



HHS Public Access

Author manuscript

Comput Human Behav. Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

Comput Human Behav. 2015 November 1; 52: 331–337. doi:10.1016/j.chb.2015.06.018.

Artificial faces are harder to remember

Benjamin Balas and Jonathan Pacella

Psychology Department, Center for Visual and Cognitive Neuroscience, North Dakota State University, Fargo, ND 58102

Abstract

Observers interact with artificial faces in a range of different settings and in many cases must remember and identify computer-generated faces. In general, however, most adults have heavily biased experience favoring real faces over synthetic faces. It is well known that face recognition abilities are affected by experience such that faces belonging to “out-groups” defined by race or age are more poorly remembered and harder to discriminate from one another than faces belonging to the “in-group.” Here, we examine the extent to which artificial faces form an “out-group” in this sense when other perceptual categories are matched. We rendered synthetic faces using photographs of real human faces and compared performance in a memory task and a discrimination task across real and artificial versions of the same faces. We found that real faces were easier to remember, but only slightly more discriminable than artificial faces. Artificial faces were also equally susceptible to the well-known face inversion effect, suggesting that while these patterns are still processed by the human visual system in a face-like manner, artificial appearance does compromise the efficiency of face processing.

Keywords

Artificial faces; face recognition; face memory

1. Introduction

High-quality artificial faces are increasingly prevalent in movies, video games, and other media. In most of these settings, these faces are created with the intent that observers will be able to perform the same recognition tasks with them that they can perform with real faces. For example, artificial faces should be easy to categorize as male or female. The race and age of an artificial face should also be recognizable. Artificial faces should be able to communicate emotion and other non-verbal social cues. Finally, high-level properties like personality traits and motivation should be recognizable in an artificial face. Understanding the differences in how real and artificial faces are processed by the visual system is critically important in developing synthetic agents that can achieve these goals. In some cases,

Corresponding Author: Benjamin Balas, 1210 Albrecht Blvd., North Dakota State University, Fargo, ND 58102, Phone: 701-231-6105, Benjamin.balas@ndsu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

artificial faces can dramatically fail to support performance in these tasks. For example, artificial faces can in some cases be perceived as “uncanny” (Mori, 1970; Segama & Nagayama, 2007), leading observers to feel disquieted by their appearance. Feelings of disquiet and disgust obviously compromise a synthetic face’s value for entertainment, etc. so understanding how it arises is important. Perceived uncanniness appears to depend on several specific features including the size of the eyes within an artificial face and the skin tone (MacDorman et al., 2009), suggesting some straightforward ways to avoid the “uncanny valley.” However, the fact that artificial faces can so dramatically fail to meet the standards of “face-ness” implemented in the human visual system suggests that real and artificial faces can in some cases be processed very differently (Green et al., 2008). While there have been significant technical advances in rendering high-quality synthetic faces that closely approximate the static features of human faces (e.g. skin-tone (Krishnaswamy & Baronovsky, 2004; Giard & Guitton, 2010) and their dynamic properties, there remain many open questions regarding how artificial faces are processed relative to real faces. Here, we chose to examine whether or not participants’ biased experience with real faces over the course of their lifetime led to measurable deficits in face memory and discrimination for artificial faces compared to real faces. That is, is the human visual system ‘tuned’ to real appearance in such a way that artificial faces are excluded from efficient or expert-like face processing?

Biased experience with faces is known to impact face recognition abilities. One of the most well-known examples of this phenomenon is the “other-race effect” (ORE). The ORE refers to observers’ generally superior abilities to recognize, remember, and distinguish between faces belonging to their own race or to the race they most encounter (Malpass & Kravitz, 1969). Other-race faces are generally harder to distinguish from one another than own-race faces and elicit different neural responses than own-race faces (Balas & Nelson, 2010; Wiese, Stahl & Schweinberger, 2009). Observers also appear to maintain a categorical boundary between own- and other-race faces (Levin & Beale, 2000), which leads to increased discriminability for faces that cross that boundary relative to faces that have roughly equal dissimilarity but lie within one category or the other. Other-race categorization also appears to be faster than own-race categorization (Levin, 1996), which may underlie differences in visual search for own-race vs. other-race faces (Sun et al., 2013). In sum, observers’ experience with primarily own-race faces leads to substantial differences in a range of recognition tasks, suggesting that exposure influences how broadly or narrowly face representations capture variability in appearance across ethnic and racial groups. Critically for the present study, similar differences in processing are evident when performance is compared across other face categories including those defined by age (Kuefner et al., 2008; Rhodes & Anastasi, 2012), and species (Taubert, 2009). We suggest, based on these results, that artificial faces may also constitute a class of “other-group” faces as well and may also be processed less effectively.

There is substantial evidence that the visual system is sensitive to the differences in appearance between real and artificial faces.. Observers are generally able to accurately label faces as real or artificial (Farid & Bravo, 2011; Balas & Tonsager, 2014), even when relatively high-quality synthetic faces are used. Moreover, like other-race faces, observers appear to maintain a category boundary between real and artificial faces (Looser &

Wheatley, 2010; Balas & Horski, 2012), such that gradual physical changes in appearance lead to a non-linear change in perceived category membership. Real vs. artificial appearance also modulates other face categorization tasks, suggesting that the artificial appearance of synthetic faces compromises or otherwise modulates a range of processes. For example, observers' ability to categorize faces according to sex is affected by artificial appearance such that artificial faces are perceived as more female and vice-versa (Balas, 2013). This result suggests that artificial appearance is "gendered" in a similar manner as race (Johnson, Freeman & Pauker, 2012) due to covariance between the visual features that define male/female categories and real/artificial categories. The interaction between real/artificial appearance and other face categories extends to neural responses as well. For example, the extent to which human/animal differences in neural processing are evident depends on real vs. artificial face appearance. Species effects on the N170 component (a face-sensitive neural response measured using EEG) are evident for real faces, but not artificial faces (Balas & Koldewyn, 2013), suggesting that species categorization is tuned to natural appearance. While the evidence for distinct neural processing of real and artificial faces is to some extent mixed (Wheatley et al., 2011; Looser, Guntupalli, & Wheatley, 2012; Balas & Koldewyn, 2013), real vs. artificial appearance appears to be a perceptually and neural real dimension of face variability (Koldewyn, Hanus, & Balas, 2014) that may lead to differential processing of synthetic faces as though they represent an "other-group" of faces.

In the present study, we examined whether or not artificial faces are processed less effectively than real ones by comparing observers' memory for real vs. artificial faces (Exp. 1) and by comparing discrimination abilities for real and artificial faces (Exp. 2). Critically, we used commercial software that made it possible to render artificial faces from photographs of real faces, ensuring that our stimuli were identical in terms of category membership and identity across our real and artificial appearance conditions. We hypothesized that artificial faces would be more poorly remembered than real faces, and also that discrimination performance may differ between stimulus categories as well. We discuss our results in the context of other reports describing the impact of expertise on face recognition, and the use of artificial faces in multiple domains.

2. Experiment 1

In our first experiment, we investigated whether artificial faces would be more poorly remembered than real faces, even when both sets of faces depicted the same individuals. If artificial faces do constitute an "out-group" relative to real faces, we would expect recall performance to be lower for these faces.

2.1 Methods

2.1.1 Subjects—We recruited 18 participants (11 female) to take part in Experiment 1. All participants were members of the NDSU undergraduate community and were between the ages of 19–26 years of age. All participants reported normal or corrected-to-normal vision and gave written informed consent prior to participating in the experiment.

2.1.2 Stimuli—We created images of real and artificial faces to use in Experiment 1 using a database of images depicting male and female undergraduate students. The original

photographs of these individuals were full-color images that were 720×480 pixels in size. Artificial faces were created using FaceGen, a commercial software package for creating and manipulating artificial faces that relies on a morphable model of appearance (Banz & Vetter, 1999). To create artificial faces, we imported each of the 90 faces selected for use in Experiment 1 into FaceGen using the PhotoFit tool. This tool requires the user to identify and match fiducial points on the original image (front and profile views) so that a least-squares fit to the estimated shape and pigmentation of the face can be achieved using the appearance model implemented in FaceGen. The resulting artificial face images were 300×300 pixels in size. To remove the external outline of the face, both real and artificial images were cropped so that only the internal features of the face were visible within a circular region in the middle of the face (Figure 1).

2.1.3 Procedure—After providing written consent to participate, each participant was told that they were going to take part in a face memory task. Participants were randomly assigned to either the Real or Artificial face condition. In both cases, the testing session began with a study phase during which participants saw each of 45 randomly chosen faces one at a time. Each face was presented on-screen for 2 seconds, and participants were asked to categorize each face by sex as quickly as possible while still being accurate. The order of faces presented during the task was pseudo-randomized for each participant.

Following the study phase, participants completed a test phase after a short delay of approximately 2–3 minutes. In this part of the task, participants were told that they would now see a total of 90 faces presented one at a time, half of which would be faces presented during the study phase and half of which would be new faces the participants had not seen before. Participants were asked to indicate whether each face was “old” or “new.” Each face remained on-screen until the participant made a response. Faces presented during the test phase were also presented in a pseudo-randomized order.

Participants completed both the study and test phase of the experiment in a darkened, sound-attenuated room. Stimuli were presented on a MacBook Pro laptop with a resolution of 1200×900 pixels. Participants were seated comfortably at a viewing distance of approximately 40cm, and faces subtended a visual angle of approximately 5 degrees of visual angle at this viewing distance. Head and eye movements were neither controlled nor recorded during completion of the task. All stimulus display and response collection routines were controlled by custom scripts written using the Matlab Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

2.2 Results

2.2.1 Study phase—We first examined participants’ performance in the study phase of Experiment 1, in which they were asked to categorize each face according to sex. For each participant, we calculated the proportion of correct responses as well as the median response time for correct responses. We found that participants who viewed artificial faces were significantly less accurate than participants who viewed real faces ($M_{\text{artificial}}=82\%$, $s.d. = 14\%$; $M_{\text{real}}=93\%$, $s.d.=4.2\%$, $t(20)=2.39$, $p=0.037$, 95% CI of the difference between means = $[2.2\% - 20\%]$, two-tailed, independent-samples, t-test). We also found that participants

who viewed artificial faces were significantly faster to respond correctly than participants who viewed real faces ($M_{\text{artificial}}=540\text{ms}$, $s.d.=160\text{ms}$; $M_{\text{real}}=746\text{ms}$, $s.d.=280\text{ms}$, $t(20)=2.18$, $p=0.043$, 95% CI of the difference between means = [7ms 408ms], two-tailed, independent-samples t-test). This suggests a possible speed-accuracy trade-off affecting sex categorization in this task, rather than a true difference in categorization performance. To address this issue, we chose to examine performance in the study phase using inverse efficiency scores (Townshend & Ashby, 1983), which combine response latency and accuracy. These scores did not significantly differ between groups ($t(20)=0.58$, $p=0.57$), suggesting there is no true difference in performance between the two categories. Rather, observers have a tendency to respond more quickly, but less accurately, to artificial faces.

2.2.2 Test Phase—To compare memory performance for real and artificial faces, we calculated both d' and the response criterion, c , for each of our participants in both groups. Both of these measures are commonly used to quantify behavior in the context of Signal Detection Theory (Green & Swets, 1966). In a typical signal detection task, participants must try and detect the presence or absence of a particular stimulus. In this case, we can consider a previously seen face to be the signal that observers must try to detect. During the test phase of our task, participants' responses will belong to one of four categories: (1) "Hits" - correctly labeled "old" items, (2) "Misses" - "old" items that observers failed to label as such. (3) "False Alarms" - new items that are incorrectly labeled as "old." And (4) "Correct Rejections" - correctly labeled new items. To compute d' and c , we use the proportion of hits made to "old" items (the hit rate) and the proportion of false alarms made to new items (the false alarm rate). In each case, the hit rate and the false alarm rate are z-scored to obtain the final value (see the equations below). Observers' sensitivity to the signal they are trying to detect is described by d' , which is large when observers obtain many hits and few false alarms. Observers' tendency to make one response more than another regardless of the stimulus (e.g. an overall bias to say "old" across all trials) is described by the response criterion, c .

$$d' = z(HR) - z(FAR)$$

$$c = -\frac{1}{2}(z(HR) + z(FAR))$$

Our analysis of d' values revealed a significant difference in face memory between our real and artificial participant groups ($t(16)=4.07$, $p=0.001$, 95% CI of the difference between means = [0.16 0.52], two-tailed independent samples t-test). Specifically, participants who saw real faces had higher d' scores ($M=0.71$, $s.d.=0.2$) than participants who saw artificial faces ($M=0.37$, $s.d.=0.15$), suggesting better memory for real faces. Both of these results differed significantly from chance performance, however (Real faces: $t(8)=10.48$, $p<0.001$; Artificial faces: $t(8)=7.45$, $p<0.001$, two-tailed single-sample t-tests), which demonstrates that participants were capable of doing the task in both cases. We also compared criterion values across groups, which revealed no difference between real and artificial faces ($t(16)=-.483$, $p=0.64$).

2.3 Discussion

The results of Experiment 1 demonstrated that observers do have poorer memory for artificial faces than real faces even when race, sex, and identity are matched between stimulus categories. While we observed no true difference in sensitivity to gender information during the study phase, we did find that d' values differed significantly during the test phase of the experiment.

There are several interesting features of the data that are worth commenting on in relation to prior work with real and computer-generated faces as well as prior results comparing in-group and out-group face recognition. First, the lack of an effect of real/artificial appearance is at odds with prior work demonstrating how sex categorization can be affected by artificial appearance (Balas, 2013) and related work demonstrating that other-race faces are categorized by sex less effectively than own-race faces (O'Toole et al., 1996). There are, however, several differences between the current experiment and prior work, including the appearance of the stimuli and a lesser emphasis on speeded responses in the current study. By giving participants in this study a long time to view the image as preparation for a subsequent memory test, we may have compromised our ability to see the effect of artificial appearance on sex categorization. In the context of the current results however, our lack of a difference in inverse efficiency scores for sex categorization across participant groups suggests that both groups of observers were engaged with the task, and thus that subsequent decreased memory performance for artificial faces was likely not simply a function of reduced attention to these faces during the study phase.

It is also interesting to see that we did not observe a difference in response criterion across participant groups. In prior reports describing differences in memory for own-race and other-race faces, one feature of the "other-race effect" is an increased tendency to report novel items as being familiar or 'old' (Meissner & Brigham, 2001), possibly resulting from a sense that other-race faces "all look alike," which leads novel faces to more closely resemble members of the study set. In some cases, however, other-race faces appear to suffer from the opposite problem: Faced with variability in lighting, view, and other ecologically relevant sources of appearance variability, other-race faces "all look different" even when they are multiple images of the same individual (Zhou, Laurence & Mondloch, 2014). Our data do not support either an increase or a decrease in response bias to artificial faces, suggesting that they don't function as an out-group in this sense.

Finally, we note that overall performance in our memory task was rather low. This may be the result of our use of a relatively large study set and also only providing participants with a relatively short amount of time to encode unfamiliar faces. Also, the lack of an external contour on our faces may have lowered performance, since the external contour plays a substantial role in the recognition of unfamiliar faces (Johnston & Edmonds, 2009).

In Experiment 2, we examined whether identity-matched real and artificial faces would also differ in terms of participant's ability to discriminate between different individuals in a matching task. If this were the case, participants' poorer memory for artificial faces may be the result of increased perceptual similarity between study and test faces in the artificial faces condition rather than an effect of artificial appearance on memory *per se*. Besides

measuring discriminability for real and artificial faces, we also chose to manipulate face orientation in Experiment 2 so that we could compare the size of the well-known face inversion effect (poorer recognition of upside-down faces -Yin, 1969; Rossion et al., 2000) in stimulus sets comprised of real and artificial faces. In some instances, researchers have used face inversion as a means of determining whether or not observers are using 'holistic' or 'configural' processing to encode facial appearance (Maurer, Le Grand, & Mondloch, 2002). More conservatively, inversion can be used as a tool to match a broad range of low-level image properties (luminance, contrast, color, e.g.) while disrupting observers' ability to efficiently use face-specific mechanisms that are tuned to upright faces. Performance with inverted faces thus may reflect observers' capabilities when only low-level mechanisms and more general mechanisms for object recognition are available, making these stimuli a useful control for raw image similarity, etc.

3. Experiment 2

In our second experiment, we examined whether or not real vs. artificial appearance led to a measurable difference in discriminability using a match-to-sample task. Also, to determine whether artificial faces were processed in the same manner as real faces, we compared performance with upright and inverted faces. Inversion typically reduces face recognition performance substantially (Yin, 1969) while generally not affecting performance with other object categories. The inversion effect has also been reported to be reduced when out-group faces are used, including faces belonging to different race categories (Balas & Nelson, 2010) and different species categories (Taubert, 2009). We therefore predicted that discrimination performance may be reduced for artificial faces, and that the impact of inversion may be less.

3.1 Methods

3.1.1 Subjects—We recruited a total of 12 participants from the NDSU undergraduate community to participate in this study. Participant age ranged from 19–23 years and all participants reported normal or corrected-to-normal vision. Written informed consent was obtained from all participants prior to the testing session.

3.1.2 Stimuli—The same images of real and artificial faces described in Experiment 1 were used in this task.

3.1.3 Procedure—Participants completed the task in two separate blocks, one in which real faces were presented and another in which artificial faces were presented. On each trial, participants were presented with a single face in the center of the screen for 250ms, and after a delay of 750ms two test faces (one of which matched the sample face) were presented to the left and right of center (Figure 3). Participants were asked to indicate which of these two faces matched the sample by pressing either the left or right shift key on the keyboard as quickly as possible. The test faces remained on screen until participants made a response. Within each block, the orientation of the sample and test faces was pseudo-randomized. Note however that within a trial, sample and test images always had the same orientation: Either all images were presented upright or all were presented inverted. Participants completed a total of 180 trials per condition, for a grand total of 720 trials across both

experimental blocks. Stimuli were presented according to the same display parameters described in Experiment 1.

3.2 Results

For each participant we computed the proportion of correct responses in each condition as well as the median response time for correct responses. In both cases, these values were submitted to a 2×2 repeated-measures ANOVA with face type (real or artificial) and image orientation (upright or inverted) as within-subjects factors.

3.2.1 Accuracy—Our analysis of participant accuracy revealed a significant main effect of face type ($F(1,11)=5.04$, $p=0.046$, partial eta-squared=0.31) such that real faces were matched more accurately than artificial faces ($M_{\text{real}}=92.5\%$; $M_{\text{artificial}}=91.0\%$). We also observed a significant main effect of orientation ($F(1,11)=51.7$, $p<0.001$, partial eta-squared=0.82), with higher accuracy for upright faces ($M_{\text{upright}}=93.8\%$) than inverted faces ($M_{\text{inverted}}=89.7\%$). The interaction between face type and orientation was not significant ($F(1,11)=0.195$, $p=0.67$). Average values in each condition are displayed in Figure 4.

3.2.2 Reaction Time—Our analysis of participant response latencies to correct responses revealed only a main effect of orientation ($F(1,11)=42.0$, $p<0.001$, partial eta-squared=0.79) such that participant's made faster correct responses to upright images ($M=725\text{ms}$) compared to inverted images ($M=815\text{ms}$). The main effect of face type did not reach significance ($F(1,11)=1.98$, $p=0.19$, partial eta-squared=0.15), nor did the interaction between these two factors ($F(1,11)=0.001$, $p=0.98$).

4. General Discussion

The results of Experiments 1 and 2 suggest that face memory and face discrimination abilities are both reduced for artificial faces. This is consistent with the hypothesis that artificial faces are to some extent an “other-group” face category like other-race and other-age faces. That is, observers' experience with real faces has led to a representation of face appearance that may exclude even realistic artificial faces from expert-like face processing. An important caveat, however, is that we observed no difference in the magnitude of the inversion effect in real and artificial faces. This suggests that artificial faces are still likely processed in a facelike way by the human visual system, in contrast to results with composite faces (faces made using forensic face production tools - Carlson et al., 2012). These composite faces appear to be processed differently than real faces in some contexts. Artificial faces are thus still perceived as faces, but deviate enough from natural appearance for memory and discrimination to be less efficient. We also note that the magnitude of the observed difference in discrimination performance that we observed in Exp. 2 for real and artificial faces is quite small (~1%) and thus the larger effect we observed in Experiment 1 likely is not solely the result of reduced discriminability. Instead, artificial faces may differ from real faces specifically in terms of how faces are represented in memory. This basic result raises several important questions and also has implications for a range of applications in which synthetic agents are used.

If experience is indeed the driving factor that determines observers' memory and discrimination abilities for artificial faces vs. real faces, it is interesting to consider both the time course of learning as well as the potential for training to change the effects of artificial appearance observed here. For other-race faces, experience-based effects on discrimination are not evident until late in the first year of life (Kelly et al., 2009) and before that, infants demonstrate equal competence with faces belonging to different race, species, and age groups. Differential processing of own- and other-race faces continues to develop into early and middle childhood as well, suggesting that there is some longer-term refinement of the representations that support face recognition (Balas, Piessig & Moulson, 2015). Does artificial face processing follow a similar developmental time course? To our knowledge there are few studies directly comparing infants' or children's perception of real faces and artificial faces. Examining how differential processing develops in childhood would be an intriguing way to determine how experience affects the processing of artificial faces. Similarly, quantifying observers' exposure to artificial faces and investigating the relationship between our effects and experience viewing synthetic agents in video games or movies would be a powerful means of establishing how experience changes the boundaries between own- and other-group faces in this context. Daily exposure to faces belonging to different categories does appear to modulate recognition behavior for other examples of other-group faces (Wiese, Komes, & Schweinberger, 2012), which may mean that individual differences in exposure to artificial faces in movies or video games similarly impacts the effects reported here. Training with other-race faces can also be effective in ameliorating the other-race effect (Young et al., 2011) and individuation training in particular appears to be a powerful means of influencing subsequent face perception skills in infancy (Scott & Monesson, 2009) and adulthood (Tanaka & Pierce, 2009). Therefore, we might expect that sufficient training or exposure to artificial faces could reduce differences in performance between real and artificial faces. Alternatively, enhancing the distinctiveness of artificial faces via caricaturing (Irons et al., 2014) could also lead to better performance and suggest ways to systematically change how face images are rendered for various applications so they will be remembered. Determining the most effective interventions to achieve such a result would likely reveal how experience is used to shape face and object representations and also lead to useful techniques for developing effective synthetic agents.

We also note that while other-race face recognition has important implications for how social and perceptual categories interact, it is unknown how artificial faces function as a social category. Social category membership can have profound effects on face perception. Indeed, other-group effects on face processing can be induced simply by introducing a social category (Bernstein, Young, & Hugenberg, 2007; Hourihan, Fraundorf & Benjamin, 2013). Randomly labeling faces as being fans of one sports team or another can lead to other-race effects on face recognition, for example. This suggests that facial appearance does not solely determine how a face is processed – social labels play an important role. Likewise, changing various aspects of appearance is typically sufficient to signal social category membership (Balas & Nelson, 2010). Social categories and the attitudes associated with them also impact representations of facial appearance (Dotsch et al., 2008) such that faces belonging to out-group categories may be represented in terms of an appearance model that emphasizes negative affect. What then are the consequences of artificial appearance on social

categorization and inferences about personality or behavior? Determining whether or not artificial appearance systematically biases judgments about personality or affect would both inform our understanding of how facial appearance is used to estimate social variables and reveal important consequences of using artificial faces in applications where observers interact with synthetic agents.

Finally, there are multiple aspects of face processing that would be intriguing to examine in the context of artificial face perception that would help us determine how these images are processed relative to real faces. For example, other-race faces appear to be less susceptible to the composite face effect (Michel et al., 2006), suggesting that holistic processing makes a smaller contribution to the recognition of out-group faces. Would this be the case for artificial faces? If so, it may suggest that different perceptual strategies are employed for artificial vs. real face processing. Similarly, further examination of how real vs. artificial appearance affects the perception of other categories like race and age may reveal important properties regarding the representation of appearance when experience with out-groups is limited.

We close by noting some key limitations of the current study. First, we have only considered one model of artificial appearance in the current study. While FaceGen images (and stimuli closely resembling them) are widely used in face recognition studies, we also acknowledge that our results may vary depending on the particular synthetic faces used for comparison to real face processing. Prior results with artificial faces have in some instances used dolls (Looser & Wheatley, 2010), cartoon faces (Chen et al., 2010), schematic faces (Wilson, Loffler & Wilkinson, 2002), as well as computer-generated faces like those used here. Determining how these different kinds of artificial appearance do or do not differentially affect visual processing is an important question for future work. Second, in both Experiments 1 and 2, participants could theoretically succeed at both tasks by using image-level analysis (matching pixels between sample and test images). Further examining the difference between real and artificial face processing when invariance to lighting, viewpoint, or expression is required would be a useful way of isolating the contribution of high-level processes from low-level pattern matching mechanisms that may have partially supported performance in the current task. Still, we argue that the current results contribute to our knowledge of how artificial appearance affects face recognition and suggest many avenues for further research.

Acknowledgments

This research was supported in part by NIGMS grant P20 103505. Thanks also to Dan Gu, Alyson Saville, and Jamie Schmidt for their help with technical support and participant recruitment.

References

1. Balas B. Biological sex determines whether faces look real. *Visual Cognition*. 2013; 21:766–788.
2. Balas B, Koldewyn K. Early visual ERP sensitivity to the species and animacy of faces. *Neuropsychologic*. 2013; 51:2876–2881.
3. Balas B, Horksi J. You can take the eyes out of the doll, but. *Perception*. 2012; 41:361–364. [PubMed: 22808589]

4. Balas B, Nelson CA. The role of face shape and pigmentation in other-race face perception: An electrophysiological study. *Neuropsychologia*. 2010; 48:498–506. [PubMed: 19836406]
5. Balas B, Peissig JJ, Moulson MC. Children (but not adults) judge own- and other-race faces by the color of their skin. *Journal of Experimental Child Psychology*. 2015
6. Balas B, Tonsager C. Face Animacy is Not All in the Eyes: Evidence from Contrast Chimeras. *Perception*. 2014; 43:355–367. [PubMed: 25109004]
7. Bernstein MJ, Young SG, Hugenberg K. Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*. 2007; 18:706–712. [PubMed: 17680942]
8. Blanz V, Vetter T. A Morphable Model for the Synthesis of 3D Faces. *SIGGRAPH '99 Conference Proceedings*. 1999:187–194.
9. Brainard DH. The Psychophysics Toolbox. *Spatial Vision*. 1997; 10:433–436. [PubMed: 9176952]
10. Carlson CA, Gronlund SD, Weatherford DR, Carlson MA. Processing differences between feature-based facial composites and photos of real faces. *Applied Cognitive Psychology*. 2012; 26:525–540.
11. Chen H, Russell R, Nakayama K, Livingstone M. Crossing the ‘uncanny valley’: adaptation to cartoon faces can influence perception of human faces. *Perception*. 2010; 39:378–386. [PubMed: 20465173]
12. Dotsch R, Wigboldus DHJ, Langner O, van Knippenberg A. Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*. 2008; 19:978–980. [PubMed: 19000205]
13. Farid H, Bravo MJ. Perceptual discrimination of computer generated and photographic faces. *Digital Investigation*. 2011; 8:226–235.
14. Giard F, Guitton MJ. Beauty or realism: The dimension of skin from cognitive sciences to computer graphics. *Computers in Human Behavior*. 2010; 26:1748–1752.
15. Green RD, MacDorman KF, Ho C-C, Vasudevan SK. Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior*. 2008; 24(5):2456–2474.
16. Hourihan KL, Fraundorf SH, Benjamin AS. Same faces, different labels: generating the cross-race effect in face memory with social category information. *Memory & Cognition*. 41:1021–1031. [PubMed: 23546969]
17. Irons J, McKone E, Dumbleton R, Barnes N, He X, Provis J, Ivanovici C, Kwa A. A new theoretical approach to improving face recognition in disorders of central vision: Face caricaturing. *Journal of Vision*. 2014; 14(2):1–29. 12.
18. Johnson KL, Freeman JB, Pauker K. Race is Gendered: How Covarying Phenotypes and Stereotypes Bias Sex Categorization. *Journal of Personality and Social Psychology*. 2012; 102:116–131. [PubMed: 21875229]
19. Johnston RA, Edmonds AJ. Familiar and unfamiliar face recognition: A review. *Memory*. 2009; 17:577–596. [PubMed: 19548173]
20. Kelly DJ, Liu S, Lee K, Quinn PC, Pascalis O, Slater AM. Development of the other-race effect in infancy: Evidence towards universality? *Journal of Experimental Child Psychology*. 2009; 104:105–114. [PubMed: 19269649]
21. Koldewyn K, Hanus D, Balas B. Visual adaptation of the perception of “life”: Animacy is a basic perceptual dimension of faces. *Psychonomic Bulletin & Review*. 2014; 21:969–975. [PubMed: 24323739]
22. Krishnaswamy A, Baronoski GVG. A Biophysically-based Spectral Model of Light Interaction with Human Skin. *Computer Graphics Forum*. 2004; 23:331.
23. Kuefner D, Macchi Cassia V, Picozzi M, Bricolo E. Do All Kids Look Alike? Evidence for an Other-Age Effect in Adults. *Journal of Experimental Psychology: Human Perception and Performance*. 2008; 34:811–817. [PubMed: 18665727]
24. Levin DT. Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1996; 22:1364–1382.
25. Levin DT, Beale JM. Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics*. 2000; 62:386–401. [PubMed: 10723217]
26. Looser CE, Wheatley T. The tipping point of animacy. How, when, and where we perceive life in a face. *Psychological science*. 2010; 21(12):1854–1862. [PubMed: 21097720]

27. Looser CE, Guntupalli JS, Wheatley T. Multivoxel patterns in face-sensitive temporal regions reveal an encoding schema based on detecting life in a face. *Social cognitive and affective neuroscience*. 2012
28. MacDorman K, Green R, Ho C, Koch C. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*. 2009; 25:695–710. [PubMed: 25506126]
29. Malpass RS, Kravitz J. Recognition of faces of own and other race. *Journal of Personality and Social Psychology*. 1969; 13:330–334. [PubMed: 5359231]
30. Maurer D, Le Grand R, Mondloch CJ. The many faces of configural processing. *Trends in Cognitive Sciences*. 2002; 6:255–260. [PubMed: 12039607]
31. Meissner CA, Brigham JC. Thirty years of investigating the own-race bias memory for faces: a meta-analytic review. *Psychology, Public Policy, & Law*. 2001; 7:3–35.
32. Michel C, Rossion B, Han J, Chung C-S, Caldara R. Holistic Processing is Finely Tuned for Faces of One's Own Race. *Psychological Science*. 2006; 17:608–615. [PubMed: 16866747]
33. Mori M. The uncanny valley. *Energy*. 1970; 7:33–5.
34. O'Toole AJ, Peterson J, Deffenbacher KA. An 'other-race effect' for categorizing faces by sex. *Perception*. 1996; 25:669–676. [PubMed: 8888300]
35. Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*. 1997; 10:437–442. [PubMed: 9176953]
36. Rhodes MG, Anastasi JS. The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological bulletin*. 2012; 138(1):146–174. [PubMed: 22061689]
37. Rossion B, Gauthier I, Tarr MJ, Despland B, Bruyer R, Linotte S, Crommelinck M. 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*. 2000; 11:69–74. [PubMed: 10683832]
38. Scott LS, Monesson A. The origin of biases in face perception. *Psychological Science*. 2009; 20:676–680. [PubMed: 19422630]
39. Seyama J, Nagayama RS. The uncanny valley: The effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*. 2007; 16(4):337–351.
40. Sun G, Song L, Bentin S, Yang Y, Zhao L. Visual search for faces by race: A cross-race study. *Vision Research*. 2013; 89:39–46. [PubMed: 23867566]
41. Tanaka JW, Pierce LJ. The neural plasticity of other-race face recognition. *Cognitive, Affective, and Behavioral Neuroscience*. 2009; 9:122–131.
42. Taubert J. Chimpanzee faces are 'special' to humans. *Perception*. 2009; 38:343–356. [PubMed: 19485131]
43. Townshend, JT.; Ashby, FG. *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press; 1983.
44. Wheatley T, Weinberg A, Looser C, Moran T, Hajcak G. Mind perception: real but not artificial faces sustain neural activity beyond the N170/VPP. *PloS one*. 2011; 6(3):e17960. [PubMed: 21483856]
45. Wiese H, Komes J, Schweinberger SR. Daily-life contact affects the own-age bias in elderly participants. *Neuropsychologia*. 2012; 50:3496–3508. [PubMed: 23022078]
46. Wiese H, Stahl J, Schweinberger SR. Configural processing of other-race faces is delayed but not decreased. *Biological Psychology*. 2009; 81:103–109. [PubMed: 19428974]
47. Wilson HR, Loffler G, Wilkinson F. Synthetic faces, face cubes, and the geometry of face space. *Vision Research*. 2002; 42:2909–2923. [PubMed: 12450502]
48. Young SG, Hugenberg K, Bernstein MJ, Sacco DF. Perception and Motivation in Face Recognition: A Critical Review of Theories of the Cross-Race Effect. *Personality and Social Psychology Review*. 2011; 16:116–142. [PubMed: 21878608]
49. Zhou X, Laurence S, Mondloch C. They all look different to me: Within-person variability affects identity perception for other-race faces more than own-race faces. *Journal of Vision*. 2014; 14(10): 1263.

Highlights

- We compared memory and discrimination for real and artificial faces.
- Artificial faces were identity-matched to real faces.
- Memory performance was poorer for artificial faces.
- Discrimination was also poorer, but this effect is smaller.

Original Face



CG Face



Figure 1.

An example of a real face used in the current study and its counterpart artificial face made using FaceGen. Both images have been cropped from a larger face image that depicted the outline of the face and head.

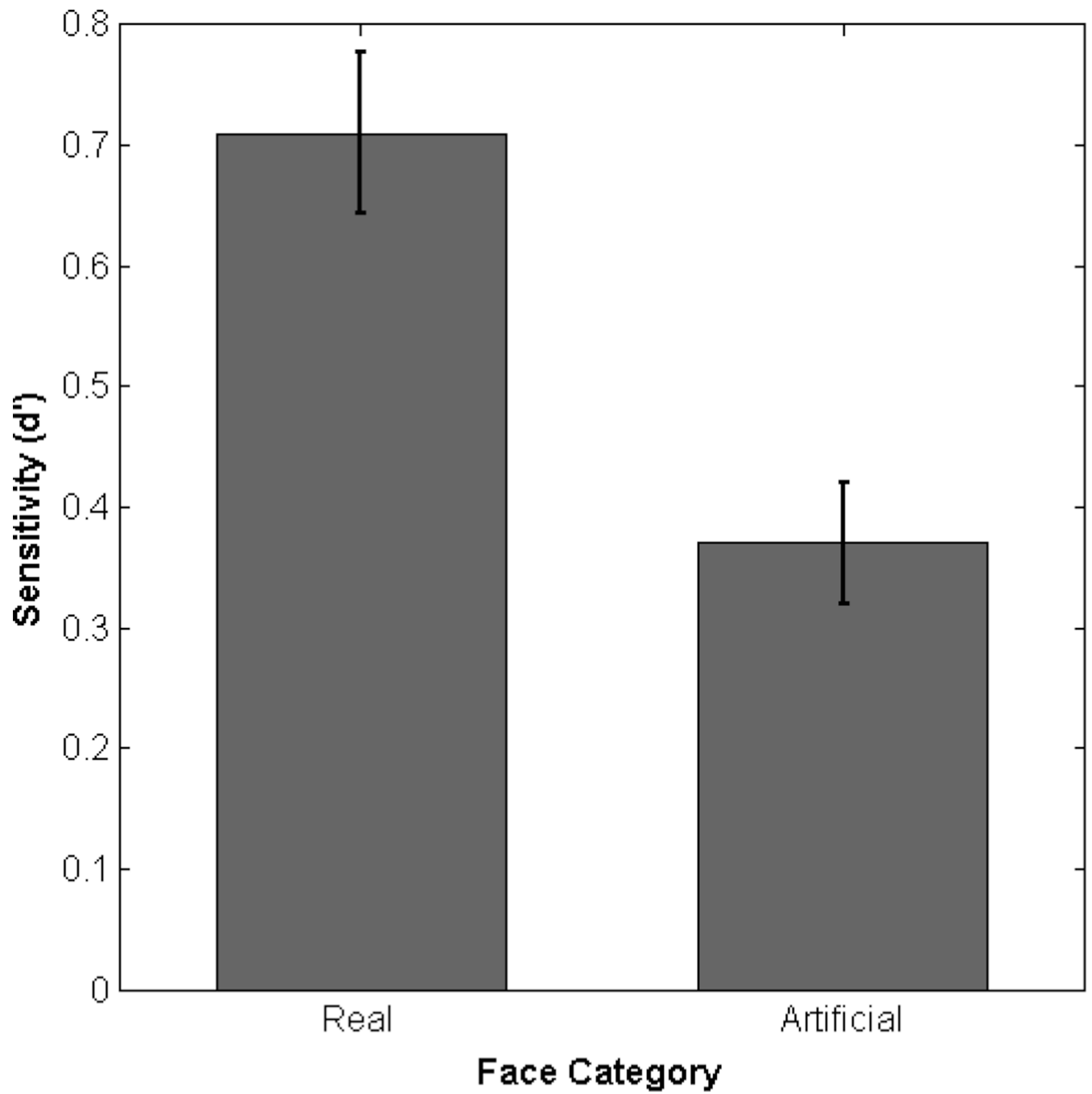


Figure 2. Average d' values for real and artificial faces. Participants were significantly worse at correctly remembering artificial faces relative to real ones.

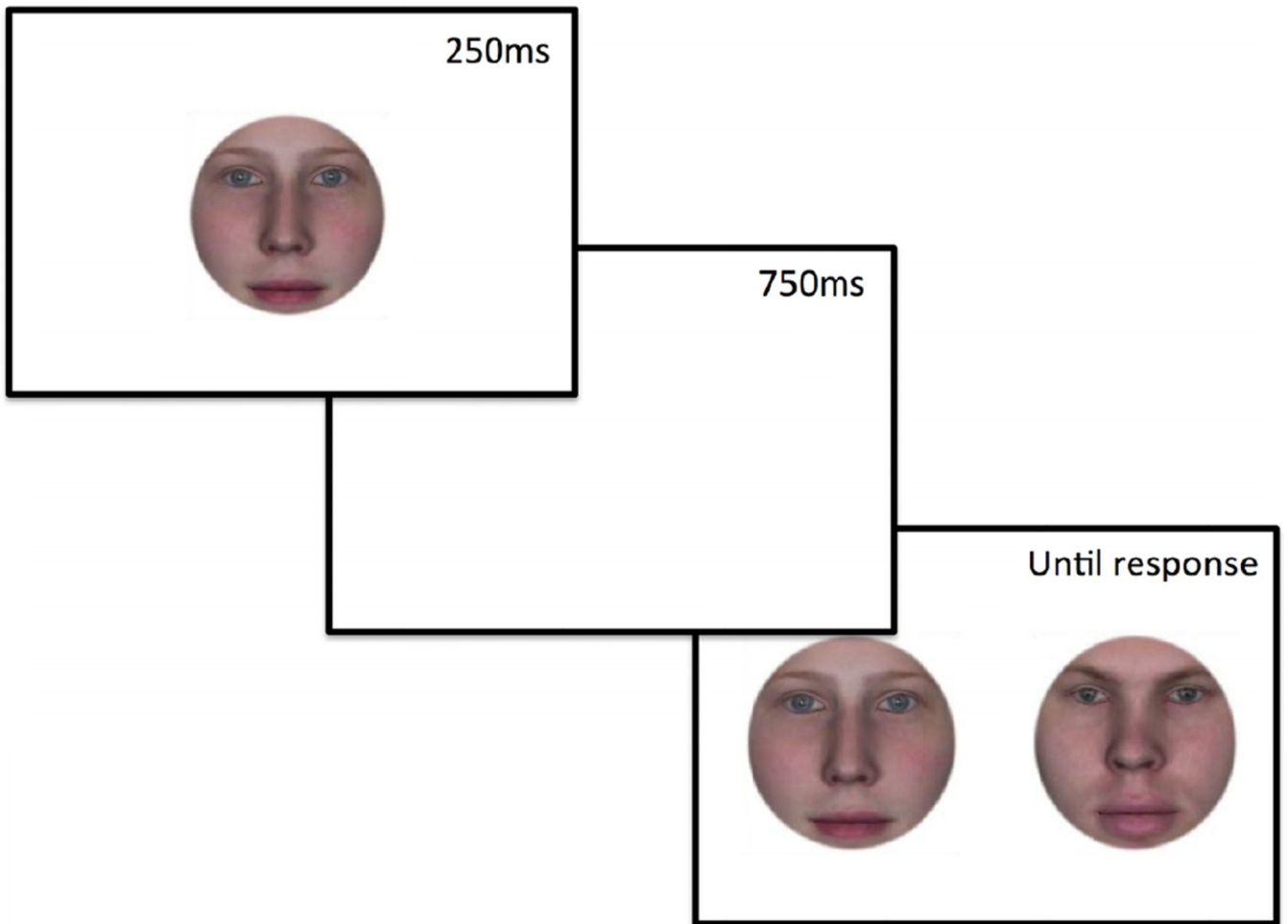


Figure 3.

A schematic depiction of the match-to-sample task used in Experiment 2. Participants were asked to indicate which of the two test faces matched a previously shown sample face.

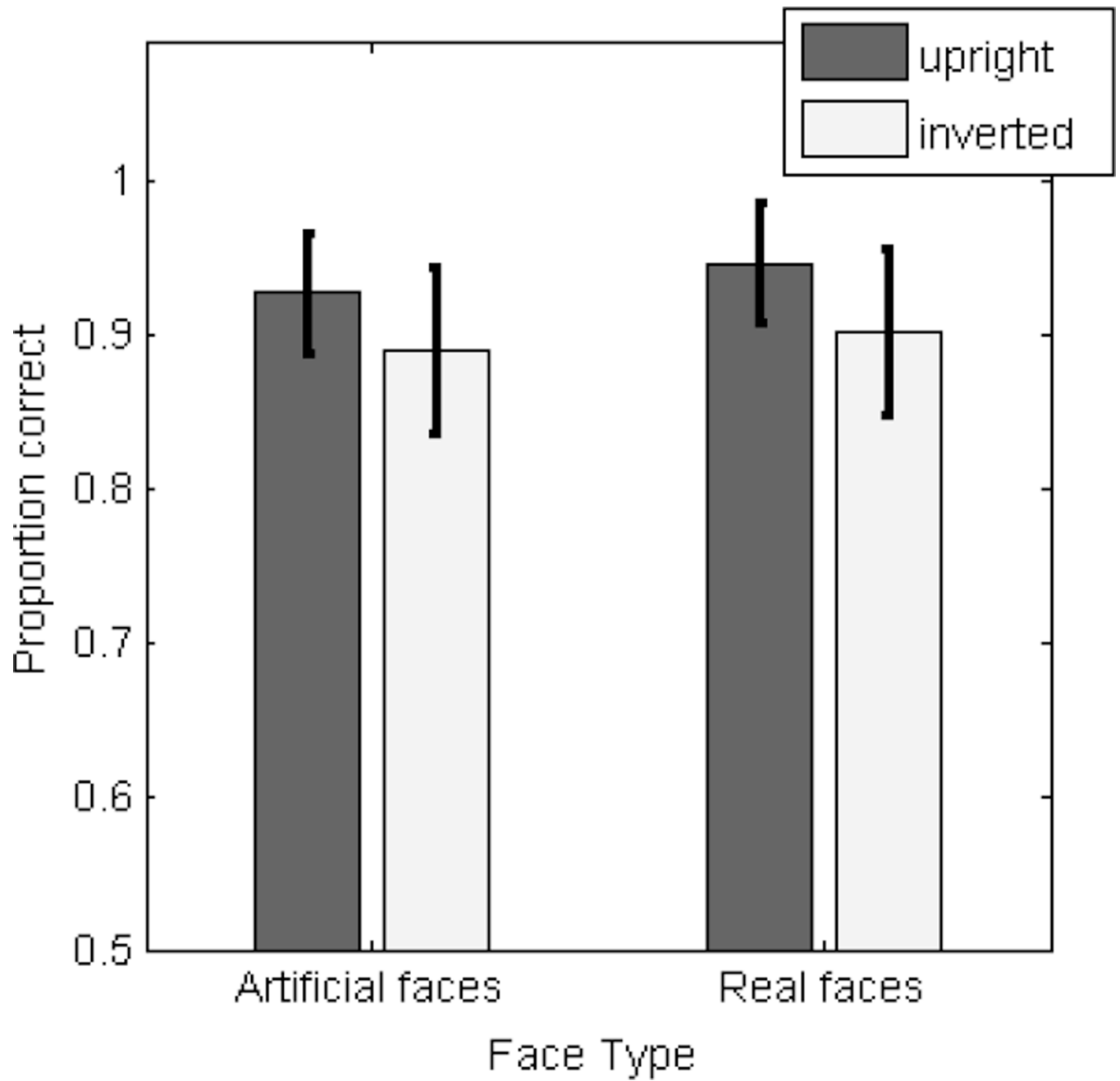


Figure 4. Average proportion correct values across participants as a function of face type and orientation. Error bars represent ± 1 std. dev.