, 2001). In a similar vein, *attractiveness* of a face modulates responses in reward circuitry in the nucleus accumbens (Aharon, Etcoff, Ariely, Chabris, O'Connor, & Breiter, 2001) and *trustworthiness* modulates neural activity in the amygdala and the superior temporal sulcus (Winston, Strange, O'Doherty, & Dolan, 2002). Additionally, faces that are more *familiar* (i.e., famous faces) cause differential activity and repetition effects in fusiform face regions (Eger, Schweinberger, Dolan, & Henson, 2005). Here, we will show through a behavioral study (Experiment 2) that several facial traits (including trustworthiness, attractiveness, and familiarity) explain a significant amount of the variance in face memorability. With knowledge on the perceptual traits that influence face memory, future neuroscientific study could examine the relations between these different processing modules in the brain.

Finally, memorability could provide insights into the content specificity (e.g. face, scene, object) of different regions of the MTL. Interestingly, memory work has found some regions that appear more selective for the processing of faces (Kleinhans et al., 2007), scenes (Staresina, Duncan, & Davachi, 2011) and objects (Buckley & Gaffan, 2006; Devlin & Price, 2007) in the amygdala, hippocampus, and perirhinal cortices, respectively (Liang, Wagner, & Preston, 2012; Litman, Awipi, & Davachi, 2009; Preston, Bornstein, Hutchinson, Gaare, Glover, & Wagner, 2010). Knowing the memorability of images of object, scene and face stimuli allows one to look for the neural sites potentially sensitive to an abstract representation of memorability beyond item-specific perceptual features. This would also allow one to determine if such sites are generic versus specific to certain stimulus categories.

The Current Study

Here, we conduct two behavioral experiments that characterize face memorability at a large scale. In Experiment 1, we assemble a large set of memorability scores using an online memory game with a novel, natural face photograph database. We find that these memorability scores are highly consistent across the population, demonstrating that face memorability can be used as a singular measure.

In Experiment 2, we determine the contributions of twenty facial attributes to memorability. We find that several facial attributes significantly contribute to a model for predicting memorability. However, even after accounting for these attributes as well as noise in the data (e.g., differences in subjective experience), there is still a large amount of variance to memorability, indicating that memorability cannot solely be reduced to a combination of these facial attributes.

This two-fold look at memorability is essential, as it allows both the decomposition of memorability into specific traits, but also the usage of memorability as its own metric. Whereas most work has looked at memory at the level of the individual observer, our study provides a benchmark for studying face memorability and its relations to facial traits at the specific item level. This perspective lends itself to novel applications of the study of human memory, including the manipulation or "training" of memory by altering stimuli based on their memorability, leading to potential innovations in the domains of education, memory rehabilitation, computer science, data storage, and neuroscience.

Experiment 1: The Face Memory Game - Do people find the same faces memorable or forgettable?

Most studies on human visual memory have evaluated observers' performance on picture memory, examining its time frame or capacity. Whereas work on word memory has looked at recognition performance based on item-effects (Hintzman & Hartry, 1990; Freeman et al, 2010), only a few studies have looked at the memorability of images themselves. Isola and collaborators (2011a, 2011b) presented participants on Amazon Mechanical Turk with a visual memory game designed to estimate the probability that a given photograph will be recognized after a single view. They found that memorability is an intrinsic and stable attribute of images that is shared across different viewers and contexts. However, pictures in these large-scale memorability studies were very diverse, covering hundreds of semantic categories (like Standing, 1973). Faces, on the other hand, belong to a single semantic category and are perceptually similar. Are specific face photographs universally memorable or forgettable?

Method

Stimuli: The 10k US Adult Faces Database

It is well-known that face recognition depends on the observer's experience: we are more sensitive to faces within our own age group (Anastasi & Rhodes, 2005) and race (Chiroro & Valentine, 1995; Meissner & Brigham, 2001). In order to have an unbiased platform for the study of face memorability (Figure 1), we built a 10,168-image database of face photographs representative of the adult United States population (Bainbridge, Isola, Blank, & Oliva, 2012). Using an online random name generator based on the 1990 US Census name distribution (Kleimo, 2011; US Census, 1990), we randomly sampled 25,000 first and last names and automatically downloaded from Google Image Search several color face photographs associated with random pairings of these first and last names. Five observers (two authors) deleted from the database recognizable celebrities, low quality images, children, and faces occluded with objects or with unusual makeup. The resulting 10k US Adult Faces Database has 10,168 individual faces, following gender, age and race distributions of the adult US population (see Figure 1 and Table 1).

Face Memory Game

We conducted a large-scale visual memory experiment using online participants (877 workers) from the crowdsourcing tool Amazon Mechanical Turk (AMT), following the protocol of Isola et al. (2011a). The "game" was structured as up to thirty levels of 120 photos each. Each level lasted 4.8 minutes and participants could take brief breaks between them. Although labeled "levels" to give a sense of progress to the participant, the levels did not differ from each other in difficulty or stimulus type. Participants could end the game at any time, and their data were used up to that point. Participants could also restart the game until the completion of the thirty levels. With each restart, participants saw only new stimuli.

In each level, the participants were shown a timed sequence of face images (1 second per face, with a 1.4 second inter-stimulus interval) and asked to press ‘r’ when the same image repeated (Figure 2). Images could repeat within and across levels. From the 10k US Adult Faces Database, 2,222 photos were randomly selected as *target* images, while 6,468 were used as *filler* images. Measures of memorability are taken as hit rates (HR) and false alarm rates (FAR) for repetition detection on the target photos, where repetitions were spaced 91-109 photos apart. This broad space between images (about 4.5 minutes) allows capturing of memory well beyond short term memory and working memory. Isola, Xiao, Parikh, Torralba, & Oliva (in press) tested how image memory varies when the second image exposure ranges from 36 seconds (or ~15 intervening images) to 40 minutes (~ 1,000 intervening images) after the first image, and found that memorability rank remains stable across these various delays. Therefore, in Experiment 1, we choose to test memory at approximately four minutes after the first face exposure, which allows the testing of many stimuli in a short period of time, while tapping into long-term storage. Repetition with the filler photos acted as a “vigilance task” to test the reliability of participants, with repetitions spaced 1-7 photos apart. The rest of the fillers were used as spacing between the target photos. In order to ensure the quality of participants, only workers with an AMT approval rate of at least 95% were allowed to participate. Workers who did poorly on the vigilance task (with poor performance being measured as greater than 50% false alarms on the last 30 non-repeat images or less than 50% hits for the last ten vigilance repeats) were also prevented from continuing with the task. Only workers with a computer IP address within the US were allowed in the game in order to match worker demographics with the stimulus demographics.

Stimuli Demographics Survey

To collect basic demographic information (age, race, and gender) on the face stimuli, we ran a survey on AMT with twelve respondents for each of the 2,222 target faces. For age, participants could choose the ranges of less than 20 years of age, 20-30, 30-45, 45-60, and over 60 years of age, while for race, they could choose white, black, Hispanic, East Asian, South Asian, Middle Eastern, or other. These options were selected based on common AMT demographics, and the same choices were available when memory task participants were asked to indicate their own demographics before beginning the game. The demographics of the faces match the demographics of both AMT workers on the memory task and the US Census (see Table 1), diminishing memory effects related to the own-race bias (Chiroro & Valentine, 1995; Meissner & Brigham, 2001) or the own-age bias (Anastasi & Rhodes, 2005).

Results

Each target photo with its repetition was seen by an average of 81.7 workers. On average, target faces were correctly recognized in 51.6% of trials (*SD*=12.6%). The average rate of false alarms was 14.4% (*SD*=8.7%).

Are memorability scores reliable between different groups of observers?

We assessed the reliability of these memorability scores (HR and FAR) across face images by looking at the correlation between split-half rankings. For both HR and FAR, we ranked the images according to one random half of the participants and compared them to the scores given by the other half of the participants (Figure 3). Over 25 of these random split-half trials, the average Spearman's rank correlation between scores given by the two halves of the participants was $\rho = 0.68$ for HR (*min* = 0.66, *max* = 0.69) and $\rho = 0.69$ for FAR (*min* = 0.67, *max* = 0.71). There is also sizable variation in face photo memorability, with HR ranging from 15.5% (the most forgettable photos) to 89.9% (the most memorable photos) and FAR ranging from 0% to 51.5%. The strength of these correlations, in spite of individual differences and other potential sources of noise, demonstrates that we have characterized real differences between face photos and that both HR and FAR are varied yet reliable measures across the population. We also looked at d-prime, a metric that combines both HR and FAR, and found a similarly high average Spearman's rank correlation of $\rho = 0.69$.

Are memorability scores consistent across individual observers?

The above analysis describes the population-level reliability of memorability scores, but it remains unclear to what degree a single observer's performance will be consistent with the population. In order to measure this, we used leave-one-out cross-validation over participants in our experiment. We measured how well the scores of N-1 participants on our task predict the scores of the N^{th} participant. For a given set of N-1 participants, we computed the average HR and FAR for all images that could have produced a hit or false alarm, respectively, in the N^{th} participant. We then used logistic regression to predict whether or not the N^{th} participant gave a hit (or a false alarm) based on the average rates from the N-1 participants. We repeated this analysis for each possible N^{th} participant (excluding participants who scored less than 2 images), giving us 854 estimates of the regression coefficient β_1 (the slope of the logit function) for HR and 876 estimates for FAR. Some participants are more consistent with the population than others. In order to estimate an average β_1 across the population of participants, we took a weighted mean over the β_1 estimates for each individual participant, weighting by the inverse variance of the β_1 estimate for each participant. We weighted in this way since different participants scored different numbers of images and so the β_1 estimates for different participants have widely varying precision; weighting by inverse variance minimizes the variance of the estimated population mean. We found a weighted mean β_1 of 3.77 for HR ($p < 0.01$), and 6.42 for FAR ($p < 0.01$). The p-values here refer to the probability of observing effects with magnitude at least this large, under the null hypothesis that the true β_1 actually equals 0 for all subjects. We simulated the null distribution using a within-subjects Monte-Carlo permutation test[1]. These β_1 values suggest that an individual observer will tend to exhibit a fair amount of consistency with the population in terms of which images he or she finds more and less memorable.

Is there reliability amongst categories of face memorability?

Can we separate the signals for false memories and true memories in our data? If a photo receives both a high HR and a high FAR, it may be highly memorable, but it could also just be a highly familiar-looking face. An alternate version of memorability can be made for photos with high HR and low FAR. This is because in recognition memory, memorable items often evoke both higher HR and lower FAR than forgettable items – what is termed a "mirror effect" (Glanzer & Adams, 1985, 1990).

To isolate truly memorable photos, we split the faces along the median HR and FAR to create groupings of four performance profiles, consisting of high / low HR and FAR (Figure 4).

[1] In more detail, we simulated chance consistency by randomly shuffling each participant's responses amongst all the images that participant responded to. We ran this simulation 1,000 times and calculated resulting weighted mean β_1 values to give an estimate of the null distribution of these values. Modeling the p-value as the success rate parameter of a binomial distribution, we found that $p < 0.01$ contains the 99.9% confidence interval for the p-values of our observed β_1 measurements for both HR and FAR (using the Clopper-Pearson method to estimate the confidence interval).

Over 25 split-halves, subjects' performance profile assignments agreed on average 55.4% of the time (within the 25 trials, *min* = 54.1%, *max* = 57.3%; compared to a chance level of 25%), with similarly high agreement levels for each quadrant. This agreement shows that there are different categories of faces in terms of memorability, for example, distinctive and highly memorable faces (high HR, low FAR), typical faces (high HR, high FAR), highly forgettable faces (low HR, low FAR), and faces that evoke many false memories but few true ones (low HR, high FAR; Figure 4).

Discussion

Here we find that intrinsic memorability is a reliable measure of face photographs, consistent across viewers, and therefore separate from individuals' subjective experiences. Such a consistency may appear surprising at first, given the plethora of memory work showing individual differences between viewers and images on memory (Anastasi & Rhodes, 2005; Chiroro & Valentine, 1995; Lewis & Johnston, 1997; Meissner, Brigham, & Butz, 2005). However, there is still remaining variance between observers, and we examine the contributions of various observer-based and item-based factors on memorability in Experiment 2.

The high consistency of face memorability observed in Experiment 1 lends itself to several applications, ranging from neuroscience studies such as exploring modulation of MTL activity by intrinsic image memorability, to computer graphics applications such as automatically manipulating the memorability of specific faces. While the memorability rank of generic photographs has been shown to be a stable function over short to long time delays, stretching up to 40 minutes after presentation (Isola et al., in press), future work may investigate the lasting power of face memorability.

Experiment 2: What components make up face memorability?

Given that memorability is a varied and reliable characteristic of faces, can it be explained by other facial attributes? Previous studies suggest that face recognition may be affected by distinctiveness of faces, as well as familiarity and subjective ratings of memorability (Bartlett et al., 1984; Deffenbacher, Johanson, Vetter, & O'Toole, 2000; Light, Kayra-Stuart, & Hollander, 1979; Vokey & Read, 1992). Other high-level attributes, such as attractiveness, have been argued in both directions to be linked with facial averageness (Langlois & Roggman, 1990) but also uniqueness (Alley & Cunningham, 1991), while untrustworthy faces have been found to be more memorable (Rule et al., 2012). However, no work so far has examined a large spread of facial attributes to explain what may influence the memorability of particular items.

Method

Attribute Selection

We started with a quasi-exhaustive set of facial attributes by compiling a collection of twenty face traits from three sources. First, we selected the fourteen personality traits found by Oosterhof & Todorov (2008) to influence face evaluation. In their work, Oosterhof & Todorov asked 55 participants to write free-flow descriptions of a set of faces, and found that these traits could be classified into fourteen different overarching trait dimensions. Importantly, these traits correspond to spontaneously inferred judgments individuals made from seeing a facial portrait, listed here by decreasing order of frequency of use: *attractive, unhappy, sociable, emotionally stable, mean, boring, aggressive, weird, intelligent, confident, caring, egotistic, responsible, trustworthy*. In addition, we selected three memory-related attributes described by Vokey &

Read (1992), namely, *memorability*, *typicality*, and *familiarity*. In their work, Vokey & Read found that typicality ratings of faces could be decomposed into components related to memorability and familiarity as well as HR and FAR performance on a memory task. Finally, we added *commonness*, *emotional magnitude* and *friendliness* to the list, as these attributes were found significantly correlated with memorability in a previous pilot study.

Attribute Antonyms Survey

In order to control for possible biases of attribute valence on memorability, we designed an AMT survey to choose antonyms for each of the attributes used by Oosterhof & Todorov (2008), Vokey & Read (1992) and our pilot experiment. Twenty workers were asked to select the best antonym for each attribute when describing a face, with the list of possible antonyms for each attribute chosen from Thesaurus.com (2013). The majority response was selected as the corresponding antonym for each attribute.

Facial Attributes Survey

Armed with the twenty facial and personality traits, each described by a positive and a negative valence word, we ran an AMT survey for each of the 2,222 target faces used in Experiment 1. The original words used in the previous studies and their antonyms were randomly split across two versions of the attribute-labeling survey, and fifteen different participants were recruited for each version. Ratings were conducted on a 9-point Likert scale, ranging from 1 (not at all) to 9 (extremely), as used in Oosterhof & Todorov (2008). In order to assist workers, each question included a "?" that could be clicked for a pop-up window with a dictionary definition of the word in that question. The survey also included a "catch" question to eliminate workers who were answering at random, asking them to indicate a number displayed on the screen (randomly chosen from 1 to 9). When participants failed the catch question, we removed the data for that entire survey from the analyses (only 0.87% of surveys). Only AMT workers with over 95% approval ratings and IP addresses within the US were allowed to participate in the survey. Each survey of 21 questions paid $0.07, for an hourly rate of approximately $3. 1,274 workers participated in this study, and their demographics closely match those of Experiment 1, the 10k US Adult Faces Database, and the 1990 US Census (see Table 1).

When the survey was complete, we looked at Pearson's correlations between attribute-antonym pairs and found that all pairs had significant negative correlations as expected ($p < 10^{-4}$). Thus, responses to antonyms were combined with the original words (after aligning them on the same scale by subtracting corresponding antonyms from 10), producing a total of 30 ratings for each of the twenty attributes, for all 2,222 faces (approximately a total of 66,660 responses per attribute).

Data Analyses Summary

We looked at the influence of these facial traits on memorability by running several multiple linear regression models. First, because HR and FAR are proportions and thus bounded between 0 and 1, we logit-transformed them in order to be able to form linear statistics with them. Some FAR scores were initially 0, and were re-coded as 0.01 to prevent logit-transformed scores of infinity. Attribute scores were normalized into standardized z-scores, with mean zero and standard deviation of one.

How do the attributes independently contribute to the scores of memorability? To answer, we ran two multiple linear regression models separately on the HR and FAR from Experiment 1. The facial attributes were the independent variables, while HR and FAR were dependent variables for the models. One multiple linear regression was run on the set of twenty attributes, and a second one was run on the set of fourteen facial traits emphasized by Oosterhof & Todorov (2008). In order to get a comparative measure of noise that could account for remaining variance in the models, the same linear regressions were also run including memorability score regressors from random split-halves of the participants.

Results

The multiple linear regression results are detailed in Tables 2 (HR) and 3 (FAR), which summarize the statistics and models run with the attributes. Pearson's correlations between all attributes, as well as with the memorability scores can be found in the Supplemental Material (Tables S1 and S2). As a note, the beta (β) values of the regression analyses only differ in sign when running the model based on phrasing them using the original attribute names versus their antonyms. Thus, for clarity in this paper, we frame these attributes based on their valence (i.e., positive or negative), with all betas reported as an absolute value. For all of these models, we looked at model residual plots to confirm a linear model was indeed a good fit, and found for all that the residuals were normally distributed.

Which facial attributes are significant predictors of recognition success?

The multiple linear regression model run on the twenty attributes with the HR shows that the combination of traits is able to explain 23.5% of the variance in correct recognition of a face photograph. As shown in Table 2, the eleven significant contributors for HR are (in order of high to low β): subjectively memorable (β = 0.18), irresponsible (β = 0.13), kind (β = 0.12), unhappy (β = 0.11), atypical (β = 0.09), trustworthy (β = 0.09), unintelligent (β = 0.08), unattractive (β = 0.06), emotional (β = 0.05), uncommon (β = 0.05), and unfamiliar (β = 0.03). This model does a good job at describing memorability ($F(2201, 20) = 35.06$, $p < 10^{-115}$), with an overall model fit of adjusted $R^2 = 0.235$. Our results align with those of Vokey & Read (1992) that find typical and unmemorable faces to be linked with lower HRs. However, other attributes stand out in our model that are not mentioned in memory literature, such as irresponsible, kind, and unhappy, indicating that face memorability has some personality and social components to its determination.

The multiple linear regression of the fourteen social and personality traits of Oosterhof & Todorov (2008) explains 14.5% of the variance of HR. The six significant contributors were (in order by β): interesting (β = 0.21), irresponsible (β = 0.13), kind (β = 0.10), unhappy (β = 0.09), emotionally unstable (β = 0.08), and unattractive (β = 0.05). This model also significantly describes HR ($F(2207, 14) = 28.58$, $p < 10^{-69}$), with an adjusted $R^2 = 0.145$. While this amount of described variance is not as high as the full model, it is still significant, indicating that memorability may not solely be determined by memory-related traits (i.e., typicality, familiarity, commonness, subjective memorability), but is also influenced by other personality and social traits, even at such a brief image presentation. Note that in this reduced set, attributes such as interesting and emotional are now significant. As "interesting" is the highest loaded trait, this may encompass the various measures of typicality that were removed (i.e., a typical face is boring while an atypical one is interesting). In the full, twenty-attribute model, all the variance of "interesting" was likely described by a combination of all the other traits, causing it to not be a significant contributor. Similarly, emotional stability may account for the “emotionality” term, which was a significant contributor in the full model.

Which facial attributes are significant predictors of false memories?

The multiple linear regression model run on the twenty attributes with the FAR shows that the combination of twenty traits is able to explain 16.4% of the variance in false alarms on a face photograph. Thirteen attributes were found to be significant contributors (in order of β): responsible (β = 0.28), uncertain (β = 0.17), kind (β = 0.15), introverted (β = 0.14), intelligent (β = 0.13), atypical (β = 0.11), trustworthy (β = 0.10), attractive (β = 0.09), familiar (β = 0.08), unemotional (β = 0.08), caring (β = 0.07), unhappy (β = 0.07), and friendly (β = 0.06). Interestingly, the top-loaded attributes here are higher-level personality traits rather than memory-related traits. This model significantly explains FAR ($F(2201, 20) = 22.76$, $p < 10^{-74}$), but with a lower explained variance compared to the model for HR (adjusted $R^2 = 0.164$). Using only Oosterhof & Todorov’s fourteen attributes, we find that six attributes are significant contributors: boring (β = 0.28), attractive (β = 0.21), happy (β = 0.19), responsible (β = 0.16), confident (β = 0.10), and humble (β = 0.09). This model is significant ($F(2207, 14) = 22.82$, $p < 10^{-54}$), but with a lower explained variance (adjusted $R^2 = 0.121$). Like with HR, “boring” now takes the lead term, likely to account for the missing memorability attributes.

What is the remaining variance in these models?

While these multiple linear regression models are all significantly able to predict HR and FAR, there is still a relatively high amount of unexplained variance left over. In order to look at how much of this unexplained variance is noise, we also ran the models including regressors for the memorability scores along 25 random half-splits of the participants. For example, with the HR twenty attributes model, an additional “attribute” that included a random half of the HR data (split by participants) was used to predict the other half of the HR data, and this split and model was run 25 times to get a range of statistics for the models. A summary of the statistics and beta values for these models can be seen in the Supplemental Material (Tables S3-S6).

For the full HR model with twenty attributes and a HR regressor, all 25 splits significantly described HR (*mean* $F(2200,21)$ =108.46, *min* = 99.62, *max* = 115.16), with an average p-value of 3.21×10^{-301} (*min* ~ 0, *max* = 4.23×10^{-300}). The adjusted R^2 also doubled with this additional regressor, going from 0.235 in the original model to a mean adjusted R^2 of 0.504 (*min* = 0.483, *max* = 0.519). These values serve as the upper bound performance of the model; while 49.6% of HR variance appears to be noise (encompassing individual differences in subjective experience, environment, etc), there is still 26.9% of the variance in HR that is reliable across participants yet unexplained by the full-encompassing set of twenty facial attributes. The model performs similarly well using only the fourteen Oosterhof & Todorov (2008) attributes with an additional HR regressor (*mean* $F(2206, 15)$ = 144.62, *min* = 133.04, *max* = 153.48; *mean* p = 2.43×10^{-296}, *min* ~ 0, *max* = 6.07×10^{-295}), with a mean adjusted R^2 of 0.492 (*min* = 0.470, *max* = 0.507). This leaves 34.7% remaining variance in HR after accounting for the fourteen attributes and noise.

We find similar results with false alarms. The 25 splits of the full FAR model with twenty attributes and a FAR regressor also significantly described FAR (*mean* $F(2200,21)$ = 88.84, *min* = 85.72, *max* = 95.04; *mean p* = 4.45×10^{268}, *min* = 2.32×10^{-289}, *max* = 8.92×10^{-267}). The adjusted R^2 is almost three times higher with the additional FAR regressor (*mean adjusted* R^2 = 0.454, *min* = 0.445, *max* = 0.471), leaving a remaining variance in FAR, after accounting for the twenty attributes and noise, of 29.0%. For the fourteen attributes FAR model plus the FAR regressor, the model has similarly high performance (*mean* $F(2206, 15) = 121.30$, *min* = 116.52, *max* = 129.32; *mean* $p = 2.23 \times 10^{-267}$, *min* = 1.39×10^{-288}, *max* = 5.14×10^{-266}), with a mean adjusted R^2 of 0.448 (*min* = 0.437, *max* = 0.465) and a remaining variance in FAR after the fourteen attributes and noise of 32.7%.

Discussion

These results give an interesting look into the influences on how well certain faces are remembered. First, our results support previous literature that have found correlations of atypicality with HR and familiarity with FAR (Vokey & Read, 1992; Bartlett et al., 1984), while discovering that some personality and social traits contribute to memorability. While other works have used single attributes to examine memorability, (e.g., correlating untrustworthiness with memorability, see Rule et al., 2012), our work offers a more comprehensive landscape of the traits that may influence memorability, using a large-scale, natural dataset.

It is interesting to note the valence tendencies of the attributes and their connection with HR and FAR. Recognition success is associated with a mix of positive and negative attributes, whereas false memories seem to be mainly positive. We find that HR is more dependent on memory-related metrics (with memorable, atypical, unfamiliar, and uncommon all receiving strong beta weights), while FAR is more dependent on social and personality traits.

So, what makes a face memorable? To create a familiarity effect, where a face is recognized whether it was seen previously or not (maximizing both HR and FAR), the face should be one that has increased values of kindness and trustworthiness, but also some atypicality. While false alarms are sometimes viewed as noise or response errors, intentionally elicited false alarms have potential to become a powerful tool in social or marketing contexts (i.e., also making a face more responsible, intelligent, attractive, and unemotional). We also find several attributes with a classical mirror effect (high HR, low FAR; Vokey & Read, 1992; Glanzer & Adams, 1985, 1990), where a face is correctly remembered with no false memories (high HR, low FAR; specifically, faces that are irresponsible, unhappy, unintelligent, unattractive, and unemotional).

While some facial attributes contribute to memorability, after accounting for these attributes as well as noise (likely made up of participant differences in subjective experience, environment, memory ability difference, etc) there is still a remaining variance reliable across participants in the memorability scores. Essentially, the variance in HR can be seen as a combination of 23.5% personality, social, and memory-related attributes, 49.6% variance between participants (e.g., individual differences, environment differences), and 26.9% unexplained variance between images. Similarly, the variance of FAR is a combination of 16.4% attributes, 54.6% variance between participants, and 29.0% unexplained variance between images.

This remaining unexplained variance between images indicates that while some face attributes are related to memorability, there is still more to memorability than just these factors. This unexplained variance across participants exists even when using the model including attributes believed to be closely tied to memorability, such as typicality, commonness, and familiarity. Whereas the nature of this unexplained variance warrants further study, it suggests that memorability can be used as a high-level attribute intrinsic to face images, which cannot be simply reduced to a combination of other face-related attributes. Thus, the combined study of the singular property of memorability and the attributes contributing to memorability gives a two-fold benefit: first, it allows quantifications, predictions, and comparisons of memorability across images and observers, and second, it suggests which attributes to alter to manipulate face memorability in future work.

General Discussion

Memorability as an intrinsic high-level facial attribute

Here we establish a large-scale database of 10,168 natural, representative face photographs of the US adult population, with objective memorability scores, and high-level attributes motivated by previous psychology literature for 2,222 of those faces. Whereas previous research has noted that memorability of a face may differ based on a few isolated attributes singled out for examination, such as matching race (Chiroro & Valentine, 1995; Meissner et al., 2005), distinctiveness (Valentine, 2001), or previously experienced images (Lewis & Johnston, 1997), the current study shows surprising reliability across people of diverse backgrounds viewing a widespread distribution of photos.

The present work brings three contributions: first, there are similarities across participants in the relative memorability of different face photos; second, a proportion of memorability can be described by a combination of facial attributes; and third, even after accounting for these attributes and noise, there is still a large amount of unexplained variance to memorability reliable across participants, indicating it is not only a composite of other facial attributes. Together, these findings suggest that face memorability can be used as a metric of interest in the study of faces and memory.

The idea of quantifying memorability of a face lends itself to many useful applications of psychology research to mainstream society in future work (Oliva, Isola, Khosla, & Bainbridge, 2013). Memory research has been mostly subject-focused, but an item-focused approach enables several possible innovations in both research and industry; instead of only improving our own memory capacities, we can also work to make our worlds easier to remember. Algorithms could automatically identify the most memorable face from an album to use in textbooks, magazines, or even social network profiles. Movie studios could use memorability to generate memorable main characters and forgettable extras in a 3D film. Smartphone applications could teach people how to apply makeup to maximize their memorability. Besides offering novel applications of basic cognitive psychology, predicting face memorability also opens a rich panorama of research questions in the human neuroscience of memory and face perception.

Applications of Memorability to Neuroscience Study

Decades of neuroscience research have established a critical role of the MTL in memory. Memorability, or the probability of remembering an event after a single exposure, is a question not only of recollection but also of perception. A critical and unexplored question is the impact of memorability on MTL structures and content-sensitive brain regions (e.g. the fusiform face area, the parahippocampal place area). Memorability postulates the existence of intrinsic perceptual features that determine what is going to be remembered and forgotten, independent of context and an observer's personal history. This new theoretical spin not only works toward uniting the often separately studied fields of memory and perception, but also lends itself to new analyses in the neuroscience domain. For example, memorability allows one to study memory without testing an individual's memory – a possible tool for studying patients with memory, social, or facial processing impairments (e.g., prosopagnosia, autism, Alzheimer's, etc), as we could examine how cortical activation differs from normal observers when viewing highly memorable faces (compared to forgettable ones). Like perceptual tasks, studies of image memorability also only require a single exposure per image, and they can be blocked by memorability score, rather than using an event-related design (which is required by most memory studies). This design not only increases the power of the study, but also allows examination of how MTL structures act during the encoding of the image. This design also avoids repetition suppression, as it has been postulated that activation may differentially change over subsequent image presentations based on memory strength (Henson, Shallice, Gorno-Tempini, & Dolan, 2002). Memorability of an image offers a new metric for studying the representational role of MTL structures to the very first exposure of an image, and may challenge the theoretical view that perception and memory processes are anatomically distinct. What can memorability teach us about the representational capabilities of MTL structures, and the neuroanatomical segregation or overlap of memory and perceptual neural processes?

In functional magnetic resonance imaging (fMRI) studies, whereas selective cortical regions have been found for faces (Kanwisher & Yovel, 2006), scenes (Epstein & Kanwisher, 1998; Dilks, Julian, Paunov, & Kanwisher, 2013) and objects (Grill-Spector, Kushnir, Edelman, Avidan, Itzchak, & Malach, 1999; Konkle & Oliva, 2012), there is an ongoing discussion on to what degree MTL structures exhibit content selectivity. The perirhinal cortex has been associated with object perception (Buckley & Gaffan, 2006; Devlin & Price, 2007), and dissociations between perirhinal and parahippocampal cortices are found respectively for object and scene stimuli (Litman et al., 2009; Staresina et al., 2011). Whereas the anterior hippocampal and subiculum responses seem content general, the posterior hippocampus discriminates scenes better than other stimuli (Liang et al., 2013; Preston et al., 2010). The anterior MTL is also found to be more selective to faces compared to the posterior MTL (Liang et al., 2013), and the amygdala has been linked with face recognition (Young, Aggleton, Hellawell, Johnson, Broks, &

Hanley, 1995; Kleinhans et al., 2007). However, these regions are incredibly difficult to image and only recently has work in this field begun incorporating multivariate techniques. There is still the open question of whether these MTL structures have category specificity as we see in more perceptual regions (e.g., the fusiform face area versus the parahippocampal place area), or whether they have a more graded difference. One could identify regions that differentiate between high and low memorability for different stimulus types, and examine where in the cortex these differences appear.

Lastly, the item-centric approach of memorability lends itself to even further analyses. Using items with known memorability scores allows one to do multivariate analyses with memory. Previously, some such studies have been conducted based on using an observer's memory performance at the retrieval phase to decode hippocampus activity during the encoding phase (Shrager et al., 2008), however using population-based memorability instead may allow us to produce multivariate models that can be generalized to a wider range of participants. Using the item-centric approach aids in the selection of stimuli (as one can select ahead of time items that will be remembered and forgotten for cleaner multivariate analyses), and also allows one to easily look at differences across observers. Finally, given the large collection of attributes for faces, this work gives the potential to look at pattern differences between these attributes and memorability in neuroscience studies, to examine what features may be processed upstream or downstream of memorability.

Conclusion

This study serves as a large-scale, empirical look at face photograph memorability. Not only did we find that people did remarkably well at identifying repeated face images after a single exposure, but that memorability scores were highly consistent across different observers. Whereas a proportion of memorability is influenced by a combination of high-level facial traits, even when accounting for these traits and noise (including observer differences), a large amount of variance in memorability is still left unexplained. These results indicate that memorability is, in itself, a predictable, singular measure of a face picture. We have outlined several directions in which the neuroscience field can utilize face memorability as a tool to examine questions on the neuroscience of memory encoding. Memorability is a novel and well-poised topic for future study in psychology, computer vision, and neuroscience.

Acknowledgements

The 10k US Adult Faces Database, as well as the names used to query the faces, demographic information, the memorability scores, and attribute labels for the target faces are available online for public use (http://www.wilmabainbridge.com/facememorability.html). Preliminary data were presented in Bainbridge et al., 2012 at the Cognitive Science Society, 2012.

W.A.B. is supported by the Department of Defense, through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. P.I. is funded by a National Science Foundation Graduate Fellowship. This work is funded by National Science Foundation grant (1016862), Google and Xerox Research Awards to A.O. The authors would like to thank Constance Bainbridge and Idan Blank for helping with stimuli generation, as well as Marc Howard, Timothy Brady, Daniel Dilks, Ran Hassin, and three anonymous reviewers for their insightful advice with the analyses and the work as a whole.

References

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron, 32*(3), 537-551.

Alley, T. R., & Cunningham, M. R. (1991). Averaged faces are attractive, but very attractive faces are not average. *Psychological Science, 2*(2), 123-125.

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12*(6), 1043-1047.

Bainbridge, W. A., Isola, P., Blank, I., & Oliva, A. (2012). Establishing a database for studying human face photograph memory. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 1302-1307.

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12*(3), 219-228.

Bonnici, H. M., Kumaran, D., Chadwick, M. J., Weiskopf, N., Hassabis, D., Maguire, E. A. (2012). Decoding representations of scenes in the medial temporal lobes. *Hippocampus 22*(5), 1143-1153.

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

Bruce, V., Burton, M. A., & Dench, N. (1994). What's distinctive about a distinctive face? *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 47*(1), 119-141.

Buckley, M. J., & Gaffan, D. (2006). Perirhinal cortical contributions to object perception. *Trends in Cognitive Sciences, 10*(3), 100-107.

Busey, T. A. (2001). Formal models of familiarity and memorability in face recognition. In M.J. Wenger & J.T. Townsend (Eds.), *Computational, geometric, and process perspectives on* *facial cognition: contexts and challenges*. Mahwah, NJ: Lawrence Erlbaum Associates, 147-192.

Bussey, T. J., & Saksida, L. M. (2007). Memory, perception, and the ventral visual-perirhinal-hippocampal stream: thinking outside of the boxes. *Hippocampus*, *17*, 898-908.

Cahill, L., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature, 377*(6547), 295-296.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology*, *48*(4), 879-894.

Deffenbacher, K. A., Johanson, J., Vetter, T., & O'Toole, A. J. (2000). The face typicality-recognizability relationship: encoding or retrieval locus? *Memory & Cognition*, *28*(7), 1173-1182.

Devlin, J. T., & Price, C. J. (2007). Perirhinal contributions to human visual perception. *Current Biology*, *17*(17), 1484-1488.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area (OPA) is causally and selectively involved in scene perception. *Journal of Neuroscience, 33*, 1331-1336 .

Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology*, *23*(5), 714-747.

Eger, E., Schweinberger, S. R., Dolan, R. J., & Henson, R. N. (2005). Familiarity enhances invariance of face representations in human ventral visual cortex: fMRI evidence. *NeuroImage*, *26*, 1128-1139.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*, 598-601.

Fink, B., Grammer, K., & Thornhill, R. (2001). Human (Homo sapiens) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology*, *115(*1), 92-99.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Science*, *11*(2), 77-83.

Foa, U. G., & Foa, E. B. *Societal structures of the mind*. Springfield, IL: Charles C Thomas.

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language, 62*(1), 1-18.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*(1), 8-20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 5-16.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience, 4*, 845-850.

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of Comparative Psychology*, *108*(3), 233-242.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187-203.

Henson, R. N. A., Shallice, T., Gorno-Tempini, M. L., & Dolan, R. J. (2002). Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex*, *12*(2), 178-186.

Hintzman, D. L. & Hartry, A. L. (1990). Item effects in recognition and fragment completion: Contingency relations vary for different subsets of words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6), 955-969.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representations of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(16), 9379-9384.

Isola, P., Xiao, J. X., Parikh, D., Torralba, A., & Oliva, A. (in press). What makes a photograph memorable? *IEEE Pattern Analysis and Machine Intelligence*.

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. *Paper presented at the 25th Conference on Neural Information Processing Systems (NIPS)*. Granada, Spain.

Johnson, K. J., & Fredrickson, B. L. (2005). “We all look the same to me”: positive emotions eliminate the own-race bias in face recognition. *Psychological Science*, *16*(11), 875-881.

Kanwisher, N., Chun, M. M., McDermott, J., & Ledden, P. J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research, 5*, 55-67.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2109-2128.

Kleimo, A. (2011). The Random Name Generator, from http://www.kleimo.com/random/name.cfm.

Kleinhans, N. M., Johnson, L. C., Mahurin, R., Richards, T., Stegbauer, K. C., Greenson, J., Dawson, G., & Aylward, E. (2007). Increased amygdala activation to neutral faces is associated with better face memory performance. *Neuroreport, 18*(10), 987-991.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Scene Memory Is More Detailed Than You Think: The Role of Categories in Visual Long-Term Memory. *Psychological Science, 21*(11), 1551-1556.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Conceptual Distinctiveness Supports Detailed Visual Long-Term Memory for Real-World Objects. *Journal of Experimental Psychology: General, 139*(3), 558-578.

Knight, R. T. (1996). Contribution of human hippocampal region to novelty detection. *Nature, 383*, 256-259.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1*(2), 115-121.

LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *The Journal of Neuroscience, 33*(13), 5466-5474.

Lee, A. C. H., Yeung, L., & Barense, M. D. (2012). The hippocampus and visual perception. *Frontiers in Human Neuroscience, 6,* 91.

Liang, J. C., Wagner, A. D., & Preston, A. R. (2012). Content representation in the human medial temporal lobe. *Cerebral Cortex, 2012*.

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 212-228.

Litman, L., Awipi, T., & Davachi, L. (2009). Category-specificity in the human medial temporal lobe cortex. *Hippocampus, 19,* 308-319.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience, 5,* 910-916.

Meissner C. A., & Brigham J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3-35.

Meissner, C. A., Brigham, J. C., Butz, D. A. (2005). Memory for own- and other- race faces: a dual-process approach. *Applied Cognitive Psychology, 19*(5), 545-567.

Monin, B. (2003). The warm glow heuristic: when liking leads to familiarity. *Journal of Personality and Social Psychology, 85*(6), 1035-1048.

Murray, E. A., & Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Current Opinion in Neurobiology, 11*(2), 188-193.

Oliva, A., Isola, P., Khosla, A., & Bainbridge, W. A. (2013). What makes a picture memorable? *SPIE Newsroom Article, 7 May 2013*. Retrieved May 28, 2013, from SPIE website: http://spie.org/x93734.xml.

Oosterhof, N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, USA*, *105*(32), 11087-11092.

Preston, A. R., Bornstein, A. M., Hutchinson, J. B., Gaare, M. E., Glover, G. H., & Wagner, A. D. (2010). High-resolution fMRI of content-sensitive subsequent memory responses in human medial temporal lobe. *Journal of Cognitive Neuroscience*, *22*(1), 156-173.

Qin, S., van Marle, H. J. F., Hermans, E. J., & Fernández, G. (2011). Subjective sense of memory strength and the objective amount of information accurately remembered are related to distinct neural correlates at encoding. *Journal of Neuroscience, 31*(24), 8920-8927.

Roget's 21st Century Thesaurus, Third Edition. Retrieved April 08, 2013, from Thesaurus.com website: http://thesaurus.com/

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of personality and social psychology, 9*(4), 283-294.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69-72.

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition, 125*(2), 207-218.

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research, 14*(3), 398-409.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*(1), 11-21.

Shrager, Y., Kirwan, C. B., & Squire, L. R. (2008). Activity in both hippocampus and perirhinal cortex predicts the memory strength of subsequently remembered information. *Neuron, 59*(4), 547-553.

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*(2), 195-231.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253*(5026), 1380-1386.

Standing, L. (1973). Learning 10,000 Pictures. *Quarterly Journal of Experimental Psychology, 25*, 207-222.

Staresina, B. P., Duncan, K. D., & Davachi, L. (2011). Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. *The Journal of Neuroscience, 31*(24), 8739-8747.

Sjöstrand, K., Clemmensen, L. H., Larsen, R., & Ersbøll, B. (in press). SpaSM: A Matlab toolbox for sparse statistical modeling. *Journal of Statistical Software*.

Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience, 18*(9), 1488-1497.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*, 67(2): 301-320.

Todorov, A. (2011). Evaluating faces on social dimensions. In A. Todorov, S. T. Fiske & D. A. Prentice (Eds.), *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind*. New York, NY: Oxford University Press.

U.S. Census Bureau (1990). *Frequently occurring first names and surnames from the 1990 Census.* Retrieved from http://www.census.gov/genealogy/names/names_files.html.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quaterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 43*(2), 161-204.

Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology*, *54*(4), 298-303.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20*(3), 291-302.

Wan, H., Aggleton, J. P., & Brown, M. W. (1999). Different contributions of the hippocampus and perirhinal cortex to recognition memory. *The Journal of Neuroscience, 19*(3), 1142-1148.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5,* 277-283.

Young, A. W., Aggleton, J. P., Hellawell, D. J., Johnson, M., Broks, P., & Hanley, J. R. (1995). Face processing impairments after amygdalotomy. *Brain*, *118*(1): 15-24.

Figure 1. 200 randomly selected faces from the 10k US Adult Faces Database.
Each photo in the database was resized to 256 pixels in height, and cropped with an oval surrounding the head to reduce background features.

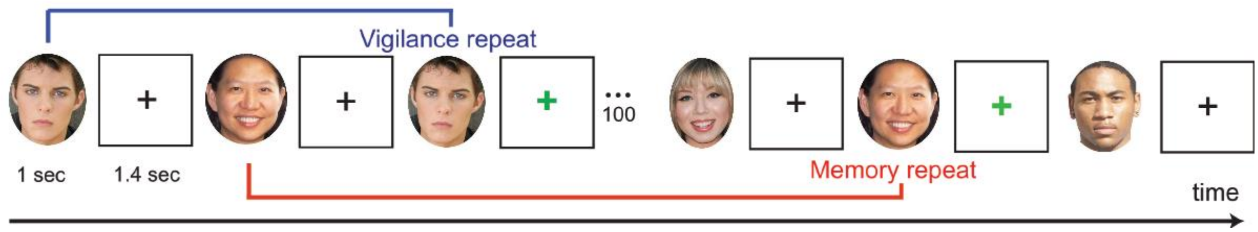


Figure 2. A flowchart of the visual memory task.
Each image was presented for 1 second, followed by a 1.4-second fixation cross. Participants pressed “r” when they saw a repeat of a face image. A green cross appeared as feedback for correct positive responses (hit) and a gray X appeared to indicate a key press error (false alarm).

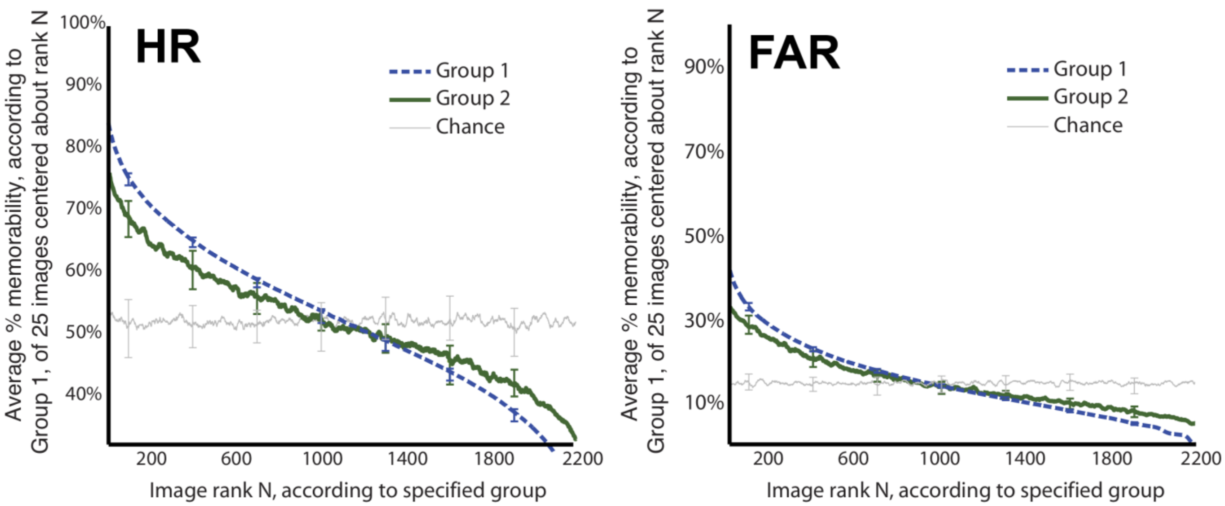
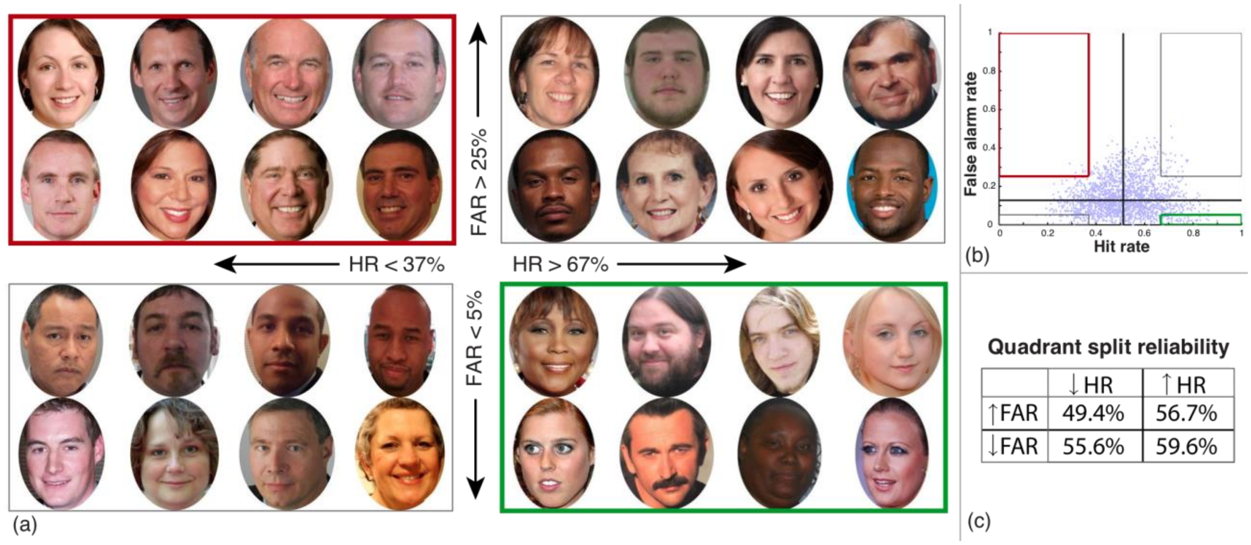


Figure 3. Left: reliability graph for Hit Rate (HR); Right: reliability graph for False Alarm Rate (FAR).

Split-half reliability depicted using memorability score rank-size plots, averaged over 25 random splits. For each split, participants were separated into two random groups: Group 1 and Group 2. Face photos were arranged along the x-axis in rank order from highest to lowest HR (left) and highest to lowest FAR (right) according to Group 1 participants (dotted line). The memorability scores of Group 2 were then plotted for the same ranking of photos (solid line). If the two groups are in perfect agreement, the solid and dotted lines should coincide. On the other hand, if no consistency exists across the participants, the solid line should coincide with the thin chance line, which is the result of assigning the images random ranks (i.e. randomly permuting the x-axis). Plots are smoothed by convolving each line with a length-25 box filter (i.e. the y-value of each plotted point is the average of the y-values of all points up to 12 ranks above and 12 ranks below the rank of the plotted point). Error bars give 80% confidence intervals estimated with a bootstrap (that is, each interval contains 80% of the 25 samples given by the random splits).



	↓HR	↑HR
↑FAR	49.4%	56.7%
↓FAR	55.6%	59.6%

Figure 4. Reliability amongst different quadrants of faces grouped by high and low HR and FAR.

a) Random examples of faces at the extremes of each quadrant of high/low HR and FAR. The faces in the green box have both higher HR and lower FAR than the faces in the red box. This comparison demonstrates a "mirror effect" (Glanzer & Adams, 1985), suggesting the faces in the green box are more memorable than the faces in the red box.

b) A chart showing the distribution of the face space in terms of HR and FAR, with the boxes showing the extremes from which the random faces of a) were taken.

c) The 25 random split-half reliability for each quadrant (the percent agreement of each quadrant between two random halves of the population, over 25 repetitions).

Table 1

Comparison of demographics of the 10k US Adult Faces Database, Amazon Mechanical Turk workers in Experiments 1 and 2, and the US Census

	10k US Adult Faces Database	AMT Workers (Experiment 1)	AMT Workers (Experiment 2)	US Census (1990)
# of People	10,168	877	1,274	2.49 million
Median Age	30 – 45 years	29 years	30 – 45 years	32.8 years
Gender				
Male	57.1%	43.0%	49.8%	48.7%
Female	42.9%	55.9%	50.2%	51.3%
Race				
White	83.7%	77.5%	80.1%	80.3%
Black	9.9%	9.0%	8.3%	12.1%
Asian	3.1%	5.5%	5.9%	2.9%
Hispanic	3.2%	3.1%	4.8%	(9.1%)*
Other	---	5.0%	0.8%	3.9%

Notes. Demographics for the 10k US Adult Faces Database were determined by an AMT demographics study involving twelve workers per face. AMT worker demographics were assembled from demographics surveys attached to the main tasks of Experiments 1 and 2. The racial, age, and gender distributions are very similar across the four different samples of people. * The 1990 US Census asks about Hispanic origin as a separate question from race, so there is likely overlap with other races.

Table 2

Multiple linear regressions run on logit-transformed hit rate (HR) scores.

HR

	Attribute	HR β	HR t	HR p		HR β	HR t	HR p
Oosterhof & Todorov attributes	interesting / **boring**	-0.01	-0.43	0.67	Oosterhof & Todorov attributes only	0.21	11.01	0.00
	calm / **aggressive**	0.01	0.34	0.74		0.01	0.25	0.80
	caring / cold	-0.05	-1.07	0.29		-0.05	-1.00	0.32
	confident / uncertain	0.02	0.67	0.50		0.03	1.17	0.24
	humble / **egotistic**	-0.01	-0.40	0.69		-0.02	-0.72	0.47
	emotionally stable / unstable	0.00	-0.08	0.94		-0.08	-3.15	1.70E-03
	intelligent / unintelligent	-0.08	-3.41	7.00E-04		5.56E+12	0.43	0.67
	sociable / introverted	0.05	1.47	0.14		0.02	0.69	0.49
	kind / **mean**	0.12	2.73	0.01		0.10	2.39	0.02
	responsible / irresponsible	-0.13	-4.38	0.00		-0.13	-4.14	0.00
	trustworthy / untrustworthy	0.09	2.38	0.02		0.06	1.66	0.10
	attractive / unattractive	-0.06	-2.96	3.10E-03		-0.05	-2.72	0.01
	happy / **unhappy**	-0.11	-2.34	0.02		-0.09	-2.08	0.04
	normal / **weird**	0.02	0.73	0.46		-5.56E+12	-0.43	0.67
Vokey & Read	**typical** / atypical	-0.09	-4.92	0.00				
	memorable / forgettable	0.18	9.69	0.00				
	familiar / unfamiliar	-0.03	-2.07	0.04				
Prelim. Study	**common** / uncommon	-0.05	-3.09	2.00E-03				
	emotional / unemotional	0.05	2.73	0.01				
	friendly / unfriendly	0.01	0.15	0.88				
	Constant	0.07	7.13	0.00	**Constant**	0.07	6.72	0.00

Model stats			**Model stats**	
F(2201,20)	35.06		F(2207, 14)	28.58
p	1.30E-116		p	4.87E-70
SSE	496.43		SSE	555.89
SSR	158.15		SSR	100.78
Adj r^2	0.235		Adj r^2	0.145

Notes. This table shows the relationships of various z-score transformed attributes with logit-transformed HR scores. Attributes are grouped based on origin – the first fourteen are from Oosterhof & Todorov (2008), the next three are from Vokey & Read (1992), and the last three are significant attributes that were found in a preliminary study. The attributes are presented here as "positive trait / negative antonym," with the attribute used from the original literature in bold. A multiple linear regression was run on all twenty attributes (left), and a second multiple linear regression was also run on only the fourteen Oosterhof & Todorov attributes (right). The *β* weight, *t* statistic, and corresponding *p-value* are reported for each attribute. Colored cells indicate cells with significant values, with darker cells significant at $p < 0.01$, lighter cells at $p < 0.05$, and white cells are non-significant. Cell coloring corresponds to the direction of significant values as well as the valence, with green indicating positive values (aligning with the positive attribute in green) and red indicating negative ones (aligning with the negative attribute in red).

Table 3

Multiple linear regressions run on logit-transformed false alarm rate (FAR) scores.

FAR

	Attribute	FAR β	FAR t	FAR p	FAR β (Oosterhof & Todorov attributes only)	FAR t (Oosterhof & Todorov attributes only)	FAR p (Oosterhof & Todorov attributes only)
Oosterhof & Todorov attributes	interesting / **boring**	-0.06	-1.40	0.16	-0.28	-9.89	0.00
	calm / **aggressive**	0.00	-0.01	1.00	-0.06	-1.53	0.13
	caring / cold	0.07	1.97	0.05	0.02	0.28	0.78
	confident / uncertain	-0.17	-4.07	0.00	0.10	2.38	0.02
	humble / **egotistic**	-0.02	-0.61	0.54	0.09	2.27	0.02
	emotionally stable / unstable	0.00	-0.02	0.99	-0.03	-0.73	0.46
	intelligent / unintelligent	0.13	2.91	3.70E-03	9.44E+12	0.48	0.63
	sociable / introverted	-0.14	-2.55	0.01	0.08	1.73	0.08
	kind / **mean**	0.15	5.08	0.00	-0.07	-1.11	0.27
	responsible / irresponsible	0.28	3.86	1.00E-04	0.16	3.48	5.00E-04
	trustworthy / untrustworthy	0.10	2.44	0.01	-0.10	-1.77	0.08
	attractive / unattractive	0.09	3.41	7.00E-04	0.21	7.96	0.00
	happy / **unhappy**	-0.07	-2.26	0.02	0.19	2.86	4.30E-03
	normal / **weird**	-0.03	-0.31	0.75	-9.44E+12	-0.48	0.63
Vokey & Read	**typical** / atypical	-0.11	-3.21	1.40E-03			
	memorable / forgettable	-0.05	-0.68	0.50			
	familiar / unfamiliar	0.08	3.38	7.00E-04			
Prelim. Study	**common** / uncommon	0.06	1.66	0.10			
	emotional / unemotional	-0.08	-2.61	0.01			
	friendly / unfriendly	0.06	2.42	0.02			
	Constant	-1.99	-127.32	0.00	-1.99	-124.09	0.00

Model stats			**Model stats**	
	F(2201,20)	22.76	F(2207, 14)	22.82
	p	1.34E-75	p	1.73E-55
	SSE	1190.10	SSE	1254.40
	SSR	246.08	SSR	181.53
	Adj r^2	0.164	Adj r^2	0.121

Notes. This table shows the relationships of various z-score transformed attributes with logit-transformed FAR scores. See the caption for Table 2 on how this Table is organized.