Journal of Experimental Psychology: Learning, Memory, and Cognition

The Memorability of People: Intrinsic Memorability Across Transformations of a Person's Face

Wilma A. Bainbridge

Online First Publication, December 12, 2016. http://dx.doi.org/10.1037/xlm0000339

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000339

The Memorability of People: Intrinsic Memorability Across Transformations of a Person's Face

Wilma A. Bainbridge
Massachusetts Institute of Technology

When encountering new people for a brief instant, some seem to last in our memories while others are quickly forgotten. *Memorability*—whether a stimulus is likely to be later remembered—is highly consistent across different group of observers; people tend to remember and forget the same face images. However, is memorability intrinsic to just the picture of a face, or to a person's identity, generalizable across views and emotions? Thousands of participants completed an online experiment testing face identity recognition over five different emotional and viewpoint transformations (neutral, happy, angry, 3/4 view, and profile view). Memorability was found to be highly consistent within each image, as well as across transformations—if a face was remembered in one image, it was also likely to be remembered in another. Most other face attributes, including what participants *thought* would be memorable, did not show consistency within an identity. Overall, these results support the existence of memorability as a uniquely intrinsic, core attribute to a person, stable across images.

Keywords: memorability, face memorability, identity, face recognition

Supplemental materials: http://dx.doi.org/10.1037/xlm0000339.supp

There are certain people that, upon your first meeting, you remember forever. For others, it can be a struggle to remember them, even after meeting them dozens of times. If you were to meet the two people in Figure 1, chances are that you will remember Person A and forget Person B. These two faces differ in their *memorability*—a predictive value of whether a stimulus is likely to be later remembered or forgotten (Isola, Xiao, Torralba, & Oliva, 2011), based on average performance in a large-scale memory study. Memorability is linked to the perceptual distinctiveness of a face (Bartlett, Hurry, & Thorley, 1984; Light, Kayra-Stuart, & Hollander, 1979; Vokey & Read, 1992), and faces that match the observer's demographics tend to be remembered more easily (e.g., the own-race effect; Chiroro & Valentine, 1995). Despite individual observer variability in memory performance (Burton, 2013), memorability is found to be a highly consistent trait over images (Isola et al., 2011)—people tend to remember the same scene images (Isola, Xiao, Parikh, Torralba, & Oliva, 2014), face images (Bainbridge, Isola, & Oliva, 2013), and charts and graphs (Borkin et al., 2013). Memorability persists over different time scales (Isola et al., 2014), is robust to different contexts (Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015), and can be predicted by computational models (Khosla, Raji, Torralba, & Oliva, 2015).

Wilma A. Bainbridge is supported by the Department of Defense, through the National Defense Science & Engineering Graduate Fellowship Program. Many thanks to Aude Oliva, for her mentorship and support, as well as to Santani Teng and Caitlin Mullin for their help in greatly improving the manuscript.

Correspondence concerning this article should be addressed to Wilma A. Bainbridge, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, MIT 32-D430, 32 Vassar Street, Cambridge, MA 02139. E-mail: wilma@mit.edu

However, in the real world, we almost never see the exact same view of a person twice. It is unclear to what degree the face memorability effects found previously (Bainbridge et al., 2013) are *image*-specific versus *identity*-specific. This is an important distinction, as these two alternatives have different implications for our understanding of memorability as a perceptual attribute, as well as what attributes are fundamental properties of a face. Previous work has only tested memorability for repeated images (Bainbridge et al., 2013; Borkin et al., 2013; Isola et al., 2011). It is thus possible that stable memorability of an image across observers only occurs in specific test cases where the same image is repeated, but cannot be generalized to the real world, where faces are constantly changing. If memorability is only *image-specific,* then memorability effects are likely to be driven by a combination of low-level visual features that exist in an image but are sensitive to transformations in viewpoint or expression. Memorability would thus be an early perceptual feature that changes constantly as faces change dynamically in real life. In contrast, if memorability is *identity-specific,* then memorability is likely to be a higher-level perceptual feature with invariance to changes in the face. Rather than being an emergent property of image features, memorability would be a part of the "essence," or set of fundamental properties, of a face or entity. This would have large implications for perception (e.g., what are key traits we pick up on when perceiving a face?), memory (e.g., to what degree do stimulus-driven vs. observer-driven factors influence memory of a face?), as well as the transition between these two processes—this high-level perceptual feature that determines memory encoding may serve as a key to the intersection between these often separately studied phenomena (Bussey & Saksida, 2007). This distinction of image-specific versus identity-specific can also be easily applied to any other face attribute (e.g., attractiveness, trustwor-

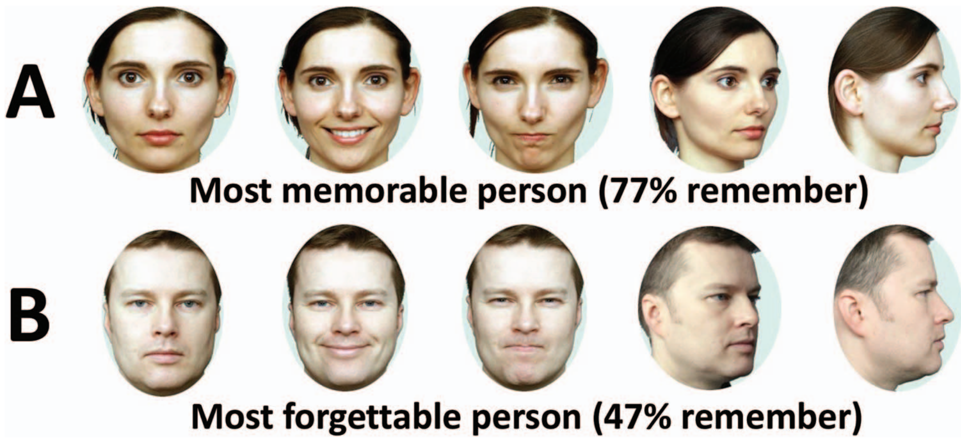

Figure 1. The most memorable (A) and forgettable (B) people found in this study. Images were printed with signed consent. See the online article for the color version of this figure.

thiness, subjective memorability), to see which attributes also persist within a person across changes.

Previous work on unfamiliar identity recognition might predict that memorability should be image-specific. While some works have found that people are able to successfully recognize identity across transformations such as emotion or viewpoint changes (D'Argembeau, Van der Linden, Comblain, & Etienne, 2003; Moses, Ullman, & Edelman, 1996; Patterson & Baddeley, 1977), many have found greatly diminished performance for identity recognition of novel faces (Bruce, 1982; Hancock, Bruce, & Burton, 2000; Hill, Schyns, & Akamatsu, 1997; Jenkins, White, Montfort, & Burton, 2011). Unfamiliar face recognition is a difficult task even when matching simultaneously presented images from the same identity (Henderson, Bruce, & Burton, 2001; Megreya & Burton, 2006). Additionally, perceptions of other face attributes such as dominance and trustworthiness are known to be influenced by changes within a face, such as emotional expression (Montepare & Dobish, 2003; Oosterhof & Todorov, 2009). Given that memory performance is low and highly variable for unfamiliar identity recognition, it seems unlikely for memorability to stay consistent within an identity, across large changes in viewpoint and emotional expression. On the other hand, perhaps observers behave similarly in which identities are harder or easier to recognize. This alternate hypothesis would predict that the memorability ranking of an identity is still preserved across different images—where some people are still more easily remembered than others.

The current study investigates the degree to which memorability ranking persists across image changes in a person's face. We find that memorability is indeed a stable, intrinsic attribute of an identity, across both viewpoint and expression changes. Interestingly, we find that among a comprehensive set of face attributes, the only other attribute showing such consistency within identity is *attractiveness,* highlighting memorability as a uniquely intrinsic attribute to a facial identity.

Method

Stimuli

The stimuli for this experiment come from two face image databases: the Karolinska Directed Emotion Faces (KDEF) database (Lundqvist, Flykt, & Öhman, 1998) and the Stirling/Economic & Social Research Council (ESRC) 3-Dimensional Face Database (Hancock, 2011). These two databases were selected because they contain photographs of different viewpoints and facial expressions for a set of unfamiliar noncelebrities. These images of real people were used to avoid potential issues of highly artificial stimuli (Burton, 2013). However, for this first study of memorability consistency, it was important to have images with consistent background and lighting to avoid observer dependence on non-face-related pictorial cues (Longmore, Liu, & Young, 2008), and to include transformations that were consistent, and could thus be compared, across identities.

A unified stimulus set was created from these two image sets, consisting of 131 people, 67 from the KDEF and 64 from the Stirling/ESRC 3D Face Database, and with 5 images per person (see Figure 1): *neutral* (forward-facing and neutral expression), *happy* (forward-facing and happy expression), *angry* (forward-facing and angry expression), *3/4* (right 3/4-view and neutral expression), and *profile* (right profile view and neutral expression). All faces were autoleveled for color and cropped with an oval to minimize background. No significant differences (all $p > .05$) were found between results for images from the two different databases for any analyses in this paper, providing evidence that camera and other low-level visual differences are unlikely to affect identity memorability rankings.

The images in the stimulus set were selected to have as homogeneous demographics as possible, to diminish effects such as the other-race effect (Chiroro & Valentine, 1995). Of the 131 images used, 68 (51.9%) were female and 63 (48.1%) were male. As the KDEF only contains white faces, and in order to diminish race effects, only white faces were used in this experiment. A separate set of 86 Amazon Mechanical Turk (AMT) workers were asked to judge the age of each face (12 raters per face). The average approximate age of the stimulus faces was 29.6 years (SD = 8.1), close to the average age of participants (M = 33.2 years, SD = 10.7), suggesting relatively accurate estimates (Rhodes, 2009).

Participants

A total of 1,579 workers from the crowd-sourcing website AMT completed this study, and data were collected following the standards of the Massachusetts Institute of Technology (MIT) Institutional Review Board. Only AMT workers with an approval rate of at least 95% and a U.S. Internet Protocol (IP) address were allowed to participate in the study, to have demographics as similar to the stimulus set as possible. Originally, 5,566 workers began the study; however 3,987 workers quit partway through or failed to complete the practice session or vigilance tests. As a note, a low participation rate is not rare among AMT experiments, especially as this experiment included countermeasures to reject cheaters (Berinsky, Huber, & Lenz, 2012; Eickhoff & de Vries, 2013; Isola et al., 2011). Ultimately, 59.8% of participants were female, while 40.2% were male. In terms of race, 82.6% were white, 6.6% were black, 5.4% were Asian, 4.2% were Hispanic, and 1.2% marked "other." The average memory test participant age was 33.2 years (SD = 10.7). Before the main experiment, participants were given a practice round of 30 face images not used in the main stimulus set to learn the paradigm. In total, the experiment took approximately 5 min per participant, and participants received compensation for completing the study.

Identity Memory Game

Memorability was assessed for the stimulus images and identities by conducting multiple parallel *Identity Memory Games* on AMT, similar in method to Isola et al. (2011)'s Image Memory Game (see Figure 2a).

In this identity memory game, participants saw a stream of 150 images and were told to press a button when they saw a repeat of the same identity, regardless of whether the image was different. Images were presented for 800 ms and were separated by a fixation cross displayed for 1.2 s. Before running the experiment, a subset of 32 stimuli were randomly selected to be "target" images (17 female, 15 male; 13 from KDEF, 19 from the Stirling/ESRC 3D Face Database), while the remaining 99 were assigned as "filler." In the memory game, the first target image presentations happened

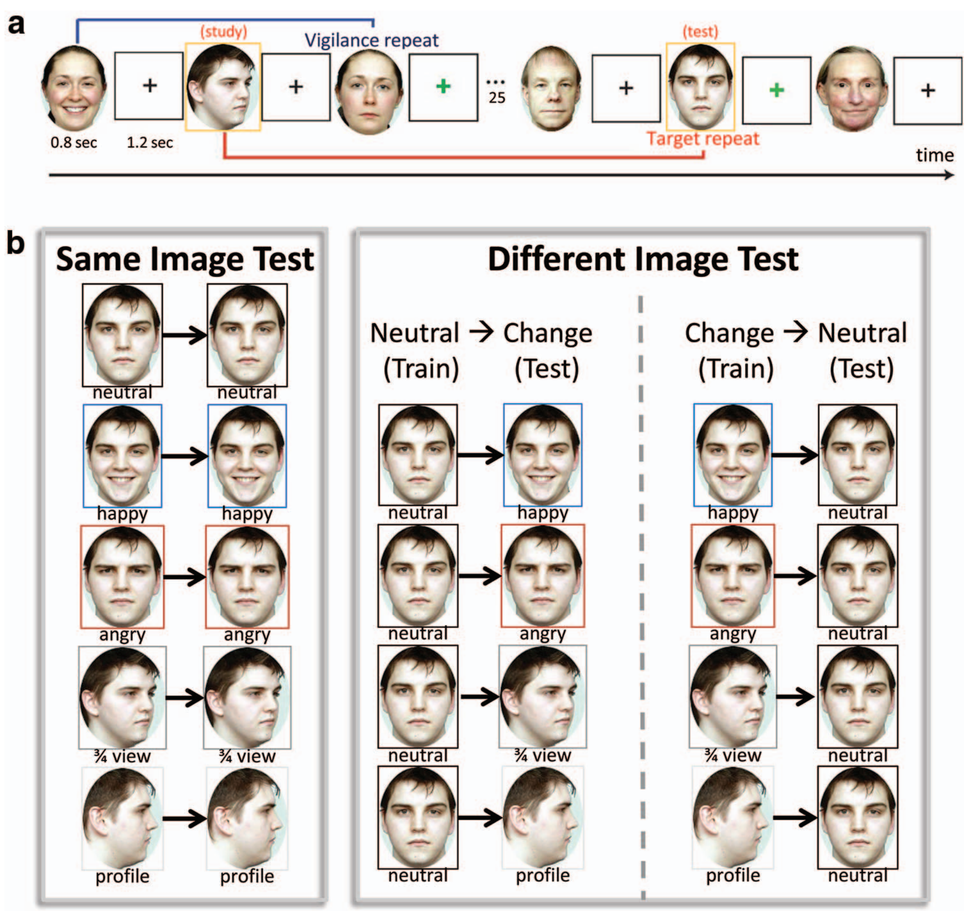

Figure 2. An illustration (a) of the methods in the Identity Memory Game. Participants had to press a key when they saw the same person repeated in a stream of images, even with a different picture. Vigilance repeats were used to ensure participants were correctly performing the task, while performance on the target repeats was used to measure memorability. A diagram (b) of the 13 tested transformations on the target faces. The first picture indicates the trained image, and the second picture indicates the tested image. The Same Image Test had vigilance repeats of the same image only, to get image-specific memorability scores. The Different Image Test had both vigilance and target repeats using different images for the same person, as illustrated in (a). See the online article for the color version of this figure.

at different, randomized (per participant) points in the experiment. Target images were then repeated approximately 25 images after their first presentation. Filler images were placed between targets and also used as "vigilance" repeats, occurring 1–7 images apart and happening in approximately 1/6 of the filler images, to test that participants were fully attending to the experiment. If participants incorrectly identified more than 50% of the last 10 vigilance repeats or made more than 50% false alarms on the last 30 nonrepeat images, then the game was ended and they were paid for their time until that point. As participants only did the game a single time, they were not aware of the distribution of target and filler images throughout the task. Image ordering was pseudorandom and varied for each participant.

The memory game was run in two versions—the "same image test" where memory for repeated images was tested to get baseline scores, and the "different image test" where trained and tested images differed for the same identity—resulting in 13 memorability values per identity (see Figure 2b for all combinations of train/test pairs). Within each memory game of either test type, participants saw a mix of image types (i.e., neutral, happy, angry, 3/4 view, profile view) among both the target and filler images. However, for the same image tests, repeats (both target and vigilance repeats) only occurred with the same image of the same identity. In contrast, for the different image tests, repeats (target and vigilance) only occurred with two different images of the same identity. Each AMT worker only participated in a single memory game (either the same image or different image test), and thus saw only one image pair per identity. Each of the 416 pairs of images (13 image pairs × 32 identities) was viewed by on average 89 observers each, close to the number of observers (approximately 80) found in previous work to result in stable consistency scores (Bainbridge et al., 2013; Isola et al., 2011).

From these identity memory games, each target image pair resulted in a false alarm rate (FAR) to the first image and a hit rate (HR) to the pair. For the current study, we define *memorability* as the HR of an image (Bainbridge et al., 2013; Isola et al., 2011). As the relationship between HR and FAR in terms of memorability is debated (Deffenbacher, Johanson, Vetter, & O'Toole, 2000; Megreya & Burton, 2007; Vokey & Read, 1992), we also include analyses for FAR and d'.

Attribute Labeling

A second set of 2,511 participants on AMT were asked to make subjective judgments for each face image on a comprehensive set of 20 personality and memory-related attributes. These attributes were compiled from face attribute and recognition literature (Oosterhof & Todorov, 2008; Vokey & Read, 1992), and used in prior work on face image memorability (Bainbridge et al., 2013). These attributes included: *attractive, unhappy, sociable, emotionally stable, mean, boring, aggressive, weird, intelligent, confident, caring, egotistic, responsible, trustworthy, (subjectively) memorable, typical, familiar, common, emotional, and friendly*. Participants were asked to rate a given face image for each of these attributes on a 9-point Likert-type scale, ranging from 1 (*not at all*) to 9 (*extremely*). Attributes were presented in a randomized order for every survey, to eliminate ordering effects. A random half of the participants rated the faces for the above attributes, while the other half rated the faces for their antonyms (and these scores were subtracted from 10), to diminish word valence biases. The survey also included a catch question asking participants to indicate a number displayed on the screen (a random integer from 1 to 9), in order to eliminate participants who responded at random. Ultimately, 49 participants' data were removed for failing this catch question. As with the image memory game, only AMT workers with an approval rate of at least 95% and a U.S. IP address were allowed to participate in the study. Participants were allowed to complete multiple face images (randomized across the 5 image types), but were not shown multiple images from the same identity. Each of the 160 images (32 identities × 5 images) had on average 34 participants rate it for the 20 attributes.

Results

Mean HR, FAR, and d' can be seen for each image pair type in Table 1 (and for each individual image in the online supplemental material: S1, S2, S3). On average, target image pairs were correctly recognized in 61.8% of trials ($SD = 11.8\%$), with an average performance of 69.6% ($SD = 11.5\%$) for same-image pairs and 57.0% ($SD = 11.9\%$) for different-image pairs. These image pairs showed a broad spread of HRs (min = 22.9%, max = 96.4%) with a normal distribution (see online supplemental material S4), indicating that there is a wide range of performances with both highly memorable and highly forgettable images. Interestingly, most image pairs with the neutral image first did not have a significantly different HR from the neutral-neutral image pairs (neutral-angry: $p > .1$; neutral-3/4: $p > .1$; neutral-profile: $p > .5$); however, all image pairs with a transformed (i.e., viewpoint or emotion change) image first had significantly lower HRs than neutral-neutral (happy-neutral: $t(62) = 2.11$, $p = .038$; angry-neutral: $t(62) = 2.76$, $p = .008$; 3/4-neutral: $t(62) = 5.21$, $p = 2.31 \times 10^{-6}$; profile-neutral: $t(62) = 8.06$, $p = 3.13 \times 10^{-11}$). These results show an interesting asymmetry in memory performance based on the type of image that is studied or tested (see online supplemental material S5). While this does not have direct implications on memorability, these may indicate a memory performance bias for first studying the prototypical (Bruce, Doyle, Dench, & Burton, 1991) or averaged face (Burton, Jenkins, Hancock, & White, 2005), and could warrant future investigation. For FAR (see Table 1), on average participants had false alarms to 23.3% of trials ($SD = 11.1\%$), with an average of 20.4% ($SD = 10.7\%$) for same-image pairs and 25.1% ($SD = 11.4\%$) for different-image pairs. It similarly shows a wide range (min = 4.2%, max = 35.3%), with a distribution characteristic of FAR histograms (see online supplemental material S4).

Is Memorability Consistent Within Image Pairs?

Hit rate consistency analysis. One first essential question is whether observers act similarly within study-test image pairs: if Participant A is successfully able to remember the angry version of a face after learning the neutral version, is Participant B able to as well? Within-pair consistency across observers was calculated for the HRs of each of the 13 study-test target image pairs. The consistency analysis was conducted using a 1,000-iteration random split-half rank correlation analysis, as established in Bainbridge et al. (2013) and Isola et al. (2011). Essentially, the participants were randomly split in equal halves, the average HR of all of the images

Table 1

Mean (M), Standard Deviation (SD), Average 1000-Iteration Split-Half Consistency (r), and Significance Versus a Chance Distribution (p) for Hit Rate (HR), False Alarm Rate (FAR), and d′ for Each Image Pair

Image pair	HR				FAR				d′			
	M	*SD*	*r*	*p*	*M*	*SD*	*r*	*p*	*M*	*SD*	*r*	*p*
Neutral-neutral	.67	.14	.71	.000	.18	.09	.70	.000	1.47	.57	.65	.000
Happy-happy	.70	.11	.71	.000	.17	.10	.75	.000	1.62	.56	.66	.000
Angry-angry	.73	.08	.56	.002	.22	.12	.76	.000	1.51	.58	.74	.000
3/4–3/4	.70	.13	.76	.000	.21	.13	.80	.000	1.47	.55	.56	.002
Profile-profile	.67	.12	.64	.000	.25	.10	.71	.000	1.21	.56	.53	.001
Neutral-happy	.58	.11	.62	.000	.23	.10	.76	.000	.98	.45	.57	.000
Neutral-angry	.62	.12	.66	.000	.25	.10	.71	.000	1.05	.33	.51	.002
Neutral-3/4	.63	.12	.59	.000	.22	.09	.65	.000	1.16	.50	.52	.004
Neutral-profile	.65	.07	.18	.31	.24	.09	.66	.000	1.15	.31	.21	.22
Happy-neutral	.59	.16	.78	.000	.27	.13	.78	.000	.95	.72	.74	.000
Angry-neutral	.58	.14	.67	.001	.30	.14	.78	.000	.81	.60	.69	.000
3/4-neutral	.50	.13	.64	.000	.26	.13	.81	.000	.72	.55	.59	.000
Profile-neutral	.42	.11	.68	.000	.25	.12	.77	.000	.53	.53	.65	.000
Same image average	.70	.12	.68		.20	.11	.74		1.46	.56	.63	
Different image average	.57	.12	.60		.25	.11	.74		.92	.50	.56	
Overall average	.62	.12	.63		.23	.11	.74		1.13	.52	.59	

Note. All were significantly above chance based on a Nonparametric Permutation Test ($p < .05$), except for the neutral-profile image pairs for HR and d'.

rated by one half was ranked, and the same images rated by the second half of participants was sorted along the same ordering. A Spearman rank correlation was calculated between these two rankings, and these randomized split halves were repeated 1,000 times, with the average correlation coefficient being the final consistency. If any ties occurred, their ranks were averaged. A higher consistency indicates that images tend to result in the same memory performance across different observers of the image. Chance levels (i.e., if no consistency exists) were approximated by randomly permuting the images 1,000 times across random ranks and correlating their HR rankings with those of the first split-half. If participants show no similarity in their memory behavior, then consistency should be similar to this chance level. *p* values here reflect the results of nonparametric permutation tests (i.e., the proportion of random permutations that show a consistency higher than the average split-half consistency).

Consistency was high for HR (Table 1, Figure 3), with an average consistency of $r = .68$ for the same-image task and $r = .60$ for the different-image task, in comparison to a consistency of $r = .68$ previously found for same-image face memorability (Bainbridge et al., 2013). Almost all consistencies were significantly different from chance ($p < .005$), indicating that participants have highly consistent memory performance, even when training and testing on different images of the same person. The single exception is when participants studied the neutral, frontal face and were tested on the profile face, where consistency is low and nonsignificant ($r = .18$). This image pair hints toward the boundaries at which memorability information may break down. Prior work has proposed that there is minimal shared information between forward-facing and profile views of a face (McKone, 2008; Verfaillie & Boutsen, 1995), and so it is possible that memorability is also not transferred across these views.

Note that this measure of consistency is independent of actual memory performance—observers can be consistent in being unable to recognize a person, as well as be consistent in remembering a person. For example, even though the average memory performance for the 3/4-neutral image pair is low (HR of 50%; Table 1), this image pair still shows significant consistency, $r = .64$, $p \sim 0$. While prior power analyses have shown that consistency estimates stabilize at approximately 60–80 participants (Isola et al., 2014), a subsample analysis here found that the consistency patterns were still largely preserved at smaller sample sizes. For this analysis, over 100 iterations, a random 40% of the participants were selected, and then the consistency analysis was run with them (i.e., instead of 80 observers per image, there were now approximately 34 observers per image). The consistencies over these 100 iterations were then averaged. Almost all image pairs for HR remained significantly consistent within this limited subsampled space ($p < 0.05$), with the exceptions of neutral-profile, neutral-3/4, and angry-angry (see online supplemental material S6). However, the magnitude of the consistencies generally decreased with this analysis, indicating that a higher number of participants results in a more stable consistency estimation, which eventually plateaus around 80 observers per image (Isola et al., 2014).

False alarm rate and d' consistency analysis. The same patterns emerge for FAR and d' as with HR. Unlike HR, FAR is determined on the first presentation of the image, so the image pairing combinations do not affect FAR. FAR consistency was significant for all image pairs (Table 1, Figure 3), with the lowest consistency as $r = .64$, and the average consistency was $r = .74$, higher than the average FAR consistency of $r = .69$ previously found for face images (Bainbridge et al., 2013). Sensitivity index d', combining HR and FAR, also shows significantly high consistency for almost all image pairs (Figure 3, Table 1), with an average consistency of $r = .59$. As with HR, the only image pair that shows a nonsignificant consistency is the neutral-profile image pair, with a consistency of $r = .22$. Subsampling the data to only 34 observers per image also found that consistency patterns were largely unaffected (see online supplemental material S6). While the magnitudes of consistency decreased with the smaller

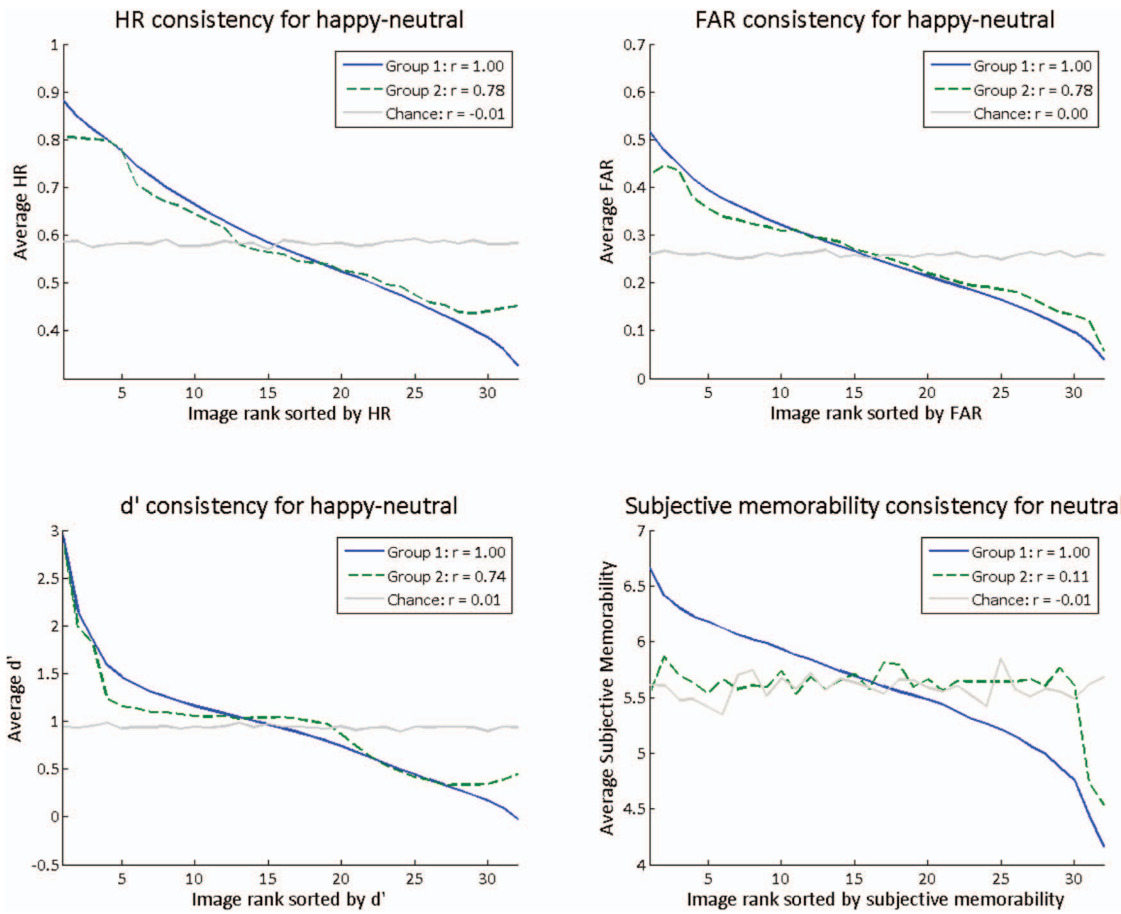

Figure 3. Chart of the consistency analyses for happy-neutral image pairs. For each, the solid blue line shows the images ranked for the random split-half Group 1, while the dashed green line shows Group 2's rankings of the same images (both averaged across 1,000 iterations). The light gray line shows an estimation of chance, by shuffling the rank orders of the second group. Here, *r* values indicate the Spearman's rank correlation value with Group 1. HR, FAR, and d' show significantly high correlations between Group 1 and Group 2 (the blue and green lines are largely overlapping), indicating consistency of HR ($p \sim 0$), FAR ($p \sim 0$), and d' ($p \sim 0$) in the happy-neutral image pair. However, subjective memorability (participants' ratings of which images they think they will remember) shows no correlation between Group 1 and Group 2 ($p = .44$; the dashed green line is more similar to the gray chance distribution than the blue line), indicating there is no consistency of subjective memorability to neutral images. See the online article for the color version of this figure.

number of observers, all image pairs remained significantly consistent for FAR, and almost all image pairs remained significantly consistent for d', with the exceptions of neutral-profile and neutral-happy. These results indicate that different metrics of memory (HR, FAR, d') all show high consistency across observers, even when they are performing identity recognition with different images.

Face attribute consistency analysis. Do other face attributes show the same degree of consistency as memorability? A similar split-half consistency analysis was conducted for each of the face attributes, except for single images instead of image pairs (as the attributes are based on ratings of a single image, while memorability scores are based on memory performance on an image pair). A summary of the attribute consistencies within the different image types can be seen in Table 2 (attribute means and standard deviations in online supplemental material S7). The attributes that show significant within-image consistency over all image types are: *egotistic, attractive, friendly,* and *happy*. However, most attributes, including those that might be tied to memorability—for example, *typical*, *subjectively memorable*, *familiar*—do not show consistency across observers (Table 2, Figure 3).

Is Memorability Ranking Preserved Across Image Pairs?

Hit rate ranking analysis. Memorability appears to be consistent within image pairs, but importantly, is it also consistent across image pairs of the same identity? That is, if a given face is the most memorable in a set of faces when neutral, is it also the most memorable when happy? Within each image pair set (e.g., neutral-happy, where the forward-facing neutral image is studied, and the happy image is tested), each identity's memorability ranking was determined, and then the ranking list was compared with the baseline neutral-neutral ranking list using a Spearman rank

Table 2
Average 1000-Iteration Split-Half Consistency (r) and Significance Versus A Chance Distribution (p) for Each of the Twenty Image Attributes

	Neutral		Happy		Angry		3/4		Profile	
Attribute	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Typical/atypical	0.20	0.263	0.21	0.226	0.25	0.189	0.34	0.071	0.13	0.491
Interesting/boring	*0.38*	*0.032*	*0.39*	*0.029*	0.27	0.141	*0.39*	*0.027*	0.30	0.095
Calm/aggressive	*0.44*	*0.014*	0.13	0.48	*0.75*	*0.000*	*0.50*	*0.002*	0.31	0.096
Caring/cold	*0.47*	*0.005*	0.31	0.085	*0.54*	*0.000*	*0.46*	*0.012*	*0.50*	*0.003*
Common/uncommon	0.28	0.119	0.16	0.362	0.32	0.073	0.07	0.693	−0.03	0.881
Confident/uncertain	0.30	0.093	*0.45*	*0.014*	0.27	0.138	0.36	0.039	0.20	0.264
Humble/egotistic	**0.35**	**0.038**	**0.44**	**0.014**	**0.44**	**0.007**	**0.62**	**0.000**	**0.36**	**0.046**
Emotionally stable/unstable	*0.42*	*0.011*	0.25	0.183	*0.62*	*0.000*	0.33	0.058	0.29	0.122
Memorable/forgettable	0.14	0.441	0.12	0.498	*0.37*	*0.038*	*0.38*	*0.035*	0.12	0.515
Intelligent/unintelligent	0.34	0.069	0.32	0.071	*0.61*	*0.000*	*0.48*	*0.004*	0.15	0.391
Sociable/introverted	*0.51*	*0.004*	*0.64*	*0.000*	0.31	0.088	*0.46*	*0.007*	*0.50*	*0.003*
Kind/mean	*0.55*	*0.003*	0.10	0.599	*0.64*	*0.000*	*0.58*	*0.000*	*0.36*	*0.044*
Responsible/irresponsible	*0.51*	*0.002*	0.18	0.335	*0.55*	*0.000*	*0.47*	*0.008*	*0.43*	*0.017*
Trustworthy/untrustworthy	*0.45*	*0.009*	−0.02	0.931	*0.56*	*0.001*	*0.41*	*0.011*	*0.44*	*0.012*
Attractive/unattractive	**0.53**	**0.005**	**0.70**	**0.000**	**0.66**	**0.000**	**0.59**	**0.001**	**0.65**	**0.000**
Emotional/unemotional	−0.13	0.480	*0.42*	*0.011*	*0.63*	*0.000*	−0.04	0.844	0.32	0.057
Familiar/unfamiliar	*0.39*	*0.031*	0.16	0.362	0.03	0.859	−0.01	0.957	0.24	0.184
Friendly/unfriendly	**0.51**	**0.001**	**0.45**	**0.015**	**0.67**	**0.000**	**0.49**	**0.004**	**0.36**	**0.046**
Happy/unhappy	**0.53**	**0.002**	**0.57**	**0.000**	**0.63**	**0.000**	**0.56**	**0.001**	**0.60**	**0.001**
Normal/weird	*0.39*	*0.037*	*0.41*	*0.017*	*0.54*	*0.002*	0.19	0.28	*0.34*	*0.050*

Note. Italicized data indicate a significantly high consistency for that attribute and image type ($p < 0.05$). Rows of bolded data indicate an attribute with significantly high within-image consistency for all image types.

correlation. Essentially, this determines if a given identity has a preserved memorability ranking across different images of that identity. Note that HR itself is not expected to remain constant across transformations, as viewpoint and expression changes have been shown to produce diminished memory (Bruce, 1982; Hancock et al., 2000; Hill et al., 1997; Jenkins et al., 2011); it is the consistency of its *ranking* that is being compared.

Almost every image pair set correlated significantly with the neutral-neutral ranking list (Figure 4; see online supplemental material S10 for correlations between all image pairs). Specifi-

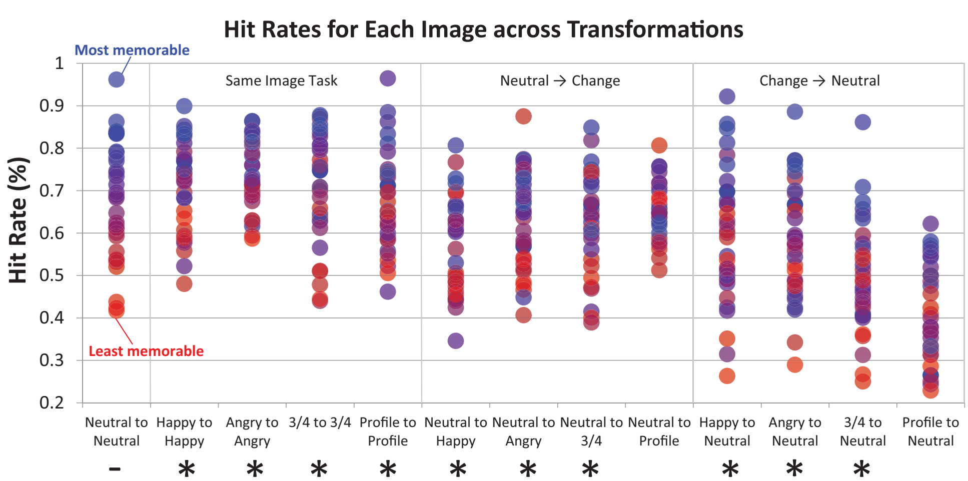

Figure 4. A chart of the hit rates (HRs) for every identity across the different experimental image transformations, showing how memorability ranking is preserved across these transformations. Each neutral-neutral image pair is colored by ranking, with the most memorable in blue and the least memorable in red. The HR for each identity is shown for all possible transformations, with circles of a given color representing a single identity, colored by its neutral-neutral ranking. The blue dots (i.e., the most memorable neutral-neutral identities) generally stay at the top across transformations, while the red dots generally stay at the bottom across transformations, indicating that memorability ranking is preserved. These results are confirmed statistically using Spearman's rank correlations against the neutral-neutral rankings. Asterisks indicate significant correlations at the level of $p < .05$. See the online article for the color version of this figure.

cally, neutral-neutral rankings were significantly correlated with happy-happy (r = .63, p = .0002), angry-angry (r = .49, p = .004), 3/4–3/4 (r = .52, p = .002), profile-profile (r = .41, p = .018), neutral-happy (r = .35, p = .050), neutral-angry (r = .50, p = .004), neutral-3/4 (r = .50, p = .004), happy-neutral (r = .52, p = .002), angry-neutral (r = .46, p = .009), and 3/4-neutral (r = .47, p = .007). The only exceptions were transformations between profile views (neutral-profile, profile-neutral), which were not significantly correlated in ranking (neutral-profile: $p > .8$, profile-neutral: $p > .05$), similar to the within image pair consistency analysis.

False alarm rate and d' ranking analyses. FAR and d' ranking are also generally preserved across these different views (see online supplemental material S8, S9, S11, S12). For FAR, the ranking of neutral images is significantly correlated with that of all image types: happy images (r = .70, p = .00001), angry images (r = .66, p = .00007), 3/4 view images (r = .66, p = .00006), and profile images (r = .55, p = .001). For d', the ranking of neutral-neutral image pairs is significantly correlated with happy-happy (r = .37, p = .036), angry-angry (r = .52, p = .002), 3/4–3/4 (r = .46, p = .008), profile-profile (r = .36, p = .042), neutral-happy (r = .35, p = .048), neutral-angry (r = .39, p = .028), happy-neutral (r = .40, p = .022), angry-neutral (r = .46, p = .008), and 3/4-neutral (r = .36, p = .041). There are, however, nonsignificant positive correlations for d' with some image pairings: neutral-3/4, neutral-profile, and profile-neutral ($p > .05$). Overall, these results indicate that in addition to HR, FAR and d' are generally consistent within an individual, in terms of both within-pair consistency and across-transformation consistency.

Face attribute ranking analyses. The same analysis was conducted on the comprehensive set of 20 attributes, to see if identity ranking within an attribute was preserved across different images of an identity (e.g., if someone is rated to be the most attractive when neutral, will they also be the most attractive when happy?). Some attributes did show a significantly preserved ranking across images, namely *intelligent, responsible, trustworthy, attractive,* and *normal*. Additionally, *calm* is significantly preserved across all images except the profile view. However, most attributes, and notably memory-related attributes such as *subjectively memorable* and *familiar,* did not show a preserved ranking across images (see Table 3). The only attribute showing significant consistency in both the within-image pair consistency analysis and across-image pair ranking analysis is *attractiveness,* presenting evidence for it as an intrinsic trait to an identity like memorability.

Table 3
Average Ranking Consistency (r) and Significance (p) for Each of the Twenty Image Attributes

Attribute	Neutral-Happy		Neutral-Angry		Neutral-3/4		Neutral-Profile	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Typical/atypical	−.02	.925	.12	.520	.17	.348	.26	.159
Interesting/boring	*.49*	*.005*	.09	.628	.34	.060	.09	.636
Calm/aggressive	*.48*	*.006*	*.44*	*.012*	*.60*	*.000*	.35	.051
Caring/cold	.33	.061	.14	.432	*.44*	*.013*	*.55*	*.001*
Common/uncommon	.16	.393	.17	.342	*.50*	*.004*	.06	.738
Confident/uncertain	.31	.081	*.39*	*.027*	*.65*	*.000*	*.45*	*.010*
Humble/egotistic	.33	.064	*.59*	*.000*	*.69*	*.000*	*.61*	*.000*
Emotionally stable/unstable	.30	.094	.19	.290	*.52*	*.002*	.14	.446
Memorable/forgettable	.18	.334	.29	.105	.34	.058	−.06	.729
Intelligent/unintelligent	**.59**	**.000**	**.51**	**.003**	**.45**	**.010**	**.38**	**.033**
Sociable/introverted	*.37*	*.036*	.33	.068	*.45*	*.010*	*.46*	*.008*
Kind/mean	.31	.082	*.39*	*.027*	*.66*	*.000*	*.59*	*.000*
Responsible/irresponsible	**.56**	**.001**	**.57**	**.001**	**.43**	**.014**	**.57**	**.001**
Trustworthy/untrustworthy	**.53**	**.002**	**.48**	**.005**	**.45**	**.010**	**.54**	**.001**
Attractive/unattractive	**.53**	**.002**	**.74**	**.000**	**.64**	**.000**	**.57**	**.001**
Emotional/unemotional	−.19	.289	.22	.218	.25	.174	.24	.194
Familiar/unfamiliar	.24	.195	−.01	.944	*.41*	*.019*	.09	.609
Friendly/unfriendly	.19	.298	.16	.388	*.58*	*.001*	*.36*	*.046*
Happy/unhappy	*.46*	*.008*	.18	.318	*.37*	*.036*	*.46*	*.008*
Normal/weird	**.40**	**.022**	**.55**	**.001**	**.55**	**.001**	**.61**	**.000**

Note. Italicized data indicate a significantly high consistency for that attribute and image type ($p < .05$). Rows of bolded data indicate an attribute with significantly high across-image (within identity) ranking consistency for all image types.

Discussion

Memorability as an Essential, Intrinsic Characteristic

These results show that beyond memorability of single images, *people* can be memorable or forgettable. Memorability is highly consistent both within and across different images of a person, and memorability ranking remains preserved even across changes in both emotional expression as well as viewpoint. This high consistency exists in true memories (HRs) and false memories (FARs), as well as the combined metric of d'. This consistency exists during both memory encoding and retrieval, even when performed on different images of the same person. This consistency also exists despite wide variance in memory performance over these views and the difficulty of unfamiliar identity recognition (Bruce, 1982; Hancock et al., 2000; Hill et al., 1997; Jenkins et al., 2011).

Interestingly, memorability is unique in that most other attributes studied in personality and memory attribute research (Oosterhof & Todorov, 2008; Vokey & Read, 1992) do not show a similar level of consistency within and across images of the same identity. Even attributes thought to be related to memorability (e.g., typicality, familiarity, or what observers think is memorable) do not show this level of consistency, nor do attributes with a long legacy of study in face perception (e.g., trustworthiness, responsibility). The only attribute showing a consistency similar to memorability is attractiveness, which is thought to be highly dependent on similarity to the average face (Langlois & Roggman, 1990) and may elicit stereotyped neural responses (Aharon et al., 2001). However, while attractiveness may also be an intrinsic property to a face identity, it is a separate property from memorability, as attractiveness has found to be only weakly correlated with memorability (Bainbridge et al., 2013; Isola et al., 2011), and in the current study, there were no significant correlations between rated attractiveness and HR for all of the image pairs (all $p > .10$). These results thus provide evidence that memorability is a unique intrinsic attribute of an identity, not solely a simple combination of other well-characterized face attributes. Indeed, previous work has found that less than half of the variance of memorability can be accounted for by this comprehensive attribute set (Bainbridge et al., 2013). Next important steps will be to identify what specific features of an image or identity are used to calculate memorability, and to be able to make precise predictions of memory performance based on characteristics of both the image and the observer.

In some cases, memorability was less clearly preserved across extreme view changes (i.e., when generalizing from the forward-facing view to profile view and vice versa). This presents a first step of quantifying the "confidence interval" (Burton, 2013) of memorability—while memorability is generally consistent, to what degree can it be altered with transformations to the face or image? Previous work in computer vision has been able to manipulate the memorability of a face image using high-level image features (Khosla, Bainbridge, Torralba, & Oliva, 2013)—similarly, would it be possible to manipulate the memorability of an identity using makeup? Also, at what point does the consistency of memorability break down—what sorts of changes in lighting, clothing, hair, and so forth cause an identity to have a different memorability, and how might this apply to the natural variations we see in real people as they change their appearance from day to day (e.g., like the naturalistic unfamiliar identity recognition task in Jenkins et al., 2011)?

While the current study helps to further characterize memorability, there are still many large, open questions about what memorability *is*. A current working hypothesis is that it is a high-level perceptual property reflecting the statistical distinctiveness of an entity in comparison to our visual world. This statistical distinctiveness may function on a multidimensional set of axes—while one image may be memorable because of a low-level visual property like large Euclidean distances of face features from the mean (Busey, 2001), another may be memorable for a high-level property like seeming extremely untrustworthy (Rule, Slepian, & Ambady, 2012). Recent neuroimaging work has found that processing of memorability in the brain is a rapid perceptual process with consistent patterns of neural activity across participants, and different neural substrates from low-level vision and memory encoding (Khaligh-Razavi, Bainbridge, Pantazis, & Oliva, submitted). Rapidly identifying the memorability of an object may be crucial to processing our world, so we can quickly identify novel (and potentially threatening) people or places that we will need for later memory, or so that we can prioritize storage or processing of items that should be remembered over those that should be forgotten. The current work shows us that memorability is not only an emergent property of specific images, but is an intrinsic property of entities, invariant to image changes such as viewpoint, as well as face changes such as emotional expression. Based on the current study, most other face-related attributes do not share this intrinsic nature to an identity. These results thus support the role of memorability as an ecologically important, intrinsic metric of an identity. Of course, further work will be necessary to fully quantify the role of memorability in perception and memory, as well as identify how stereotyped memorability processing develops in spite of our diverse visual experiences. Recent work has found memorability measures to generalize to other time lags (Isola et al., 2014) as well as to other memory tasks such as directed forgetting or differential encoding depth tasks (Bainbridge, submitted). Additional work will be needed to generalize memorability to other memory paradigms such as alternative forced choice or free recall, as well as other types of memory such as semantic versus episodic, or iconic versus long-term memory.

Applications of Identity Memorability

Taken together, these results indicate that there is an intrinsic memorability to a person, or entity. The ultimate memory performance of an observer given an image can be thought of as a conjunction of several factors building upon this intrinsic memorability, including psychological attributes of the stimulus (Bainbridge et al., 2013), image features (Isola et al., 2011), surrounding image context (Bylinskii et al., 2015), viewer differences (Chadwick, Bonnici, & Maguire, 2014), and viewing environment and timing (Isola et al., 2014). Further understanding of both intrinsic memorability and these factors will allow increasingly precise predictions of memorability and individual memory performance. While the current study uses fairly homogenous stimuli, increasingly realistic and natural stimuli would also allow us to examine the "confidence interval" (Burton, 2013) of how memorability is affected by each factor.

This new understanding of memorability as intrinsic to identities and entities allows several new applications of memorability. Identity memorability may be a useful tool to consider in crime-related applications (MacLin & MacLin, 2004; Molinaro, Arndorfer, & Charman, 2013; Valentine, 2014), to improve eyewitness identification (e.g., by understanding memorability biases a suspect's face may have) and inform criminal lineup selection (e.g., by matching people in a lineup for memorability). Dynamic memorability prediction algorithms (Khosla et al., 2015) could assess memorability of an actor based on an audition video, or help a businessperson study clients' faces that they are likely to forget. Knowing the memorability of specific faces could serve as a benchmark against which to compare memory performance and neural signatures for patient populations with impairments in face perception (e.g., prosopagnosia; Duchaine & Nakayama, 2006), memory (e.g., Alzheimer's disease; Della Sala, Muggia, Spinnler, & Zuffi, 1995), or social cognition (e.g., autism spectrum disorder; Boucher & Lewis, 1992). For example, do these populations show the same consistencies in memorability ranking? Memorability could also guide memory aides—forgettable faces could be caricaturized (Rhodes, Brennan, & Carey, 1987) to enhance memory. We must also uncover the limits of *manipulating* memorability (Khosla et al., 2013)—to what degree can this core, intrinsic memorability of a person be changed to make oneself, for an instant, a memorable spokesperson or a forgettable spy.

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,* 537–551. http://dx.doi.org/10.1016/S0896-6273(01)00491-3

Bainbridge, W. A. (submitted). *The resiliency of memorability: A predictor of memory separate from attention and priming*. Manuscript submitted for publication.

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142,* 1323–1334. http://dx.doi.org/10.1037/a0033872

Bartlett, J. C., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12,* 219–228. http://dx.doi.org/10.3758/BF03197669

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical

Turk. *Political Analysis, 20,* 351–368. http://dx.doi.org/10.1093/pan/mpr057

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics, 19,* 2306–2315. http://dx.doi.org/10.1109/TVCG.2013.234

Boucher, J., & Lewis, V. (1992). Unfamiliar face recognition in relatively able autistic children. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 33,* 843–859. http://dx.doi.org/10.1111/j.1469-7610.1992.tb01960.x

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73,* 105–116. http://dx.doi.org/10.1111/j.2044-8295.1982.tb01795.x

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38,* 109–144. http://dx.doi.org/10.1016/0010-0277(91)90049-A

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology, 66,* 1467–1485. http://dx.doi.org/10.1080/17470218.2013.800125

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51,* 256–284. http://dx.doi.org/10.1016/j.cogpsych.2005.06.003

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* (pp. 147–192). Mahwah, NJ: Erlbaum.

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. http://dx.doi.org/10.1002/hipo.20320

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

Chadwick, M. J., Bonnici, H. M., & Maguire, E. A. (2014). CA3 size predicts the precision of memory recall. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 10720–10725. http://dx.doi.org/10.1073/pnas.1319641111

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: Human Experimental Psychology, 48,* 879–894.

D'Argembeau, A., Van der Linden, M., Comblain, C., & Etienne, A. (2003). The effects of happy and angry expressions on identity and expression memory for unfamiliar faces. *Cognition and Emotion, 17,* 609–622. http://dx.doi.org/10.1080/02699930302303

Deffenbacher, K. A., Johanson, J., Vetter, T., & O'Toole, A. J. (2000). The face typicality-recognizability relationship: Encoding or retrieval locus? *Memory & Cognition, 28,* 1173–1182. http://dx.doi.org/10.3758/BF03211818

Della Sala, S., Muggia, S., Spinnler, H., & Zuffi, M. (1995). Cognitive modelling of face processing: Evidence from Alzheimer patients. *Neuropsychologia, 33,* 675–687. http://dx.doi.org/10.1016/0028-3932(95)00009-R

Duchaine, B. C., & Nakayama, K. (2006). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology, 16,* 166–173. http://dx.doi.org/10.1016/j.conb.2006.03.003

Eickhoff, C., & de Vries, A. P. (2013). Increasing cheating robustness of crowdsourcing tasks. *Information Retrieval, 16,* 121–137. http://dx.doi.org/10.1007/s10791-011-9181-9

Hancock, P. J. B. (2011). *Stirling/ESRC 3D Face Database*. Retrieved from http://pics.stir.ac.uk/ESRC/

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4,* 330–337. http://dx.doi.org/10.1016/S1364-6613(00)01519-9

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology, 15,* 445–464. http://dx.doi.org/10.1002/acp.718

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62,* 201–222. http://dx.doi.org/10.1016/S0010-0277(96)00785-8

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36,* 1469–1482. http://dx.doi.org/10.1109/TPAMI.2013.200

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 145–152. http://dx.doi.org/10.1109/CVPR.2011.5995721

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121,* 313–323. http://dx.doi.org/10.1016/j.cognition.2011.08.001

Khaligh-Razavi, S.-M., Bainbridge, W. A., Pantazis, D., & Oliva, A. (submitted). *From what we perceive to what we remember: Characterizing representational dynamics of visual memorability*. Manuscript submitted for publication.

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),* Sydney, New South Wales, Australia.

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. http://dx.doi.org/10.1109/ICCV.2015.275

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1,* 115–121. http://dx.doi.org/10.1111/j.1467-9280.1990.tb00079.x

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,* 212–228. http://dx.doi.org/10.1037/0278-7393.5.3.212

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 77–100. http://dx.doi.org/10.1037/0096-1523.34.1.77

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces—KDEF* [CD ROM]. Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, Solna Municipality, Sweden.

MacLin, M., & MacLin, O. (2004). The effect of criminality on face attractiveness, typicality, memorability, and recognition. *North American Journal of Psychology, 6,* 145–154.

McKone, E. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 310–327. http://dx.doi.org/10.1037/0096-1523.34.2.310

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition, 34,* 865–876. http://dx.doi.org/10.3758/BF03193433

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & Psychophysics, 69,* 1175–1184. http://dx.doi.org/10.3758/BF03193954

Molinaro, P. F., Arndorfer, A., & Charman, S. D. (2013). Appearance-change instruction effects on eyewitness lineup identification accuracy are not moderated by amount of appearance change. *Law and Human Behavior, 37,* 432–440. http://dx.doi.org/10.1037/lhb0000049

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27,* 237–254. http://dx.doi.org/10.1023/A:1027332800296

Moses, Y., Ullman, S., & Edelman, S. (1996). Generalization to novel images in upright and inverted faces. *Perception, 25,* 443–461. http://dx.doi.org/10.1068/p250443

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 11087–11092. http://dx.doi.org/10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion, 9,* 128–133. http://dx.doi.org/10.1037/a0014520

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 406–417. http://dx.doi.org/10.1037/0278-7393.3.4.406

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19,* 473–497. http://dx.doi.org/10.1016/0010-0285(87)90016-8

Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology, 23,* 1–12. http://dx.doi.org/10.1002/acp.1442

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition, 125,* 207–218. http://dx.doi.org/10.1016/j.cognition.2012.06.017

Valentine, T. (2014). Estimating the reliability of eyewitness identification. In T. Perfect & D. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 579–595). London, UK: Sage. http://dx.doi.org/10.4135/9781446294703.n32

Verfaillie, K., & Boutsen, L. (1995). A corpus of 714 full-color images of depth-rotated objects. *Perception & Psychophysics, 57,* 925–961. http://dx.doi.org/10.3758/BF03205454

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20,* 291–302. http://dx.doi.org/10.3758/BF03199666

Received December 28, 2015
Revision received August 9, 2016
Accepted August 12, 2016 ■