Published in final edited form as:
Vis cogn. 2013 ; 21(6): . doi:10.1080/13506285.2013.823138.

Biological Sex Determines Whether Faces Look Real

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

Abstract

Judging whether a face is real or artificial can be done relatively rapidly and accurately, even when visual information is substantially impoverished. The perception of animacy in the face also has several interesting properties that may reflect both the underlying "tuning" of face space to preferentially represent real face appearance and the diagnosticity of individual features for categorizing faces as animate or inanimate. In the current study, we examined how sex categories interact with animacy perception by separately characterizing animacy judgments as a function of stimulus sex. We find that stimulus sex affects subjective ratings of animacy and sex categorization of real and artificial faces. Specifically, female faces look more artificial and artificial faces look more female. We discuss our results in terms of the ecology of real and artificial faces and the possible role of visual experience with artificial female faces, and the objectification of female faces.

Keywords

Face perception; Sex effects; Gender Classification; Animacy Perception

Introduction

Artificial faces of many kinds serve an important function in behavioral and neural studies of face perception. Photo-realistic (Oosterhof & Todorov, 2008) and schematic (Wilson, Loffler, & Wilkinson, 2002) artificial faces are an important way to control specific aspects of facial appearance, such as 3D shape and 2D surface pigmentation (Blanz & Vetter, 1999; Balas & Nelson, 2010), while preserving many aspects of typical appearance. In many respects, artificial faces are a useful way to balance the desire to study naturalistic stimuli that have clear ecological relevance with the competing goal of being able to manipulate stimulus properties to address specific hypotheses (Felson & Dan, 2005; Rust & Movshon, 2005).

The perception of artificial faces is also an important test domain for understanding the nature of face-specific processing and high-level visual learning in general. Despite the fact that artificial faces share the first-order arrangement of discrete features into a face pattern and are typically well-matched to real faces in terms of so-called second-order geometric relationships (Maurer, Le Grand & Mondloch, 2002), artificial faces appear to occupy interesting territory at the boundary of normal face perception. Advances in graphics and image editing software notwithstanding, artificial human faces are in general readily distinguishable from real ones (Farid & Bravo, 2011). The perception of animacy in human faces is also categorical (Looser & Wheatley, 2011), meaning that perceived animacy along a continuum spanning artificial and real faces changes in a step-like fashion rather than gradually. This non-linear relationship between differences in perceived animacy and physical stimulus differences is largely driven by the appearance of the eyes (Looser & Wheatley, 2011), though the appearance of the rest of the face also influences animacy judgments (Balas & Horski, 2012). Besides consistent psychophysical evidence for a categorical boundary between real and artificial faces, in some circumstances artificial faces appear grotesque or “uncanny” (Mori, 1970) further suggesting that normal face perception is tuned to the appearance of real faces.

In the current study, our goal was to examine how the perception of real faces and artificial faces interacts with other categories, specifically face sex. The attribution of intention and agency to a face has consequences for social interactions in real and virtual environments, and we wished to determine how categories observers have extensive experience with (male and female faces) influence the categorization process by which faces are perceived as real or artificial. To the extent that there is a relationship between perceived animacy and face sex, we also wished to determine if it is symmetrical – should face sex affect the perception of a face as animate or inanimate, does the animacy of a face also influence perceived sex?

We predicted that perceived animacy would depend upon the sex of the faces being judged. Our hypothesis is motivated by related results suggesting that the perception of race is also “gendered” (Johnson, Freeman, & Pauker, 2012), demonstrating that the assignment of social and perceptual categories may be substantially influenced by real covariation between categories and cognitive biases. Darker skin tone, for example, is both diagnostic of sex category (Male faces are darker than female faces (Jablonski & Chapin, 2000)) and diagnostic of some race categories, resulting in measurable biases in categorizing faces of vary lightness according to race and sex. In the case of animacy, or perceiving “life” in a face, the properties of many types of artificial face may similarly covary with sexually dimorphic properties of real faces. There are multiple sexually dimorphic features of the human face, including the thickness and height of eyebrows (Burton, Bruce & Dench, 1993), the extent to which the eyes are recessed (Bruce et al., 1993), and the overall difference in skin albedo mentioned above. Sufficient covariation between the appearance of real and artificial faces and sexually dimorphic features like these could lead to “gendered” animacy perception. In the case of race categories there are relatively straightforward features differences between categories to consider – to what extent can we characterize the difference between real and artificial faces in a manner that is representative of the very broad class of artificial faces? Presently, there are two important relationships between artificial faces and female faces that we will highlight to motivate our hypothesis that animacy perception may be affected by the sex of the face. In both cases, the effects of cosmetics usage on appearance play a critical role, since cosmetics use as a technology for manipulating appearance (Russell, 2010) is heavily biased towards female users. Indeed, cosmetics usage tends to make faces look more feminine (Cox & Glick, 1986), which suggests that the effects of cosmetics on appearance represents another important sexual dimorphism in facial appearance that both accentuates natural dimorphisms (Russell, 2010), but may also induce other dimorphic features that could contribute to gendered animacy perception. First, we note that cosmetics use tends to homogenize face pigmentation (Mulhern et al., 2003), making faces more attractive (Fink, Grammer, & Thornhill, 2001) by covering idiosyncratic blemishes. Homogenous skin tone and pigmentation may thus be perceived as a marker of femininity that covaries with the appearance of artificial faces of many kinds (e.g. dolls, cartoons, avatars, etc. (Giard & Guitton, 2010) Second, cosmetics usage often enhances the natural dimorphism in face contrast (Russell, 2003) such that female faces have enhanced local contrast via eye-liner, mascara, or lipstick. Again, this induced (or more accurately, *enhanced*) dimorphic features is both diagnostic of femininity and potentially representative of a number of artificial faces. In both cases, we cannot (and do not) claim that all artificial faces have homogeneous skin and high local contrast – to our knowledge there is not available data on the typical appearance of a wide range of artificial faces. Nonetheless, we suggest that even weak covariance between these properties in artificial faces and female faces could be sufficient to lead to gendered perception of animacy.

We chose to test our hypothesis in two ways: (1) By comparing the function relating perceived animacy ratings to physical appearance along a real/artificial morph continuum for male and female faces, (2) measuring sex categorization for real and artificial faces along a morph continuum. In both tasks, we found that female faces look more artificial than male faces, and artificial faces look more female than real faces do; perceived face animacy is gendered. We support these conclusions by an additional task in which we measured the low-level discriminability of the faces we used in the first two experiments and conclude that low-level artifacts of the morphing procedure and our stimulus set do not contribute to the effects we observe. Finally, we attempted to replicate the results of our first experiment with a new set of artificial faces to test the generality of our effects. We discuss our results in relation to prior research regarding known sex effects in face perception, with an emphasis on the relationship between diagnostic features for judging animacy and sexually dimorphic facial features (Russell, 2009).

Experiment 1

The goal of our first experiment was to determine whether or not subjective animacy ratings of human faces interact with stimulus sex. Specifically, does the point of subjective equality (PSE) for animate/inanimate categorization vary as a function of the sex of the face?

Methods

Subjects—We recruited 19 naïve observers (10 female, 9 male) between the ages of 19 and 25 to take part in this task. All participants reported normal or corrected-to-normal vision.

Stimuli—We created 6 sequences of images that spanned a continuum between real and artificial faces for use in this task. We created each continuum by choosing a single real face (3 male, 3 female) and pairing it with a visually similar doll face. Both the human face and the doll face were cropped to remove the external contour of the face. These two faces were then blended together in 10% increments to create a collection of 11 morphed images (66 total images). The resulting full-color images were 280×400 pixels in size.

Procedure—All images were presented to our participants on a 1024×768 LED display. Participants viewed the stimuli from a distance of approximately 60cm and the images subtended a visual angle of ~3.5 degrees of visual angle in height and 2.5 degrees of visual angle in width. Each image appeared 8 times in a fully randomized order for a grand total of 528 trials. Images were presented on-screen for 500ms against a medium-gray background, and participants were asked to use a 1-7 Likert scale to judge the animacy of each face (1=definitely artificial, 7=definitely real). Participants had unlimited time to make each judgment and were encouraged to use the full range of numbers.

Results

We analyzed the data from each participant by fitting a logistic curve to the average value assigned to faces at each morph level. First, the ratings assigned to each level (e.g. all 10% human faces) were averaged together. Second, participant ratings were re-centered and re-scaled so that the 1-7 range was transformed into a 0-1 range. This transformation did not guarantee that participants' ratings spanned the full range of possible values – if the maximum value assigned to any face was 6, for example, then the maximum re-scaled value would be 0.83. Finally, the average ratings at each morph level were used to fit the parameters of a logistic function so that we could estimate the point of subjective equality (PSE) for each subject in each stimulus category. In the context of our task, the PSE corresponds to the morph level at which faces are rated as being maximally ambiguous, corresponding to the midpoint of the scale. We used the Palamedes Toolbox for Matlab (Prins & Kingdom, 2009) to fit these psychometric curves with alpha, beta, gamma, and lambda as free parameters. We obtained robust fits for all participants.

PSE analysis – All stimuli combined—First, we combined the ratings assigned to male and female faces to determine if the overall PSE was shifted towards the human end of the morph continuum, as observed in previous reports (Looser & Wheatley, 2011, Balas & Horski, 2012). This shift towards real faces may reflect the sensitivity of human face processing to distortions of typical appearance. We found that the average PSE was significantly larger than the physical midpoint of the morph sequence (M=0.56, sd=0.09, t(18)=2.60, p=0.018, two-tailed, one-sample t-test). This is in concordance with previous results, demonstrating that observers in our task were also conservative about assigning animacy to morphed faces in general.

PSE analysis -Effects of observer and face sex—We continued by comparing the PSE to male and female faces. We calculated the average ratings for each morph level for male and female faces separately (Figure 2), and then fit separate psychometric curves to each category for each participant. The result was a separate estimate of the PSE for male and female morph continua for each observer. We analyzed these values with a 2×2 mixed-design ANOVA with participant sex as a between-subjects factor and stimulus sex as a within-subjects factor. This analysis revealed only a main effect of stimulus sex ($F(1,17)=17.8$, $p=0.001$, $\eta^2=0.52$) – neither the main effect of participant sex ($F<1$, $p=0.53$) nor the interaction between the two factors ($F<1$, $p=0.47$) were significant. The significant effect of face sex resulted from a PSE for female faces (M=0.64, 95% confidence interval = [0.58 0.69] that was significantly closer to the “real” endpoint than the PSE for male faces (M=0.46, 95% confidence interval = [0.39 0.55]). Participants were therefore especially conservative about their assessments of animacy in female faces, but comparatively more lax when rating the animacy of male faces. However, by examining Figure 2, it is apparent that considering only the PSE obscures the real basis of this effect. While it is true that the PSE for female faces is shifted significantly to the right relative to male faces, this is the result of an overall baseline shift in the ratings assigned to female faces relative to male faces. Female faces are overall classified as less real than male faces across the entire real/artificial continuum.

Discussion

In Experiment 1, we replicated previous results regarding the non-linear relationship between the physical and perceived animacy of faces along a continuum between real and artificial faces. We observed a sigmoid-like function for subjective ratings of animacy along this continuum, and further found that the overall PSE for male and female faces combined was shifted towards real faces, in good agreement with previous studies. These results serve as an important validation of our stimulus set, insofar as they demonstrate that we obtain the same results as previous studies when we consider our aggregate data. Additionally, we found that animacy ratings for female faces were overall lower than male faces at all morph levels, suggesting that perceived animacy is a function of face sex. In Experiment 2, we continued by asking a complementary question: Is perceived sex also a function of animacy?

Experiment 2

The data from Experiment 1 indicated that stimulus sex modulates perceived animacy across the entire continuum between real and artificial faces. Specifically, female faces appear to be perceived as more artificial, resulting in lower animacy ratings across all morph levels. In our second experiment, we directly examined the relationship between stimulus sex and animacy by asking participants to perform a speeded sex categorization judgment with images across the real/artificial continuum we constructed.

Sex categorization is known to be affected by expertise and category membership. Participants perform more poorly at categorizing other-race faces according to sex (O'Toole, Peterson, & Deffenbacher, 1996), and both perceived and veridical confounds between race categories and sex lead to measurable impairments in sex categorization (Goff, Thomson, & Jackson, 2008) – for example, Black faces in general are perceived as more masculine, making Black women's faces harder to correctly categorize. Presently, we asked whether animacy is "gendered" (Johnson, Freeman, & Pauker, 2012) in a similar manner, which may be a key determinant of the results obtained in Experiments 1.

Methods

Subjects—We recruited 10 additional participants (6 female, 4 male) who were naïve to the purpose of this experiment and who had not participated in Experiment 1. All of these participants reported normal or corrected-to-normal vision.

Stimuli—The same stimulus set created for Experiment 1 was used here.

Procedure—Participants in this task were asked to categorize each image as male or female as quickly as possible. Stimuli were presented with the same display parameters as in Experiment 1, but were visible until participants responded. Participants were encouraged to respond as quickly as possible, but were also told not to sacrifice accuracy for speed. Each image was presented 12 times in a fully randomized order, for a grand total of 792 trials.

Results

For each participant, we calculated the average accuracy and response time (RT) to correctly labeled faces as a function of morph level and face sex. In both cases, the data was analyzed using an 11×2 repeated-measures ANOVA, with morph level and sex as within-subjects factors.

Accuracy—Our ANOVA revealed a η significant main effect of level ($F(10,90)=4.87$, $p<0.001$, $\eta^2=0.35$) and a significant interaction between sex and level ($F(10,90)=15.7$, $p<0.001$, $\eta^2=0.64$). In Figure 3, we display mean accuracy as a function of morph level and face sex, which clearly illustrates the cross-over interaction we observe in our data: Male faces are hardest to label correctly when they are artificial and female faces are hardest to label correctly when they are real. This suggests that in general, artificial faces are feminized relative to real faces.

Response Times—Our ANOVA of average response latencies to correct categorizations only revealed a marginal main effect of level ($F(10,90)=1.92$, $p=0.052$, $\eta^2=0.18$) and also a significant interaction between face sex and morph level ($F(10,90)=7.86$, $p<0.001$, $\eta^2=0.47$). A simple effects analysis of the effect of level within each sex category revealed a highly significant effect of level for female faces ($F(1,10)=4.94$, $p<0.001$) but only a marginal effect for male faces ($F(1,10)=2.10$, $p=0.033$) after correction for multiple comparisons. In Figure 4, we display the average RT as a function of morph level and sex.

Discussion

Complementing our results from Experiment 1, the data from this task reveal that real faces look less female than artificial faces. Together, both tasks demonstrate that the perception of faces as real or artificial covaries with the perception of those same faces as male or female. Sex categories and animacy each impact the perception of the other. In Experiment 2, we see that this effect appears to be somewhat stronger for female faces than for men, but we do obtain significant results for each category.

Both tasks, however, are potentially limited by the stimulus set we created. Low-level properties of the morph continua we generated could potentially be the basis of the effects we have observed thus far. For example, if the real female faces we chose are physically more similar to their corresponding doll faces than the real male faces are to their counter-parts, our data would be a simple by-product of image discriminability. We continue by implementing an image-level discrimination task in Experiment 3 to determine the perceptual similarity of stimuli across our morph sequences.

Experiment 3

In this experiment, we used an image-level discrimination task to further examine the nature of the boundary between real and artificial faces and to investigate the basis of the sex effects we have observed in Experiments 1 and 2. This task is critically important for interpreting the nature of these results.

Creating a uniform continuum between distinct stimuli is straightforward when there is an established physical variable that has known relevance to perception. For example, the categorical perception of phonemes has been demonstrated by smoothly varying voice-onset time (Liberman et al., 1957), which is known to be a meaningful physical variable that is measured by our auditory system. Similarly, wavelength can be varied smoothly to study the categorical perception of color (Bornstein & Korda, 1984). In complex stimuli like faces, however, it is less clear how to ensure that a morph sequence spanning two stimuli is comprised of uniform increments. Morphing software, such as the program we used to construct our stimuli, make it easy to generate stimuli at any point along an established continuum, but there is no guarantee that increments of the same nominal size across the continuum are actually of equal size in terms of some real physical variable, or that increments of the same size in different continua are actually perceptually matched. This latter point is particularly relevant to our results from Experiments 1 and 2. We selected the faces that comprise our stimulus set by visually matching dolls with real faces, leaving open the possibility that we chose female faces that have smaller perceptual increments along the continuum from real to artificial. Were this the case, much of our data from Experiments 1 and 2 could be predicted by simple pattern discrimination – female faces may have looked less “real” because they actually resembled artificial faces more. Alternatively, if there are not extant differences in the perceived difference between patterns along morph continua as a function of stimulus sex, we can be confident that our effects are not artifacts of the stimulus selection process.

Method

Subjects—We recruited an additional 17 observers (9 female, 8 male) to participate in Experiment 3. None of these participants had been recruited for Experiments 1 or 2, and all participants were naïve to the purpose of the experiment. All participants also reported normal or corrected-to-normal vision.

Stimuli—The same morph continua created for Experiment 1 were used in this task.

Procedure—Participants in this task were asked to perform a same/different image-level discrimination task using images at different points along the human-doll continua we created for male and female faces. On each trial, participants were shown two images sequentially and asked to report if the images were identical or if there were any differences at all between the two images. The first image was presented for 250ms, followed by a 500ms ISI and a 250ms presentation of the second image. The second image was also jittered by a random amount between 20-60 pixels both horizontally and vertically. Participants responded to each trial using a small button box (Retrolink USB).

We designed our task to measure image discriminability across the range of PSE values observed for male and female faces in Experiment 1. We chose to measure discriminability at morph levels of 30-70% by implementing "different" trials in our design with a 20% increment between the two images presented to the observer. For example, a "different" trial at a morph level of 30% would be comprised of a 30%-Human image and a 50%-Human image, with the order of sequential presentation randomized. "Same" trials at each morph level were implemented using repeated presentation of the same stimulus. The morph level assigned to "different" trials was assigned based on the first image presented to the observer in the trial sequence and corresponding false alarm rates were assigned based on the morph level of both stimuli. The rationale behind this assignment is that the initial stimulus presented to the observer in corresponding "same" and "different" trials is the same and performance in these two trials is thus more directly comparable. Also, individual participants were not shown "different" trials comprised of the same images in different sequence order: If observers were shown a trial sequence with a 30% morph followed by a 50% morph, that observer would not be shown a trial with those images appearing in the opposite order. This allowed us to measure the false alarm rate independently at each morph level so that d' values at each morph level were calculated with a unique hit rate and false alarm rate. We presented participants with 15 "different" trials and 15 "same" trials at each morph level (30-70%) and each sex (male and female for a grand total of 300 trials.

Results

We calculated the hit rate (correct detections of different images) and the false alarm rate (incorrect "different" responses to identical images) for each combination of morph level and face sex. We used these values to compute d' and the response criterion 'c' for each participant as a function of morph level for both male and female faces. If low-level discriminability is driving the effects observed in Experiments 1 and 2 (e.g. Female faces actually do look more like their artificial counterparts than male faces), we would expect to see sex effects on these measures.

Discriminability (d')—We analyzed the d' data using a 5×2 repeated-measures ANOVA with morph level and face sex as within-subject factors. This analysis revealed no main effects of sex ($F<1$) or level ($F<1$) and the interaction between these factors was also not significant ($F<1$). We display the average d' values for each morph level and sex category in Figure 6.

Response Criterion—We analyzed the response criterion, c, using an ANOVA with the same factors described above. This analysis revealed only a main effect of morph level ($F(4,16)=6.28$, $p<0.001$, $\eta^2=.28$), such that significantly more "different" responses were made to faces that were closer to the "real" end of the morph continuum.

Discussion

Our data suggest that the effects we have observed in Experiments 1 and 2 are unlikely to be primarily driven by image-level differences. Specifically, we do not find any evidence that the real and artificial female faces we selected are more similar to one another than the male faces we chose. We also do not see evidence that image-level discriminability increases across the PSEs measured in Experiment 1, suggesting that in terms of pattern discrimination, the perceptual distance across the morph continua is relatively constant. Finally, the criterion shift for artificial faces regardless of sex is in keeping with data from other-race studies demonstrating that there is a “same” bias for faces of other races (Quattrone & Jones, 2008) – members of categories we have low perceptual expertise with tend to look more alike, leading to a lower number of “different” responses. The observed bias for “same” responses for faces closer to the artificial end of the spectrum is consistent with the idea that artificial faces are not as well-represented in the face space of individuals with typical face experience.

Experiment 4

In our final experiment, we chose to determine the extent to which the sex-dependence of perceived animacy was driven by our particular choice of stimuli, particularly the artificial faces we selected as approximate visual matches for our photographs of real faces. In this final task, we used the same experimental design and procedure as Experiment 1, but created artificial faces matched to photographs of real individuals using FaceGen, a proprietary software package that implements a 3D morphable model of facial appearance. Faces created with FaceGen (and other similar applications) are more representative of the class of artificial faces typically used in current face recognition research (Oosterhof & Todorov, 2008) and the ability to estimate the appearance of an artificial face from a specific 2D image offers a principled way to match real and artificial faces along morph continua. Our goal in this task was to determine whether or not the sex effect we observed in Experiment 1 extends to another class of real and artificial faces, or if our previous results are stimulus-dependent.

Methods

Subjects—We recruited 18 participants for this study (11 female/ 7 male), none of which had participated in any of the previous tasks reported above. All participants reported normal or corrected-to-normal vision and were naïve to the purpose of the experiments.

Stimuli—We used 20 full-color photographs of men and women (10 of each sex) to create a set of synthetic 3D models of facial appearance using the “PhotoFit” capabilities of FaceGen. Briefly, the PhotoFit option allows the user to identify a set of fiducial points on a photograph of a face, which FaceGen uses to create a least-squares estimate of the 3D shape and 2D surface properties of the individual depicted in the photographs. We rendered the resulting 3D models from a frontal view and scaled the artificial and real images to the same size (300 × 300 pixels). We cropped the hairline of both kinds of faces as well, since the FaceGen application does not render hair very accurately. We then used the resulting paired images to create morph sequences in the same manner as described in Experiment 1. All original photographs depicted individuals who were free of facial hair, cosmetics, or facial piercings.

Procedure—The testing procedure for this task was identical to that described in Experiment 1, save for the larger number of stimuli in this task.

Results—We analyzed the data from this experiment using a 2×11 repeated-measures ANOVA with the sex of the face and the morph level of the stimulus as within-subject factors. This analysis revealed a significant effect of morph level ($F(10,8)=9.94$,, $p=0.002$) and a significant interaction between morph level and face sex ($F(10,8)=24.1$, $p<0.001$).

There was no significant main effect of stimulus sex ($F(1,17)=2.44$, $p=0.137$). Post-hoc tests revealed that the interaction between sex and morph level was driven by a significant effect favoring higher ratings for male faces than female faces (consistent with Experiment 1) at the “real” end of the continuum established by our morph sequences, but no difference at the “artificial” end of the continuum. (Figure 7). Specifically, we observed significantly higher ratings for male faces at morph levels of 100% ($p=0.003$), 90% ($p<0.001$) and 80% ($p=0.041$) – in each case, as assessed by a two-tailed, paired-samples t-test. We did not observe significantly higher ratings for male faces at any other points along the morph continuum. Thus, while we do observe higher animacy ratings for male faces with this stimulus set, this effect is limited in scope to morphs that are largely real-looking. This sensitivity to stimulus level was not evident in Experiment 1. Our results thus do not fully replicate the effects found in Experiment 1, in which we observed significant male/female differences in perceived animacy across the entire continuum from real to artificial faces. We also note that the variation across participants appears to be larger in this task than in Experiment 1, which may also reflect important stimulus differences between the two tasks.

Discussion—The results of our final task suggest that the effect we observed in Experiment 1 does generalize to new stimuli, but with an important qualification. Those faces that more closely resembled the artificial faces created using FaceGen did not elicit differential animacy ratings, while the faces that were more photorealistic did. This suggests that at the artificial end of the spectrum in Experiment 1, the difference between male and female artificial faces may have been driven by item-specific factors, possibly including the cosmetic details applied to the doll faces we selected in that task. However, the fact that we observed a sex difference in perceived animacy at the real end of the spectrum in this task suggests that in terms of real face appearance, the sex effects we observed are more general. We take up the discrepancy between the results of this experiment and the results of Experiment 1 more fully in our General Discussion section.

General Discussion

The data from both Experiment 1 and Experiment 2 demonstrate that animacy is gendered – female faces are harder to correctly classify when they resemble real faces more, while the opposite is true for male faces. Further, response times to correctly categorize female faces by sex are slower as faces appear more real. These data, in concert with the results from Experiment 1, indicate that under some circumstances, perceived animacy depends upon stimulus sex: Female faces look more artificial, and male faces look more real. Observers thus appear to be sensitive to covariance (either real or perceived) between face sex and animacy, and our data from Experiment 3 suggest that the results of our first two experiments are not simply an artifact of physical stimulus differences across male and female morph continua. Were this the case (if, for example, real female faces were objectively more similar to their artificial counterparts than male faces) we would expect that image discrimination should be subject to the same sex effects. . Our first three experiments thus suggest an intriguing relationship between sex categories and observers’ perception of real vs. artificial facial appearance, a result that may have applications to the design of synthetic agents in multiple domains.

The results of Experiment 4 complicate the above discussion, however, but also raise important issues for further study. Specifically, Experiment 4 is intriguing in that we only partially replicated the results of our first task – male faces were rated as more “real” than female faces at the real end of the spectrum (from 80-100% morph levels), but not beyond that. This demonstrates that while the sex effect we observed using our original stimulus set generalizes to a new and larger set of *real not* human faces, the effect does generalize to a new set of artificial faces. The former result provides important support for the idea that the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

perceived animacy of real faces (excluding artificial faces) is indeed a function of sex category, which I suggest is an interesting contribution in itself. Regardless of the results obtained at the artificial end of the morph continuum, Experiment 4 supports the conclusion that the perception of animacy in real, photographic human faces is affected by stimulus sex. The latter result, however, is also interesting insofar as the failure to obtain a sex effect for a new class of artificial faces makes it possible to draw some tentative inferences about the critical visual features that drive the effect with real faces. The main difference between the artificial faces we used in Experiment 1 (doll faces) and Experiment 4 (FaceGen photo-fit faces) is that the faces in Experiment 4 were created by using a morphable model of facial appearance to estimate the best fit to image data obtained from real faces (Blanz & Vetter, 1999). Compared to the faces in Experiment 1, these artificial faces were closely matched to the shape of the real faces at the other end of the animacy continua we defined by morphing images together. The photo-fitting process is carried out by identifying a number of shape-based fiducial points on the original image, which are then fit as closely as possible with a synthetic face. The result is that the primary source of appearance differences between real and artificial images in Experiment 4 are errors in surface pigmentation rendering (e.g. the appearance of the skin). I suggest therefore, that Experiment 4 tells us that sexually-dimorphic shape is not sufficient to drive this effect – with shape cues largely fixed across the real/artificial continuum in Experiment 4, the effect of sex disappears as surface properties become more artificial. More importantly, I suggest that the lack of a main effect of sex in Experiment 4 demonstrates an additional important point regarding face perception and the perception of animacy in real and synthetic agents: There are many different kinds of artificial face corresponding to variations in graphics algorithms, materials, and artistic technique, but only one kind of real face (at least defined in terms of material properties, etc.). The results of Experiment 4 thus demonstrate that not all artificial faces are perceived the same way, which raises a range of interesting questions about the critical visual features that support face processing in different settings. Moreover, our results from this task do support the conclusion that perceived animacy is affected by sex in real face images. Taken together, we therefore suggest that while our data do not support conclusions that are generalizable to a broad class of artificial classes, we have observed a novel characteristic of the visual processing of real human faces.

Why should animacy and sex categorization interact this way in real face images? As discussed in the introduction, we suggest that known sexually dimorphic features of the human face may indeed correspond to diagnostic information that in some instances supports the discrimination of real faces from artificial faces. The lack of a main effect of sex in Experiment 4 suggests that not all artificial faces will reflect these covarying features in the same way, but we suggest that the visual system may nonetheless have implicitly learned that these dimensions of variation are not orthogonal. Sex classification requires the integration of multiple cues (Bruce et al., 1993; Bruce & Langton, 1994), including the 3D shape of the head (O'Toole et al., 1997; O'Toole et al., 1998) and 2D surface properties (Russell, 2003), both of which may covary across the male/female and real/artificial category boundaries. The critical question is whether observers' representation of the modal appearance of artificial faces is consistent with their representation of modal appearance of female appearance. In particular, the lack of an effect at the artificial end of the spectrum in Experiment 4 suggests that sexually-dimorphic shape and pigmentation may contribute to these modes to differing degrees. In the absence of an ecological survey of artificial face appearance, we can presently only suggest that some broad trends do indicate at least a crude correspondence between the appearance of doll faces and female appearance. In terms of surface properties, male faces are known to be both darker (Frost, 1988) and redder (Tarr et al., 2001) than female faces, (see Russell, 2010 for a comprehensive review), which is consistent with the classic pale skin of porcelain and plastic dolls. The size and shape of male faces also differs from female faces (Burton et al., 1993) favoring protruding noses and wider jaws in men, but protruding cheeks and chins in women. Again, we suggest that this is broadly consistent with doll appearance regardless of sex – the term “doll-like” features, for example, typically refers to smaller noses and mouths, but larger eyes. The former is consistent with typical female 3D shape, while the latter is consistent with the use of cosmetics to enlarge perceived eye size in women (Russell, 2009). At present we are speculating about the relationship between appearance modes for gender and appearance modes for synthetic faces; a rigorous survey of the appearance of artificial faces would be required to draw firm conclusions. The discrepancy between the results of Experiments 1 and 4 also suggest that a broad behavioral survey of the stimulus conditions necessary to observe a main effect of sex on perceived animacy would clarify the critical sources of covariation that may drive the results we observed in our tasks. For example, we motivated our study with the observation that cosmetics usage among women may also have induced particularly salient dimorphic features (homogeneity of skin pigmentation and enhanced local contrast) that tended to co-vary with artificial appearance – while we suggest that these properties are likely to both be evident in the images we used in Experiments 1-3, we also observe that the artificial faces created for Experiment 4 do not necessarily share the same enhanced local contrast as our pictures of actual dolls. The lack of a sex effect on animacy ratings at the artificial end of the morph spectrum in Experiment 4 may thus be a function of the lack of cosmetic detail in the artificial faces created by FaceGen, but the presence of the effect at the real end of the spectrum may be due to increased homogeneity of the skin. Manipulating specific qualities like 3D shape of the head, 2D configuration of the features, and the texture of the skin could potentially help isolate the critical features that drive our effect. Presently, however, our data suggest that determining how diagnostic cues for animacy relate to diagnostic cues for sex, race, or other categories would add to our understanding of how face representations are tuned to the statistics of experience.

Besides this perceptual account of how animacy may become gendered, an alternative interpretation of our data is that participants’ responses reflect the extent to which female faces and bodies are objectified. This term typically refers to non-visual judgments like the subjective attribution of agency (Cikara, Eberhardt & Fiske, 2010) or mind (Loughnan et al., 2010), both of which are decreased for women relative to men. Recent results, however, suggest that female bodies are also *perceptually* processed more like objects than male bodies – inversion of female bodies, for example, does not disrupt recognition as much as it disrupts male bodies (Bernard et al., 2012), suggesting that they are treated more like objects in terms of visual perception and recognition (But see Tarr (2013) for important criticism of these results). Here, rather than using the inversion effect as a proxy for objecthood, our data speak directly to the question of whether or not female faces look real to an observer via an explicit perceptual judgment. Our results suggest that across the entire continuum between real and artificial faces, female faces look less real, which may partly reflect the construal of these faces as being more object-like. Further, the data from Experiment 2 suggests that this bias may implicitly affect other visual recognition tasks. Assigning sex categories to faces is affected by animacy in a way that is consistent with increased objecthood for female faces. While our current results cannot distinguish between purely perceptual effects and higher-level cognitive biases, measuring and incorporating objectification into the perceptual study of animacy is an intriguing direction for future study. In either case, the results of Experiment 4 again suggest that we cannot consider all artificial faces as part of the same monolithic category. Whether we choose to discuss and examine the effects we report here in terms of perceptual or cognitive mechanisms, we must acknowledge that our data suggests that the extent to which sex and animacy interact depends on the characteristics of different kinds of artificial face.

Another important issue that deserves further study is the extent to which realistic artificial faces are subject to the same face-specific behavioral effects as real faces. For example, both the composite face effect (Young et al., 1987), which is known to be race-specific (Michel et al., 2006) or photo-negation (Galper, 1970), which appears to be species-specific (Parr & Taubert, 2010) are good candidates for further characterizing the similarities and differences between real faces and artificial faces. The latter is potentially of particular interest in the context of our above discussion of the surface properties of real and artificial faces, since photo-negation reverses the polarity of specularities, which are important cues for determining surface gloss. One important way in which convincing artificial faces differ from lower-quality synthetic images is the extent to which the proper scattering of light through the skin is modeled (Krishnaswamy & Baronoski, 2004). Skin that reflects too much light will tend to have strong specularities and ultimately look plastic or waxy (MacDorman et al., 2009), suggesting that the disruption of those specularities via contrast negation may have substantial consequences for perceived animacy.

Finally, we point out important limitations of the current study. Regarding our discussion of the role material properties may play in our tasks, artificial faces come in many different varieties, and even with the inclusion of FaceGen faces in Experiment 4 we have studied a fairly limited range of synthetic faces. Considering how material (plush, fabric, or plastic) influences these phenomena would be an important contribution towards determining how representations of facial appearance and representations of textural properties each contribute to the assignment of animacy to face images. Second, by employing cropped faces in our study, we limited our analysis of perceived animacy to internal face features. External face features are known to make a substantial contribution to face perception in general, (Sinha & Poggio, 1996) and may also play an important role in animacy perception. Finally, the present study cannot speak to the extent to which animacy is a domain-general function that is specific to faces, or if the distinction between real and artificial objects is a domain-general process. Examining how animacy is perceived in both human and non-human faces (as well as other object classes) would be a useful way to understand how domain-specific animacy perception is, and how it intersects with face and object processing more broadly.

Acknowledgments

The human faces used in this study are courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition, Carnegie Mellon University, http://www.tarrlab.org/. BB was supported by COBRE Grant GM103505 from the National institute for General Medical Studies (NIGMS) and NSF EPSCoR Grant #EPS-0814442. The author would additionally like to thank Alison Carlson for her assistance with data collection, Joseph Horski for his role in creating the morphed stimuli, Alyson Saville for her help with data management and recruiting, and Alice O'Toole for helpful comments on an earlier draft of this manuscript.

References

1. Balas B, Horski J. You can take the eyes out of the doll, but…. Perception. 2012; 41:361–364. [PubMed: 22808589]
2. 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]
3. Bernard P, Gervais SJ, Allen J, Campomizzi S, Klein O. Integrating sexual objectification with object versus person recognition: The sexualized-body-inversion hypothesis. Psychological Science. 2012; 23:469–471. [PubMed: 22477107]
4. Blanz, V.; Vetter, T. A Morphable Model for the Synthesis of 3D Faces; SIGGRAPH '99 Conference Proceedings; 1999; p. 187-194.
5. Bornstein MH, Korda NO. Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. Psychological Research. 1984; 46:207–222. [PubMed: 6494375]
6. Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]

7. Bruce V, Burton AM, Dench N, Hanna E, Healey P, Mason 0, Coombes A, Fright R, Linney A. Sex discrimination: How do we tell the difference between male and female faces? Perception. 1993; 22:131–152. [PubMed: 8474840]
8. Bruce V, Langton S. The use of pigmentation and shading information in recognising the sex and identities of faces. Perception. 1994; 23:803–822. [PubMed: 7845771]
9. Burton AM, Bruce V, Dench N. What's the difference between men and women? Evidence from facial measurement. Perception. 1993; 22:153–176. [PubMed: 8474841]
10. Cikara M, Eberhardt JL, Fiske ST. From agents to objects: Sexist attitudes and neural responses to sexualized targets. Journal of Cognitive Neuroscience. 2010; 23:540–551. [PubMed: 20350187]
11. Cox CL, Glick WH. Resume evaluations and cosmetics use: When more is not better. Sex Roles. 1986; 14:51–58.
12. Farid H, Bravo MJ. Perceptual discrimination of computer generated and photographic faces. Digital Investigation. 2011; 8:226–235.
13. Felson G, Dan Y. A natural approach to studying vision. Nature Neuroscience. 2005; 8:1643–1646.
14. Fink B, Grammer K, Thornhill R. Human (homo sapiens) facial attractiveness in relation to skin texture and color. Journal of Comparative Psychology. 2001; 115:92–99. [PubMed: 11334223]
15. Frost P. Human skin color: A possible relationship between its sexual dimorphism and its social perception. Perspectives in Biology and Medicine. 1988; 32:38–58. [PubMed: 3059317]
16. Galper RE. Recognition of faces in photographic negative. Psychonomic Science. 1970; 19:207–208.
17. 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.
18. Goff PA, Thomas MA, Jackson MC. "Ain't I a woman?" Towards an intersectional approach to person perception and group based harms. Sex Roles. 2008; 59:392–403.
19. Husk JS, Bennett PJ, Sekuler AB. Houses and textures: Investigating the characteristics of inversion effects. Vision Research. 2007; 47:3350–3359. [PubMed: 17988706]
20. Jablonksi NG, Chapin G. The evolution of human skin coloration. Journal of Human Evolution. 2000; 39:57–106. [PubMed: 10896812]
21. 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]
22. Krishnaswamy A, Baronoski GVG. A Biophysically-based Spectral Model of Light Interaction with Human Skin. Computer Graphics Forum. 2004; 23:331.
23. Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology. 1957; 53:368–385.
24. Looser CE, Wheatley T. The Tipping Point of Animacy: How, When, and Where We Perceive Life in a Face. Psychological Science. 2010; 21:1854–1862. [PubMed: 21097720]
25. Louhgnan S, Halsam N, Turnae T, Vaes J, Reynolds C, Suitner C. Objectification leads to depersonalization: The denial of mind and moral concern to objectified others. European Journal of Social Psychology. 2010; 40:709–717.
26. 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.
27. Maurer D, Le Grand R, Mondloch CJ. The many faces of configural processing. Trends in Cognitive Sciences. 2002; 6:255–260. [PubMed: 12039607]
28. 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]
29. Mori M. The uncanny valley. Energy. 1970; 7:33–5.
30. Mulhern R, Fieldman G, Hussey T, Leveque J-J, Pineau P. Do cosmetics enhance female Caucasian facial attractiveness. International Journal of Cosmetic Science. 2003; 25:199–205. [PubMed: 18494902]
31. O'Toole AJ, Peterson J, Deffenbacher KA. An 'other-race effect' for categorizing faces by sex. Perception. 1996; 25:669–676. [PubMed: 8888300]

32. O'Toole AJ, Vetter T, Troje NF, Bulthoff HH. Sex classification is better with three-dimensional head structure than with image intensity information. Perception. 1997; 26:75–84. [PubMed: 9196691]
33. O'Toole AJ, Deffenbacher KA, Valentin D, McKee K, Huff D, Abdi H. The perception of face gender: The role of stimulus structure in recognition and classification. Memory & Cognition. 1998; 26:146–160. [PubMed: 9519705]
34. Oosterhof NN, Todorov A. The functional basis of face evaluation. Proceedings of the National Academy of Sciences of the USA. 2008; 105:11087–11092. [PubMed: 18685089]
35. Parr LA, Taubert J. The importance of surface-based cues for face discrimination in non-human primates. Proceedings of the Royal Society, Series B. 2010
36. Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision. 1997; 10:437–442. [PubMed: 9176953]
37. Prins, N.; Kingdom, FAA. Palamedes: Matlab routines for analyzing psychophysical data. 2009. http://www.palamedestoolbox.org
38. Russell R. Sex, beauty, and the relative luminance of facial features. Perception. 2003; 32:1093–1107. [PubMed: 14651322]
39. Russell R. A sex difference in facial pigmentation and its exaggeration by cosmetics. Perception. 2009; 38:1211–1219. [PubMed: 19817153]
40. Russell, R. Why cosmetics work. In: Adams, R.; Ambady, N.; Nakayama, K.; Shimojo, S., editors. The Science of Social Vision. Oxford University Press; New York: 2010. p. 186-203.
41. Rust NC, Movshon JA. In praise of artifice. Nature Neuroscience. 2005; 8:1647–1650.
42. Sinha P, Poggio T. I think I know that face…. Nature. 1996; 384:404. [PubMed: 8945457]
43. Tarr MJ, Kersten D, Cheng Y, Rossion B. It's Pat! Sexing Faces Using Only Red and Green. Journal of Vision. 2001; 1(3):337.
44. Tarr MJ. Perception isn't so simple: Commentary on Bernard, Gervais, Allen, Campomizzi & Klein, 2012. Psychological Science. 2013; 24:1069–1070. [PubMed: 23599305]
45. Taubert J. Chimpanzee faces are 'special' to humans. Perception. 2009; 38:343–356. [PubMed: 19485131]
46. Wilson HR, Loffler G, Wilkinson F. Synthetic faces, face cubes, and the geometry of face space. Vision Research. 2002; 42:2909–2923. [PubMed: 12450502]
47. Young AW, et al. Configurational information in face perception. Perception. 1987; 16:747–759. [PubMed: 3454432]

Figure 1.

An example morph sequence for a male face and a female face – the real face is depicted at far right and the artificial face is at far left. In between are examples of blended stimuli with varying degrees of animacy. Full sequences were generated with increments of 10%.

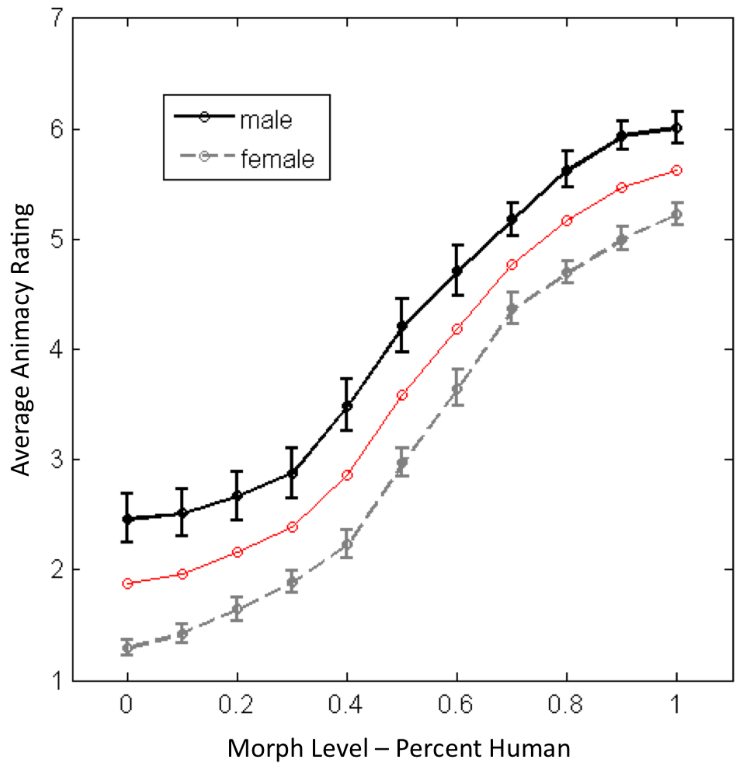

Figure 2.
Average animacy rating as a function of morph level and face sex, collapsing across participant sex. The red line illustrates the aggregate data pooled across male and female faces, while the black and gray lines illustrate the ratings for male and female face respectively. Error bars represent +/− 1 s.e.m. Female faces are reliably rated as less real than male faces.

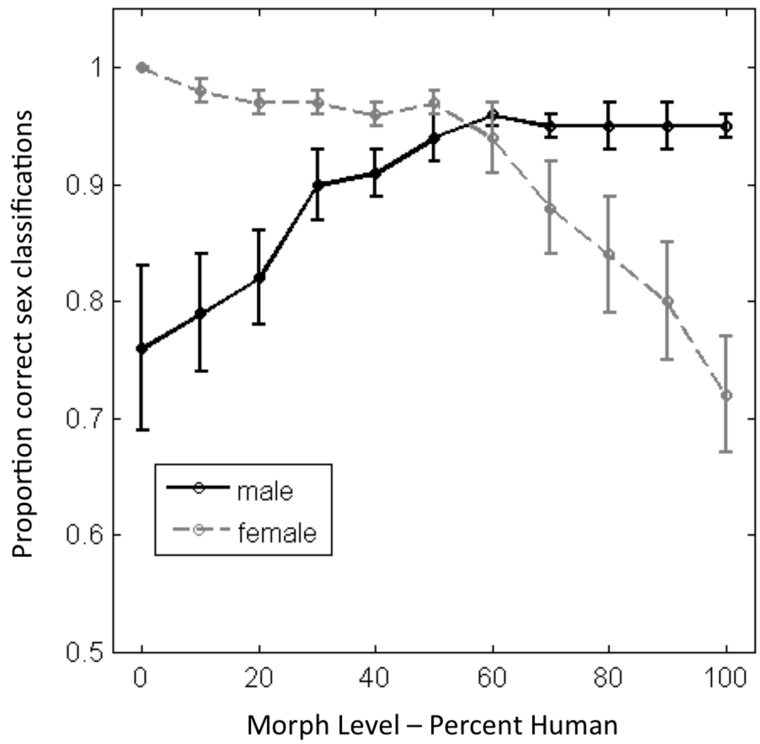

Figure 3.
Average proportion of correct sex classifications as a function of morph level and face sex. Error bars represent +/− 1 s.e.m. Male faces are easiest to classify when they are real, female faces are the easiest to classify when they are artificial. This suggests that in general, real faces are more masculine looking than artificial faces.

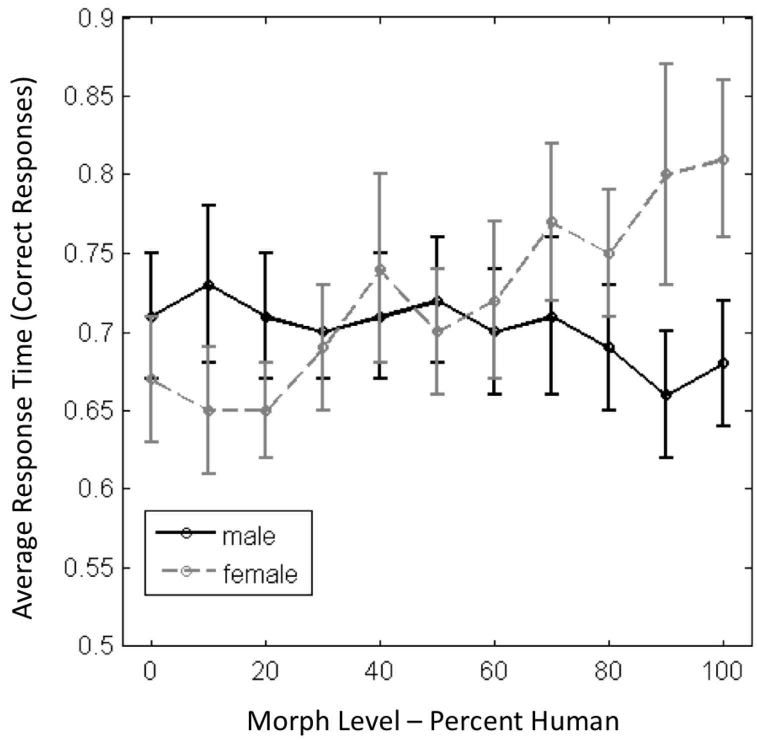

Figure 4.

Average response latency for correct classifications as a function of morph level and face sex. Error bars represent +/− 1 s.e.m. We observe a crossover interaction such that female faces are hardest to classify by sex when they are real, but the opposite is true for male faces.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

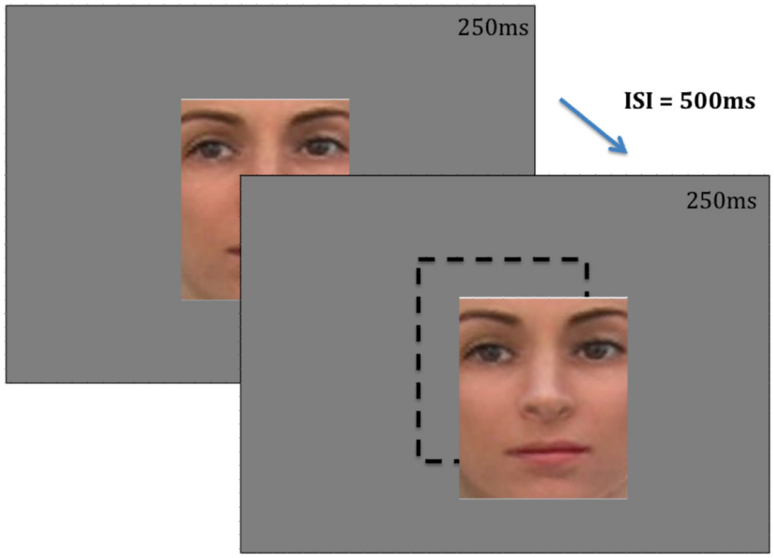

Figure 5.
A schematic view of the same/different paradigm used in Experiment 3. On each trial, two images were presented sequentially for 250ms each, separated by a 500ms ISI. The dashed lines on the second image illustrate that the second face was randomly jittered horizontally and vertically by a small amount relative to the first face.

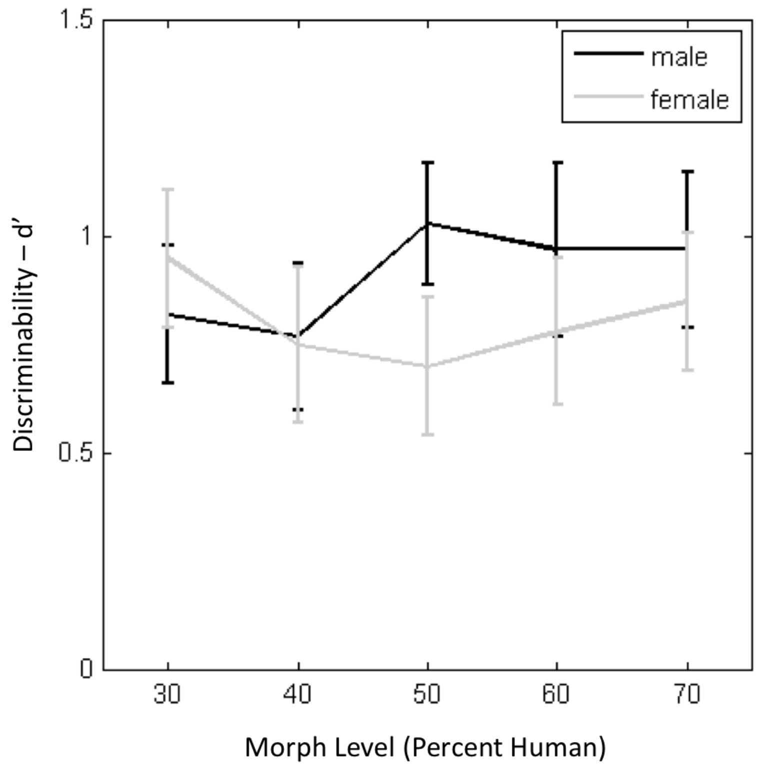

Figure 6.
Average image discriminability across the morph continuum as a function of stimulus sex. Error bars represent +/− 1 s.e.m. We found no effect morph level or sex, suggesting that low-level differences are relatively constant in all conditions.

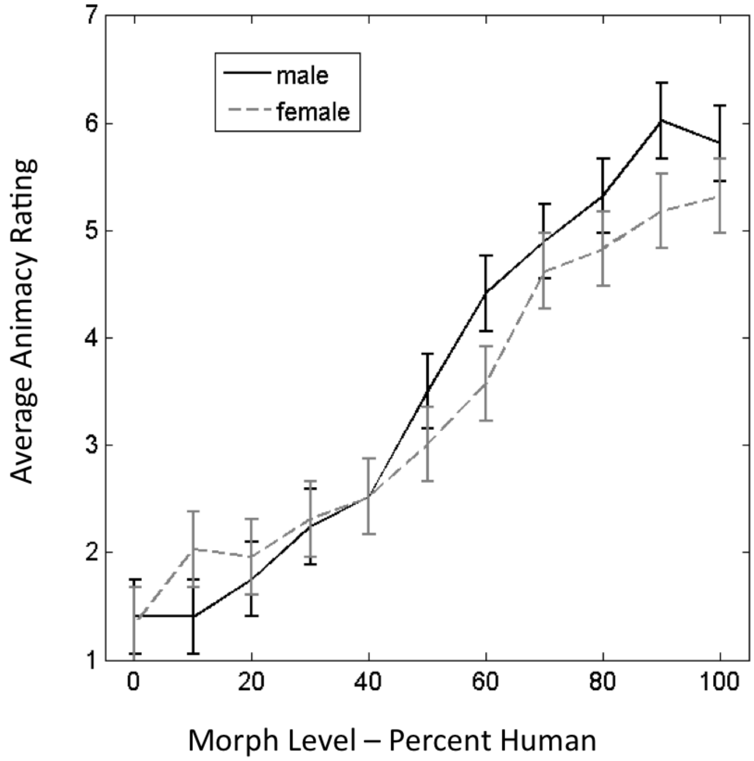

Figure 7.
Average animacy rating as a function of morph level and face sex, collapsing across participant sex. The black and gray lines illustrate the ratings for male and female face respectively. Error bars represent +/− 1 s.e.m. Female faces are reliably rated as less real than male faces, but only at the "real" end of the morph continuum.