

# Aftereffects support opponent coding of face gender

**Stephen Pond**

ARC Centre of Excellence in Cognition and its Disorders,  
School of Psychology, University of Western Australia,  
Crawley, WA, Australia

**Nadine Kloth**

ARC Centre of Excellence in Cognition and its Disorders,  
School of Psychology, University of Western Australia,  
Crawley, WA, Australia

**Elinor McKone**

Research School of Psychology, Australian National  
University & ARC Centre of Excellence in Cognition and  
its Disorders, Canberra, ACT, Australia

**Linda Jeffery**

ARC Centre of Excellence in Cognition and its Disorders,  
School of Psychology, University of Western Australia,  
Crawley, WA, Australia

**Jessica Irons**

Research School of Psychology, Australian National  
University & ARC Centre of Excellence in Cognition and  
its Disorders, Canberra, ACT, Australia

**Gillian Rhodes**

ARC Centre of Excellence in Cognition and its Disorders,  
School of Psychology, University of Western Australia,  
Crawley, WA, Australia



Many aspects of faces derived from structural information appear to be neurally represented using norm-based opponent coding. Recently, however, Zhao, Seriès, Hancock, and Bednar (2011) have argued that another aspect with a strong structural component, namely face gender, is instead multichannel coded. Their conclusion was based on finding that face gender aftereffects initially increased but then decreased for adaptors with increasing levels of gender caricaturing. Critically, this interpretation rests on the untested assumption that caricaturing the differences between male and female composite faces increases perceived sexual dimorphism (masculinity/femininity) of faces. We tested this assumption in Study 1 and found that it held for male, but not female faces. A multichannel account cannot, therefore, be ruled out, although a decrease in realism of adaptors was observed that could have contributed to the decrease in aftereffects. However, their aftereffects likely reflect low-level retinotopic adaptation, which was not minimized for most of their participants. In Study 2 we minimized low-level adaptation and found that face gender aftereffects were

strongly positively related to the perceived sexual dimorphism of adaptors. We found no decrease for extreme adaptors, despite testing adaptors with higher perceived sexual dimorphism levels than those used by Zhao et al. These results are consistent with opponent coding of higher-level dimensions related to the perception of face gender.

## Introduction

By looking at a face, we can determine many socially relevant attributes, including a person's gender, ethnicity, age, attractiveness, direction of attention, emotional state, and, if we know the person, their identity (Calder, Rhodes, Johnson, & Haxby, 2011). Recent progress in understanding how these attributes are perceptually coded has come from adaptation studies, in which exposure (adaptation) to a stimulus attribute alters subsequent neural processing and

Citation: Pond, S., Kloth, N., McKone, E., Jeffery, L., Irons, J., & Rhodes, G. (2013). Aftereffects support opponent coding of face gender. *Journal of Vision*, 13(14):16, 1–19, <http://www.journalofvision.org/content/13/14/16>, doi:10.1167/13.14.16.

produces perceptual aftereffects (for reviews, see Rhodes & Leopold, 2011; Webster & MacLeod, 2011).

Zhao, Seriès, Hancock, and Bednar (2011) have used face gender aftereffects (Webster, Kaping, Mizokami, & Duhamel, 2004) to investigate the neural mechanisms underlying the visual coding of face gender, a highly salient and socially important aspect of facial appearance. Specifically, they asked which one of two possible neural coding mechanisms, multichannel coding or norm-based opponent coding, is used to represent face gender.

In multichannel coding an attribute is represented by activation in many pools of neurons (or channels), each tuned to distinct values. In norm-based opponent coding, an attribute is represented by activation in two pools of neurons tuned to opposite ends of the stimulus dimension, with equal activation signaling a neutral point on that dimension. This neutral point is called the *norm*, and the coding is described as norm-based because the channels are tuned to deviations from the norm. Both forms of coding are used in the cortical visual system. For example, multichannel coding is used to represent tilt/orientation and spatial frequency (Blakemore & Sutton, 1969; Clifford, Wenderoth, & Spehar, 2000) and norm-based opponent coding to represent color (Webster & Leonard, 2008). For face attributes, norm-based opponent coding appears to be used for spatial configuration, identity-related dimensions, and expression (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Jeffery, Read, & Rhodes, 2013; Jeffery et al., 2010; Jeffery et al., 2011; Rhodes & Jeffery, 2006; Robbins, McKone, & Edwards, 2007; Skinner & Benton, 2010; Susilo, McKone, & Edwards, 2010), whereas a three-channel system appears to be used for gaze direction and head direction (Calder, Jenkins, Cassel, & Clifford, 2008; Lawson, Clifford, & Calder, 2011, but see Kloth & Schweinberger, 2010).

Zhao et al. (2011) sought to determine whether face gender is multichannel coded or opponent coded by examining how the size of face gender aftereffects changes with adaptor extremity. As illustrated in Figure 1, the two models make different predictions. Opponent coding (Figures 1A, B) predicts that aftereffects increase monotonically with increasing adaptor extremity over the full range of face masculinity/femininity values that occur in the real world (labeled “normal range” in Figure 1). The increase occurs because more extreme adaptors activate their preferred channel more strongly than less extreme adaptors, producing a stronger reduction in response with adaptation, and thus a larger aftereffect (larger shift in the crossover point at which the two pools are responding equally strongly). Beyond the normal range, aftereffects could either asymptote (Figure 1A) or continue to increase (Figure 1B).

In contrast, for multichannel coding (Figure 1C), aftereffects initially increase in magnitude as adaptors move away from the test image, but reach a maximum and then decrease for more extreme adaptors. We depict the decrease as occurring within the normal range (Figure 1C), on the assumption that channels tile the normal range to allow coding of natural variations in masculinity/femininity that occur in the real world. This type of nonmonotonic pattern is seen for multichannel-coded attributes like tilt and spatial frequency (Blakemore & Sutton, 1969; Clifford et al., 2000). The decrease for more extreme adaptor levels occurs because these adaptors have less impact on channels that respond to the test image than less extreme adaptors. For tilt, the maximum aftereffect occurs for relatively nonextreme adaptors that are only  $10^{\circ}$ – $15^{\circ}$  from the test orientation (Clifford, 2002). For face gender, beyond predicting that the maximum should occur somewhere within the normal range, it is difficult to predict its precise location, because this would depend on the spacing and tuning width of the channels.

Zhao et al. (2011) examined face gender aftereffects, measured as the shift in perceptual boundary (perceived androgyny), after exposure to adapting faces taken from gender continua. These continua were created by morphing between male and female averaged composite faces (labeled  $-0.5$  and  $+0.5$ , respectively) and by exaggerating, i.e., caricaturing, the physical differences between these composite faces. Their assumption was that this procedure would create continua with “extremely masculine” faces at one end ( $-2.5$ ) and “extremely feminine” faces ( $+2.5$ ) at the other (where a  $\pm 2.5$  face increased the difference from the original composite faces by 500%). They found significantly smaller gender aftereffects for more extreme (2.5, collapsed across gender) than less extreme (1.5) adaptors, and concluded that face gender is multichannel coded.

We found this conclusion surprising because there are large structural differences between male and female faces (Farkas, 1988), and, as noted above, many higher level structural attributes of faces (related to configuration, identity and expression) seem to be opponent coded.<sup>1</sup> One possibility is that Zhao et al.’s (2011) aftereffects actually reflect adaptation of low-level attributes, such as tilt and spatial frequency, which are multichannel coded in V1. They made no attempt to minimize low-level adaptation for five of their six participants. Moreover, the only participant tested with a size change between adapt and test faces had very small aftereffects that were very similar in size for 1.5 and 2.5 adaptors (no statistical comparison reported).

Another possibility is that Zhao et al.’s (2011) more highly caricatured faces (2.5) did not actually look more sexually dimorphic<sup>2</sup> than their less caricatured

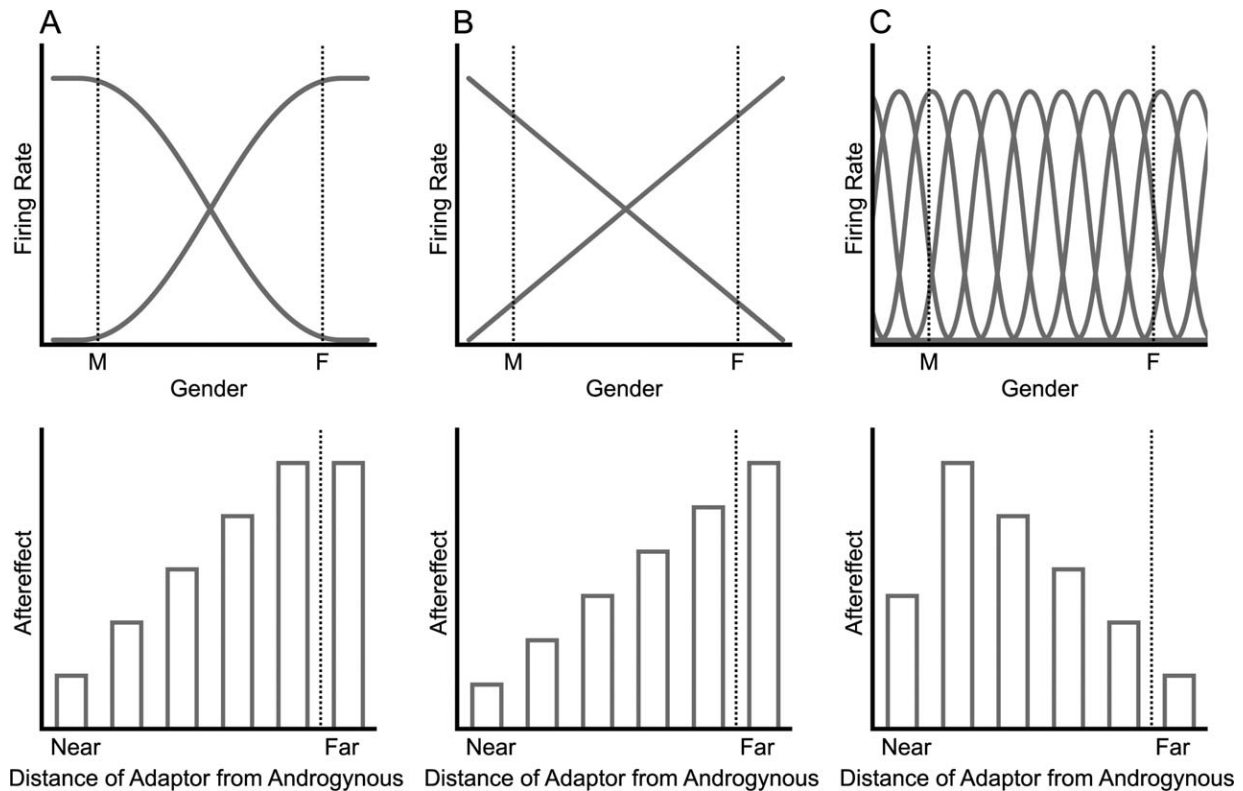


Figure 1. Top row: (A) Opponent coding of gender with one neural population responding to masculine (M) and the other to feminine (F) values. Equal activation of the two populations would signal an androgynous face. Vertical dashed lines indicate the limits of the normal range. In this version of the model, activation asymptotes at the limits of the normal range of real world faces. (B) An alternative opponent coding model in which activation continues to increase beyond the normal range. (C) A narrowband, multichannel model, with gender coded by multiple channels tuned to different values of sexual dimorphism, with many channels tuned to values within the normal range (to allow good discrimination of real world faces). It is not known whether additional channels would be tuned to values outside the normal range (as illustrated) or not. Bottom row shows the predictions of each model for how aftereffects change as a function of adaptor distance from an androgynous test face.

faces (1.5), in which case the observed decrease in aftereffects would not indicate multichannel coding of face gender. This important assumption was not tested, and inspection of Figure 2 (Rows 1–3) suggests that it may not be correct. For example, the 2.5 female faces have squarer jaws than the 1.5 faces, and a square jaw is a masculine trait. It is possible that averaged composite faces might not be ideal parent faces for creating gender continua, because the process of averaging can diminish masculine traits. For example, averaging male faces, each with a square jaw but of various widths, can produce a rounded jaw in the averaged composite. Averaging will also reduce coarse skin texture and any beard stubble, making complexions smoother. These changes will all reduce perceived masculinity, and may explain why Zhao et al.'s male average composites (−0.5) shown in Figure 2 do not look very masculine. Figure 2 also shows that some of their extreme caricatures have deformed features (e.g., ears and foreheads—see Figure 2), and a loss of realism might interfere with the perception of gender and reduce the size of aftereffects. Clearly, it is important to

understand how these faces vary in perceived sexual dimorphism and realism in order to interpret Zhao et al.'s results.

In Study 1 we obtained ratings of sexual dimorphism and realism for faces from gender continua made by caricaturing the differences between male and female averaged composite faces, including the continua used by Zhao et al. (2011). We asked whether perceived sexual dimorphism increases systematically across these continua (from most masculine to most feminine), and more specifically, whether it increases between Zhao et al.'s two critical levels (1.5 vs. 2.5). If it doesn't, then their observed decrease in aftereffects between these levels is not evidence for multichannel coding of gender.

In Study 2 we minimized the contribution of low-level, retinotopic adaptation, to determine whether higher level face-gender-related attributes are multichannel or opponent coded. Importantly, our adaptors that spanned a greater range of perceived sexual dimorphism than those used by Zhao et al (2011). If we replicate the nonmonotonic pattern found by Zhao et



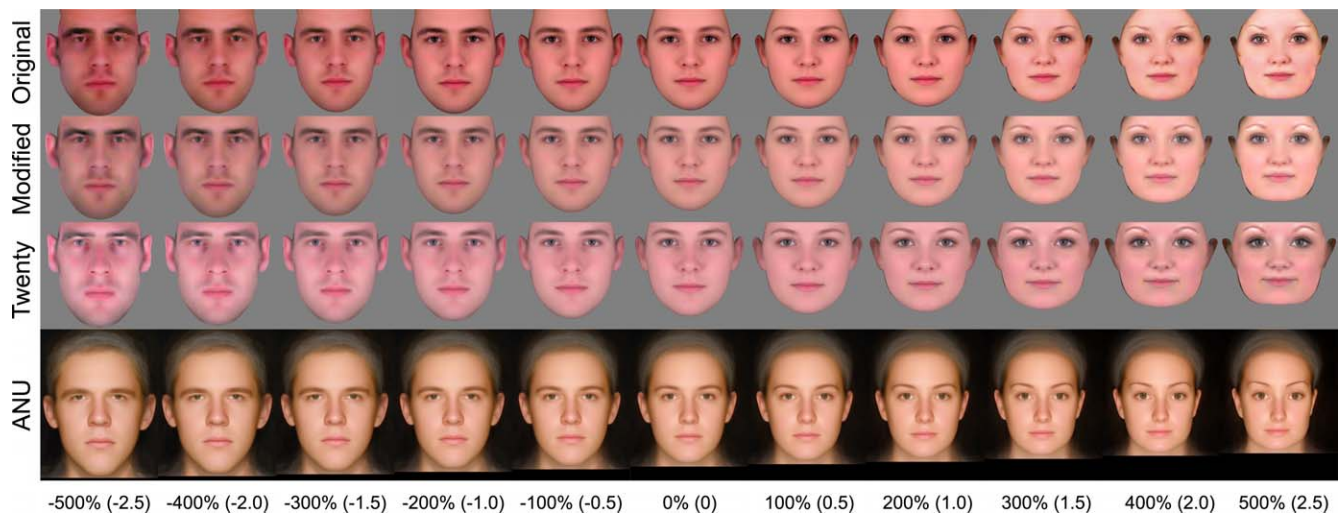


Figure 2. Faces from the three gender continua created by Zhao et al. (2011) (*original*, *modified*, and *twenty*) (adapted with permission from Zhao et al., 2011) and the *ANU* gender continuum. The *original* continuum manipulated gender by exaggerating the differences between male and female composites made from 150 images each. The *modified* continuum removed asymmetries and reduced color differences in the *original* continuum. The *twenty* continuum was identical to the *modified* continuum except that the composites were made using only 20% of the faces available for each gender. The undistorted male averaged composite is labeled  $-100\%$ , the undistorted female averaged composite is labeled  $100\%$ , and the androgynous composite is labeled  $0\%$ . The corresponding level labels used by Zhao et al. (2011) are shown in parentheses.

al. (2011), then it would suggest that face gender is multichannel coded. Alternatively, a monotonic increase would be consistent with opponent coding.

## Study 1

We asked participants to rate the perceived sexual dimorphism (masculinity/femininity) and realism of faces from: (a) the three continua used by Zhao et al. (2011), (b) faces from a new gender-caricatured continuum, and (c) natural male and female faces that varied in perceived masculinity and femininity. The new gender-caricatured continuum (*ANU* (Australian National University) in Figure 2) minimized morphing artifacts seen in Zhao et al.'s (2011) continua (e.g., misshapen male heads in their *original* continuum and ear distortions in all their continua, see Figure 2). This study provides a direct test of Zhao et al.'s critical assumption that the perceived sexual dimorphism of their adaptors increased over the range where gender aftereffects decreased (from caricature levels 1.5 to 2.5).

This study also allowed us to address several other questions of general interest. First, do continua created by caricaturing the differences between male and female averaged composites vary linearly, or at least monotonically, in perceived sexual dimorphism, as assumed? Second, does this caricaturing create faces with higher perceived sexual dimorphism than natural faces? Third, do averaged composite faces, particularly

male composites capture normal levels of sexual dimorphism?

For ease of exposition, we use morph level labels to indicate the degree to which faces have been morphed away from the undistorted male and female parent faces, which we label  $-100\%$  and  $+100\%$ , respectively (cf.  $-0.5$  and  $+0.5$  in Zhao et al.). An androgynous face has a value of  $0\%$ , and increasingly gender-caricatured male and female faces have increasing negative and positive values, respectively. In our numbering system, Zhao et al.'s critical levels of 1.5 and 2.5 become 300% and 500%, respectively.

Zhao et al. (2011) measured gender aftereffects (for Caucasian faces) in three Chinese and three Caucasian participants, with similar results for both ethnicities. We also tested Chinese and Caucasian participants to ensure optimal applicability of our ratings to interpreting their aftereffect data.

## Method

### Participants

Sixteen adult participants (eight Caucasian, five female,  $M = 19.5$  years,  $SD = 1.2$ , range 18–21; eight Chinese, five female,  $M = 25.3$  years,  $SD = 5.3$ , range 20–33), recruited from the Australian National University through posters and word-of-mouth, completed the experiment for course credit or \$5. Ethics approval had been obtained and participants gave informed consent. The sample size was decided in advance based on evidence that sexual dimorphism ratings are highly

consistent across individuals (e.g., Koehler, Simmons, Rhodes, & Peters, 2004).

### Stimuli

The stimuli consisted of front views of Caucasian faces with neutral expressions, taken from male-female continua created by caricaturing the differences between male and female composite faces using standard morphing methods (Figure 2). Three of the continua were created by Zhao et al. (2011): *original*, in which male and female averaged composites were made from 150 male faces and 150 female faces; *modified*, in which the original set was altered to enforce symmetry and reduce color differences between the male and female composites; and *twenty*, which was the same as the modified continuum except that the male and female composites were created using only 20% of the total pool of faces (i.e., 30 males and 30 females) (for details, see Zhao et al., 2011).

A new fourth continuum, labeled *ANU*, was created to avoid morphing artifacts. The parent faces were male and female averaged composites made from 26 male and 53 female face images, respectively, taken from the Australian National University face database.<sup>3</sup> Composites were created using Abrosoft FantaMorph5, with landmark points ( $N = 167$ , males;  $N = 175$ , females) placed by hand on each face and used to align the faces for morphing. Gender continua were created in FantaMorph, by morphing each composite away from the other to increase physical gender differences and by morphing between the two composites to reduce them. Due to software limitations only shape information was caricatured in the *ANU* continua, whereas color differences were also caricatured in Zhao et al.'s (2011) continua. Images were selected from each continuum to span the full range from  $-500\%$  to  $500\%$  caricatures. These consisted of 41 images (in 25% steps) from the *ANU*, *modified*, and *twenty* continua, and 25 images from the *original* continuum (148 images in total).

To determine where Zhao et al.'s (2011) stimuli fell with respect to the normal range, and whether it is possible to perceptually caricature gender beyond the normal range, we also included 20 natural Caucasian faces (10 female, 10 male) taken from a large set ( $N = 166$  male faces,  $N = 196$  female faces) that had been rated previously on perceived masculinity and femininity (Rhodes, Simmons, & Peters, 2005). They varied in perceived sexual dimorphism and spanned the full range available in that set (mean masculinity of male faces = 4.8,  $SD = 1.1$ ; mean femininity of female faces = 4.9,  $SD = 1.4$ ; seven-point scales). All face images were cropped or masked to hide any detail above the hairline or below the chin. Participants viewed the stimuli at a

distance of 45 cm, using a chin rest, so that faces subtended approximately  $16^\circ \times 19^\circ$  of visual angle.

### Procedure

Participants rated the perceived sexual dimorphism of each face using a 15-point rating scale ranging from  $-7$  (extremely masculine) through zero (androgynous) to seven (extremely feminine). The rating scale was displayed below each face, and participants responded by selecting a level on the scale using a standard computer mouse. Each face remained on screen until the participant responded. An interstimulus interval of 300 ms separated each trial. Faces from each continuum were presented in separate blocks, each of which also contained the 20 natural faces. Block order was identical for each participant (*original*, *modified*, *twenty*, and *ANU*), with trial order randomized within blocks.

After rating sexual dimorphism, participants rated how much each face looked like a real person, on a scale from one (not at all like a real person) to 10 (extremely like a real person). The procedure was the same as for sexual dimorphism ratings except for the response scale.

In each task, prior to rating any faces, participants were shown 10 example faces simultaneously for 20 s. They were selected to illustrate the range of caricature levels and types of images that would be shown (specific images shown: Zhao *original* 0.92 &  $-0.96$ , Zhao *modified* 1.54 &  $-2.38$ , Zhao *twenty* 0.36 &  $-1.6$ , *ANU*, 2.20 &  $-0.30$ , *ANU* male and female averages).

## Results and discussion

We conducted three-way repeated-measures ANOVAs for both sexual dimorphism and realism ratings, with continuum (*original*, *modified*, *twenty*, *ANU*) and morph level ( $-500\%$ ,  $-400\%$ ,  $-300\%$ ,  $-200\%$ ,  $-100\%$ ,  $0\%$ ,  $100\%$ ,  $200\%$ ,  $300\%$ ,  $400\%$ ,  $500\%$ ) as repeated-measures factors and rater ethnicity (Caucasian or Asian) as a between-participants factor.<sup>4</sup> The morph levels used in the analysis were selected from the large number that had been rated to give a set that was common to all four continua and that spanned the full range used by Zhao et al. (2011). Degrees of freedom were adjusted using Greenhouse-Geisser correction whenever Mauchly's test indicated that the assumption of sphericity had been violated. Planned  $t$  tests with Bonferroni correction for multiple comparisons were used to test for differences between morph levels.

### Sexual dimorphism ratings

Results for perceived masculinity/femininity are plotted in Figure 3. There were no differences between

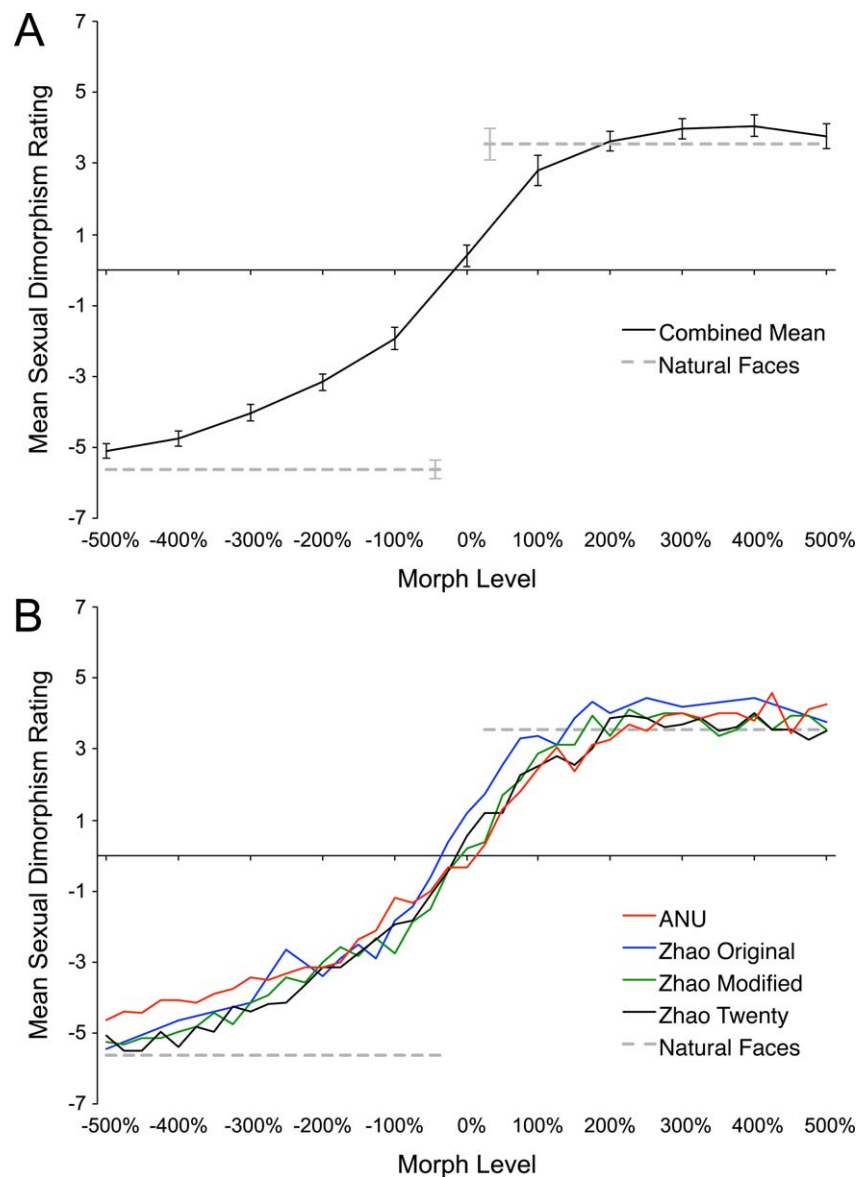


Figure 3. (A) Mean ratings of sexual dimorphism ( $-7$ , extremely masculine;  $+7$ , extremely feminine) as a function of gender morph level ( $-500$ , extremely male; zero androgynous;  $+500$ , extremely female), collapsed across continua, and for natural faces. *SEM* bars are shown. (B) Mean ratings of sexual dimorphism shown separately for each continuum and for natural faces.

the *ANU* continuum and Zhao et al.'s (2011) continua, with no significant main effect or interactions involving continuum, all  $F_s > 1.44$   $p_s > 0.25$ ,  $\eta_p^2 < 0.09$ . Thus, the morphing artifacts present in Zhao et al.'s stimuli do not seem to affect perceived sexual dimorphism. Nor does color information (present in Zhao et al. continua, absent in *ANU* continuum), suggesting possible redundancy in shape and color cues to sexual dimorphism. Finally, there was no significant main effect or interactions involving ethnicity, all  $F_s < 1.75$ ,  $p_s > 0.08$ ,  $\eta_p^2 < 0.12$ .

We found that continua created by caricaturing the differences between male and female averaged composites increased monotonically in perceived sexual

dimorphism, but only up to and not beyond the levels seen in natural faces (Figure 3). Therefore, the apparently extreme  $\pm 500\%$  ( $\pm 2.5$ ) adaptors used by Zhao et al. (2011) are not perceived as extremely sexually dimorphic, but rather fall within the range found in natural faces. There was also an asymmetry between the masculine and feminine ends, with ratings approaching natural levels more slowly for the masculine than feminine end.

ANOVA confirmed a significant main effect of morph level,  $F(2.84, 39.81) = 171.22$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.92$ . Caricaturing female composites did not increase perceived femininity, all  $p_s > 0.46$  (over levels  $+100\%$  to  $+500\%$ ). Caricaturing male composites increased



perceived masculinity, with significant increases from  $-100\%$  to  $-200\%$ ,  $-200\%$  to  $-300\%$ , and  $-300\%$  to  $-500\%$ , all  $ps < 0.05$ . However, this caricaturing only brought perceived masculinity up to the level of natural male faces, with no significant difference between  $-500\%$  caricatures ( $M = -5.1$ ,  $SD = 1.1$ ) and natural male faces ( $M = -5.6$ ,  $SD = 0.8$ ),  $t(15) = 1.62$ ,  $p = 0.127$ . Across the full continuum ( $-500\%$  to  $500\%$ ), planned contrasts indicated a significant linear trend,  $F(1, 14) = 355.05$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.96$ . There were also strong, significant quadratic, cubic, fifth, and seventh order trends, all  $F_s > 13.58$ ,  $ps < 0.002$ ,  $\eta_p^2 > 0.49$ .

Male averaged composites ( $-100\%$ ) looked substantially, and significantly, less masculine ( $M = -1.9$ ,  $SD = 1.2$ ) than natural male faces ( $M = -5.6$ ,  $SD = 0.8$ ),  $t(15) = 12.67$ ,  $p < 0.0001$ , supporting our conjecture that such composites do not capture masculine traits very well. Female averaged composites also tended to look less feminine ( $M = 2.8$ ,  $SD = 1.6$ ) than natural female faces ( $M = 3.6$ ,  $SD = 1.4$ ), but this difference was not significant,  $t(15) = 1.66$ ,  $p = 0.228$ .

For the critical levels of  $300\%$  and  $500\%$ , between which aftereffects decreased in Zhao et al. (2011), perceived masculinity was stronger for  $-500\%$  than  $-300\%$  (i.e., male) faces,  $p < 0.011$ , supporting Zhao et al.'s assumption. However, perceived femininity was not stronger for  $+500\%$  than  $+300\%$  (i.e., female) faces,  $p = 1.00$  (Figure 3).

### Realism ratings

Realism ratings, shown in Figure 4, indicated that faces right across the gender continua looked much less realistic than natural faces, especially for extreme caricatures. The highest mean rating was 5.0 (95% confidence interval [CI] of 4.1, 5.9), for  $+200\%$  faces, which was much lower than ratings for natural female ( $M = 9.0$ ,  $SD = 1.3$ ) or male ( $M = 8.9$ ,  $SD = 1.4$ ) faces,  $t_s > 6.70$ ,  $ps < 0.0001$ . There was a significant effect of morph level,  $F(2.25, 31.50) = 7.06$ ,  $p < 0.002$ ,  $\eta_p^2 = 0.34$ , with no other significant effects (continuum, rater ethnicity, interactions), all  $F_s < 1.53$ ,  $ps > 0.24$ ,  $\eta_p^2 < 0.10$ . The most extreme female morph level ( $500\%$ ) looked significantly less realistic than several other levels ( $+300\%$ ,  $+200\%$ ,  $-200\%$ ,  $-300\%$ ,  $-400\%$ ), including the critical comparison level of  $+300\%$ , all  $ps < 0.026$ . No other differences between levels were significant, using the conservative Bonferroni correction, although the most extreme male level ( $-500\%$ ) was significantly less realistic than the  $-300\%$  level, without this correction,  $p < 0.015$ . The  $\pm 500\%$  ratings were six points (on a 10-point scale) below natural faces, indicating a substantial lack of realism.

### Summary

Zhao et al.'s (2011) critical assumption that perceived sexual dimorphism increased over the range where their gender aftereffects decreased ( $300\%$  vs.  $500\%$ ) held for male, but not female faces. Therefore, if their decrease in aftereffects at the male end of their continua was significant (they collapsed across gender), then it could indicate multichannel coding. However, as noted above, this could reflect adaptation of low-level mechanisms such as those coding tilt and/or spatial frequency. We also observed a decrease in realism over the two critical levels that could have decreased any high-level (although not low-level) aftereffects by interfering with perception of gender and/or reducing face-selective adaptation.

Surprisingly, caricaturing the physical differences between male and female averaged composite faces did not increase perceived sexual dimorphism beyond the levels seen in natural faces. Therefore, even Zhao et al.'s (2011) most "extreme" ( $\pm 500\%$ ) faces were not perceived as more masculine or feminine than natural male and female faces. A possible reason for this lack of hypergender in extreme male morphs is that the averaged male composite faces did not capture some typical masculine traits (square jaw, stubble texture) and looked substantially less masculine than natural male faces. None of these findings were unique to Zhao et al.'s continua; the ANU continuum had the same properties.

## Study 2

In Study 2 we examined how face gender aftereffects change with adaptor sexual dimorphism when low-level retinotopic adaptation is minimized. Our aim was to determine whether higher level face gender aftereffects increase with adaptor extremity, consistent with opponent coding, or show an initial increase followed by a decrease, consistent with multichannel coding. Critically, we confirmed that our most extreme adaptors were perceived as more sexually dimorphic than those used by Zhao et al. (2011).

The results of Study 1 suggested that it may be difficult to increase perceived sexual dimorphism beyond natural levels by caricaturing the differences between male and female averaged composites. Therefore, we also used adaptors from single-identity continua made by exaggerating the differences between highly sexually dimorphic male and female parent faces (Figure 5). By using parent faces that looked more sexually dimorphic than the natural faces used in Study 1, we ensured that our adaptors would span a greater range than Zhao et al.'s (2011) (which did not exceed the perceived sexual dimorphism of those natural faces).

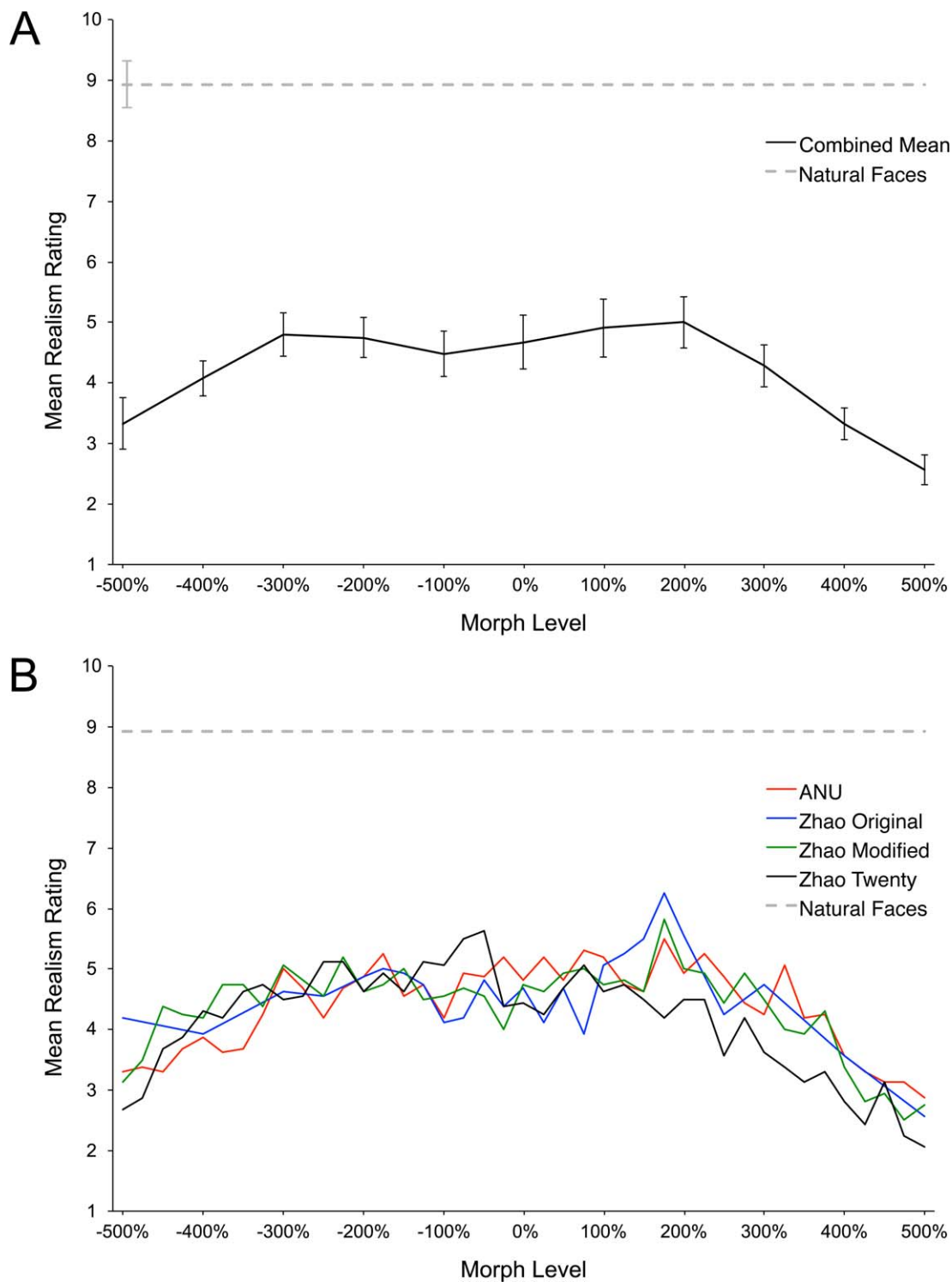


Figure 4. (A) Mean ratings of realism (1 = not at all like a real person, 10 = extremely like a real person) as a function of gender morph level (–500, extremely male; zero androgynous; +500, extremely female), collapsed across continua, and for natural faces. SEM bars are shown. (B) Mean ratings of realism shown separately for each continuum and natural faces.

**Method**

**Participants**

Thirty-six Caucasian adults (34 female,  $M = 19.8$  years,  $SD = 4.4$ , range 17–40), recruited from the

Introductory Psychology course at the University of Western Australia (UWA), completed the adaptation task for course credit. This sample size was determined in advance, based on previous aftereffect studies that estimated aftereffects at a single point as done here



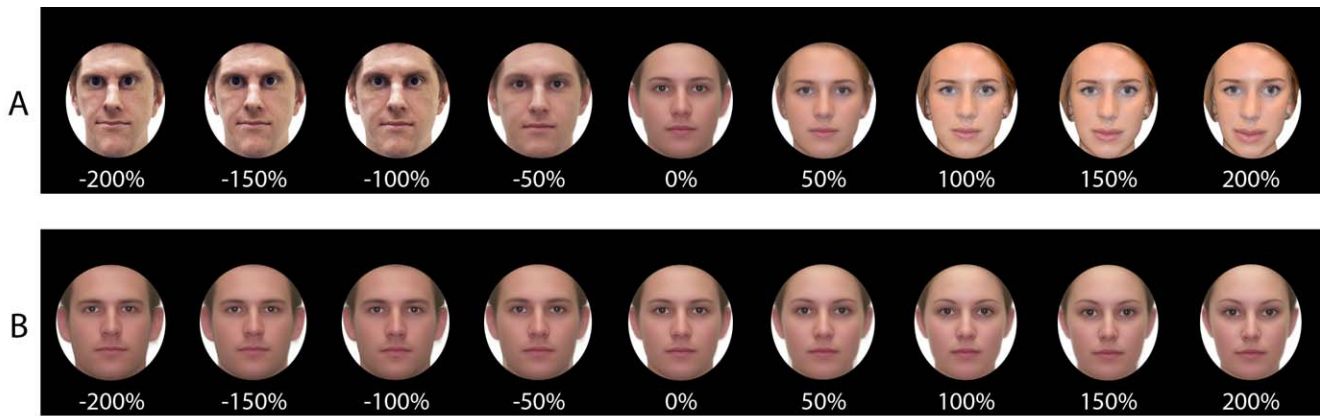


Figure 5. (A) Example of a gender continuum made using natural male (–100%) and female (100%) parent faces (single-identity continuum). The continuum shown here is for illustrative purposes only, because we do not have permission to publish the parent faces used in the study. (B) Gender continuum made using averaged-composite male (–100%) and female (100%) parent faces (UWA composite continuum). All continua were made by caricaturing the parent faces away from, and towards, a 0% androgynous composite that was the average of 24 male and 24 female faces. The single-identity parent faces were selected to be highly sexually dimorphic, and looked more masculine/feminine than the faces in Study 1 (including those used by Zhao et al., 2011).

(e.g., Fiorentini, Gray, Jeffery, Rhodes, & Pellicano, 2012; Jeffery et al., 2011). An additional 12 Caucasian adults (eight female,  $M = 30.3$  years,  $SD = 13.2$ , range 21–55) completed the rating tasks as volunteers. This smaller sample size was deemed appropriate because sexual dimorphism ratings are highly consistent across individuals (e.g., Koehler et al., 2004). Ethics approval had been obtained and participants gave informed consent.

### Stimuli

Eleven face continua, each ranging from strongly masculine (–200%) to strongly feminine (+200%), were created using 11 male and 11 female face images (see Figure 5 for examples). Ten continua were made using highly sexually dimorphic natural male and female parent faces ( $\pm 100\%$  faces in Figure 5). These highly masculine ( $M = 5.3$ ,  $SD = 0.4$ , scale maximum = 7) and highly feminine ( $M = 5.4$ ,  $SD = 0.1$ ) faces were taken from the same large database as the natural faces in Study 1 (Rhodes, Simmons, & Peters, 2005). They were more extreme than the natural male ( $M = 4.8$ ,  $SD = 1.1$ ) and female ( $M = 4.9$ ,  $SD = 1.4$ ) faces used in Study 1, and therefore more extreme than all of Zhao et al.’s (2011) faces, which never exceeded the perceived dimorphism of the natural faces in Study 1 (see Figure 3). The 11th continuum was based on male (–100%) and female (+100%) averaged composite parent faces taken from Rhodes et al. (2011). These were made from 24 male and 24 female faces, respectively. All face images were front views of Caucasian young adults with neutral expressions and

minimal facial hair or makeup. Minor alterations were made using Adobe Photoshop to remove blemishes and piercings.

To make the gender continua, each 100% parent face was caricatured towards and away from an androgynous face (0%), made by averaging the male and female averaged composites. The caricaturing was done using Abrosoft Fantamorph, so that differences of each parent face from the androgynous face (i.e., sexual dimorphism) were minimized or exaggerated, respectively. Images from these continua were exported at 50%, 150%, and 200%. The maximum caricature level was set at 200% because further caricaturing resulted in clearly unnatural-looking faces, particularly for the single-identity continua.

In total, 11 face continua were created, each with nine caricature levels (–200%, –150%, –100%, –50%, 0%, 50%, 100%, 150%, 200%). Shape information varied continuously across these levels, but color information varied only between –100% and +100% due to software limitations. A black oval mask covered most of the hair (Figure 5). These 99 faces were used as adapting stimuli. They subtended approximately  $5.5^\circ$  in height and  $4.3^\circ$ – $6.5^\circ$  in width, at the viewing distance of 70 cm. Test faces were 75% of the adaptor size, approximately  $4.1^\circ \times 3.2^\circ$ – $4.9^\circ$ . The test face used to measure aftereffects was an androgynous morph (20%) from the composite continuum. We used the 20% morph because pilot testing ( $N = 12$ ) indicated that it looked more androgynous than the 0% morph, which looked slightly male (see Figure 5). On 16% of the trials –100% and +100% faces from the averaged-composite

continuum were used as test faces to provide some easy trials and help maintain motivation.

### Adaptation procedure

Each trial began with an adaptor face for 5000 ms, followed by a fixation cross for 200 ms, and then a test face for 500 ms. Participants indicated whether the test face was male or female by pressing one of two labeled keys. There were 432 trials. The test face was the perceptually androgynous face (20%) on 360 trials, and the  $-100%$  and  $100%$  faces from the averaged-composite continuum on the remaining 72 trials. For trials with the androgynous test face, each of the 90 adaptor faces of the single-identity face continua (10 for each of the nine caricature levels) was repeated twice and each of the nine adaptor faces of the averaged-composite continuum was repeated 20 times, resulting in 180 trials for each continuum type. Trials with unambiguous test faces were also split evenly between single-identity and averaged-composite adaptors with equal numbers at each of the nine adaptor levels. The task lasted approximately 50 minutes and was split into nine blocks of 48 randomly chosen trials that differed between participants. Five practice trials randomly selected from all experimental trials preceded the experiment proper.

### Rating procedure

The rating tasks were similar to those of Study 1. Sexual dimorphism was always rated first as this was the primary variable of interest. Initially, a subset of 45 randomly ordered face images were presented one at a time for the participant to view at their own pace. Stimuli in this phase ranged across all identities and caricature levels, and participants were asked to consider how they would apply the sexual dimorphism rating scale given the range of faces in the experiment. Next, participants rated faces for how masculine or feminine they appeared on a 15-point scale ranging from  $-7$  (extremely masculine), through zero (androgynous), to seven (extremely feminine). Each trial began with a 200 ms fixation cross, followed by a face image that remained on screen until the participant responded via keyboard. There were 99 trials, with each face (nine faces from 11 continua) presented once in random order. Nine practice trials, with one face from each caricature level (all from single-identity continua), preceded the main task. After completing the sexual dimorphism ratings, participants rated all the faces for realism on a scale ranging from 1 (not at all like a real person) to 10 (extremely like a real person). The procedure was otherwise identical to the sexual dimorphism rating task. Note that we did not include a

separate set of natural faces, because  $100%$  faces in the single-identity continua are natural faces.

## Results and discussion

### Aftereffects

We measured gender aftereffects as the bias to see a perceptually androgynous test face as male (proportion of male responses minus proportion of female responses). Therefore, male adaptors should bias participants to see the test faces as female, resulting in negative valued aftereffects, and female adaptors should bias them to see the test faces as male, resulting in positive valued aftereffects. Larger absolute values indicate larger aftereffects. Because there are many adaptor levels, we focus on trend analysis rather than pairwise comparisons. However, where specific pairwise comparisons are of particular interest, we report Bonferroni-corrected significance levels.

An initial two-way repeated-measures ANOVA, with adaptor caricature level ( $-200%$ ,  $-150%$ ,  $-100%$ ,  $-50%$ ,  $20%$ ,  $50%$ ,  $100%$ ,  $150%$ ,  $200%$ ) and continuum type (single identity or composite) as repeated-measures factors, yielded a significant main effect of adaptor caricature level,  $F(2.70, 94.53) = 58.49$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.63$ , which interacted with continuum type,  $F(8, 280) = 3.37$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.09$  (Figure 6). There was no main effect of continuum type,  $F(1, 35) = 0.00$ ,  $p > 0.05$ . To explore the significant interaction, we conducted separate one-way ANOVAs for each type of continuum.

For the single-identity continua, there was a significant main effect of adaptor morph level,  $F(4.74, 165.93) = 36.02$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.51$ . Figure 6 shows that aftereffects increased from  $-200%$  through to  $+100%$ , with no further increase. The linear trend was significant,  $F(1, 35) = 102.00$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.75$ . There were also smaller, but significant, cubic,  $F(1, 35) = 14.02$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.29$ , and fourth order trends,  $F(1, 35) = 5.77$ ,  $p < 0.03$ ,  $\eta_p^2 = 0.14$ , reflecting the flattening at the female end (i.e., beyond  $100%$ ). Importantly, there was no evidence of the nonmonotonic pattern (initial increase followed by subsequent decrease) expected with multichannel coding. The observed pattern is consistent with opponent coding of gender.

For the composite continuum, there was a significant main effect of adaptor morph level,  $F(3.00, 105.12) = 39.56$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.53$ . Aftereffects increased numerically over the entire range, except for one small, nonsignificant dip on the female side near androgyny (Figure 6). The linear trend was significant,  $F(1, 35) = 72.82$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.68$ , and the only other significant effect was a seventh order trend,  $F(1, 35) = 11.71$ ,  $p < 0.002$ ,  $\eta_p^2 = 0.25$ . Again, there was no evidence of the nonmonotonic pattern expected with

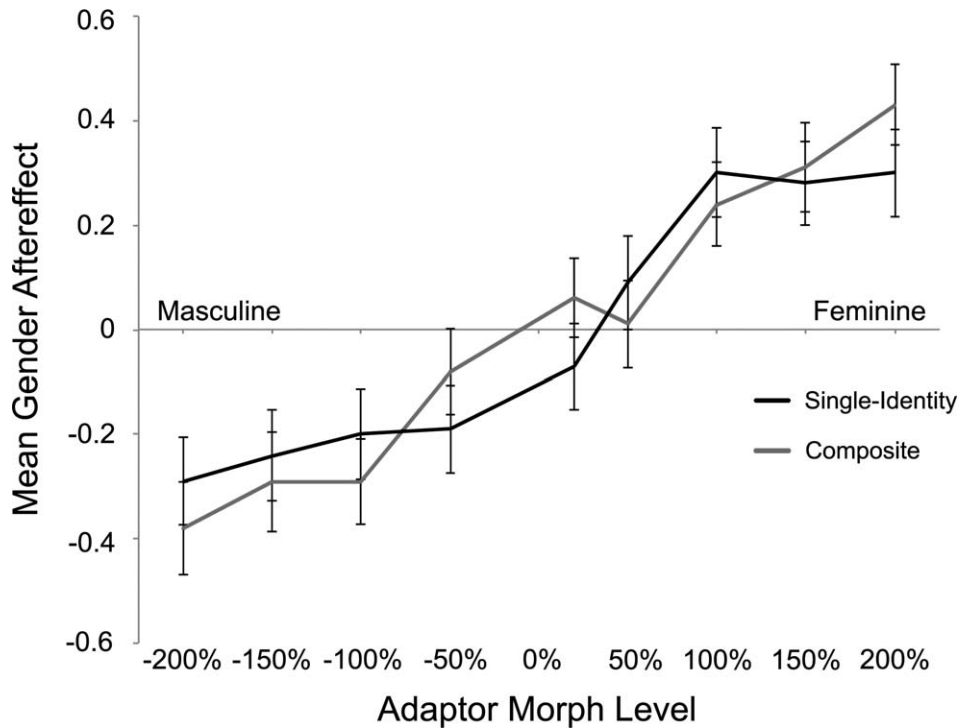


Figure 6. Mean gender aftereffects as a function of adaptor morph level for single-identity and UWA composite continua. *SEM* bars are shown. Negative values indicate that the androgynous test face was judged as female (expected for male adaptors), and positive values indicate that it was judged as male (expected for female adaptors). Larger absolute values indicate larger aftereffects.

multichannel coding. The observed pattern supports opponent coding of gender.

### Sexual dimorphism ratings

Figure 7 shows that perceived sexual dimorphism increased numerically over the entire range of the composite continuum, and increased from  $-100\%$  to  $+50\%$ , plateauing at the level of the high-dimorphism  $+100\%$  natural female faces, for the single-identity continua. Importantly, comparison with Figure 6 (aftereffects) shows that gender aftereffects increased (numerically) over the full range of morph levels for which sexual dimorphism increased, supporting opponent coding.

A preliminary two-way repeated measures ANOVA on mean sexual dimorphism ratings yielded significant main effects of adaptor morph level,  $F(2.58, 28.41) = 260.52$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.96$ , and continuum type,  $F(1, 11) = 21.09$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.66$ , and a significant interaction,  $F(3.71, 40.85) = 15.34$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.58$  (Figure 7). To follow up the interaction, we conducted separate one-way ANOVAs for each type of continuum.

For the single-identity continua, there was a significant effect of adaptor morph level,  $F(2.07, 22.77) = 464.47$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.98$ . Inspection of Figure 7 shows that masculinity ratings reached a maximum at

$-100\%$  (high-dimorphism natural male faces), and femininity ratings reached a maximum at  $+50\%$  (at the level of the  $100\%$  high-dimorphism natural female faces), with no further increases for more extreme morph levels. The slight decrease in femininity ratings beyond  $+50\%$  was not significant, all  $ps = 1.00$  (Bonferroni corrected). Therefore, our gender continua made using natural parent faces did not exaggerate perceived sexual dimorphism beyond natural levels. Overall, there was a strong linear trend,  $F(1, 11) = 692.49$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.98$ , and several other significant trends (quadratic, cubic, Orders 4, 5, 7, and 8), all  $Fs > 22.93$ ,  $ps < 0.001$ ,  $\eta_p^2 > 0.68$ .

For the composite continuum, there was a significant effect of adaptor morph level,  $F(3.53, 38.86) = 78.97$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.88$ . Sexual dimorphism ratings increased monotonically across the full range (Figure 7), with a strong, significant linear trend,  $F(1, 11) = 225.54$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.95$ . There were several other significant trends (quadratic, cubic, Order 4), but all with small effect sizes, all  $Fs > 5.35$ ,  $ps < 0.05$ ,  $\eta_p^2 > 0.33$ . In contrast to the composite continua in Study 1 (and the single-identity continua in this study), perceived femininity on this new composite continuum increased beyond normal levels, with the most extreme adaptor ( $200\%$ ) rated as significantly more feminine than high-dimorphism natural female faces ( $100\%$ , single-identity continuum),  $t(11) = 6.37$ ,  $p < 0.0001$ . As before, however, masculinity was not increased beyond



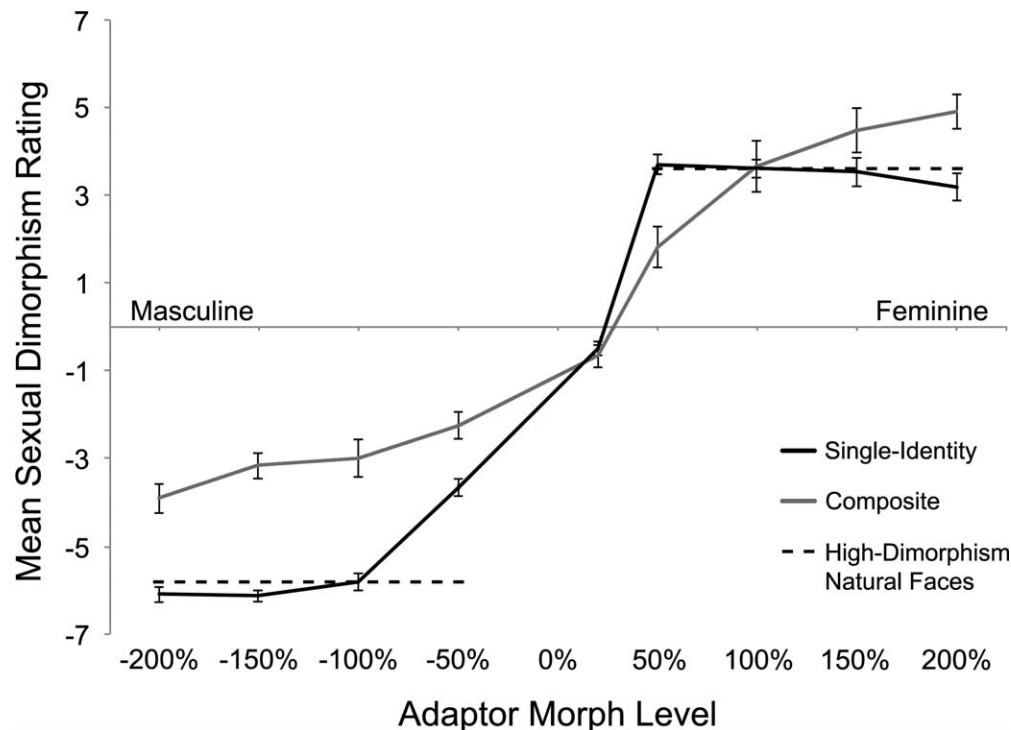


Figure 7. Mean sexual dimorphism (masculinity/femininity) rating as a function of adaptor morph level, for each continuum type. SEM bars are shown. The dashed lines indicate the ratings given to the highly dimorphic natural male (–100%) and female (100%) parent faces used to make the single-identity continua.

the level seen in natural faces (–100%, single-identity continuum), with the most extreme male adaptor (–200%) rated as marginally less masculine than the high-dimorphism natural male faces,  $t(11) = 2.32$ ,  $p = 0.08$  (Bonferroni corrected  $N = 2$ ).

### Realism ratings

Figure 8 shows that the new composite continuum produced realistic faces, unlike those examined in Study 1, whose faces were rated as much less realistic than natural faces. However, realism of exaggerated faces on the single-identity continua dropped below that of natural faces.

A two-way repeated measures ANOVA yielded a significant main effect of adaptor morph level,  $F(3.35, 36.86) = 10.41$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.49$ , and a significant interaction,  $F(8, 88) = 16.65$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.60$ . There was no main effect of continuum type,  $F(1, 11) = 3.72$ ,  $p = 0.08$ ,  $\eta_p^2 = 0.25$ . Given the significant interaction, we conducted separate one-way ANOVAs for each type of continuum.

For the single-identity continua, realism ratings varied from 3.3 to 7.6 (cf. 6.6 for natural faces), with a significant effect of adaptor morph level,  $F(1.76, 19.40) = 23.78$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.68$ . There was a complex relationship between morph level and degree of realism (Figure 8), with several significant trends (linear,

quadratic, cubic, Orders 6, 7, 8), all  $F_s > 7.36$ ,  $p_s < 0.02$ ,  $\eta_p^2 > 0.40$ . Androgynous images were rated significantly lower than neighboring levels, Bonferroni-corrected  $p_s < 0.05$ , possibly due to the low frequency of such faces in the population. More generally, realism ratings declined towards the ends of the single-identity continua (from 100% to 200%), Bonferroni-corrected  $p_s < 0.0001$ .

For the composite continuum, there was a significant effect of adaptor morph level,  $F(8, 88) = 2.48$ ,  $p < 0.02$ ,  $\eta_p^2 = 0.18$ . Only fourth and sixth order trends were significant, both  $F_s > 7.00$ ,  $p_s < 0.03$ ,  $\eta_p^2 > 0.39$ . The only significant pairwise difference was that the androgynous face looked less realistic than the female averaged composite (+100%), Bonferroni-corrected  $p < 0.05$ . It is clear from Figure 8 that realism levels were relatively high (i.e., equal to natural faces) right across the composite continuum, and were similar to those of natural faces (–100%  $M = 7.0$ ,  $SE = 0.4$ ; 100%  $M = 6.1$ ,  $SE = 0.4$ ).

### Relating gender aftereffects to perceived sexual dimorphism and realism of adaptors

Comparison of Figures 6 and 7 suggests a strong correspondence between the size of gender aftereffects and the perceived sexual dimorphism of adaptors. To more formally assess this relationship, we correlated



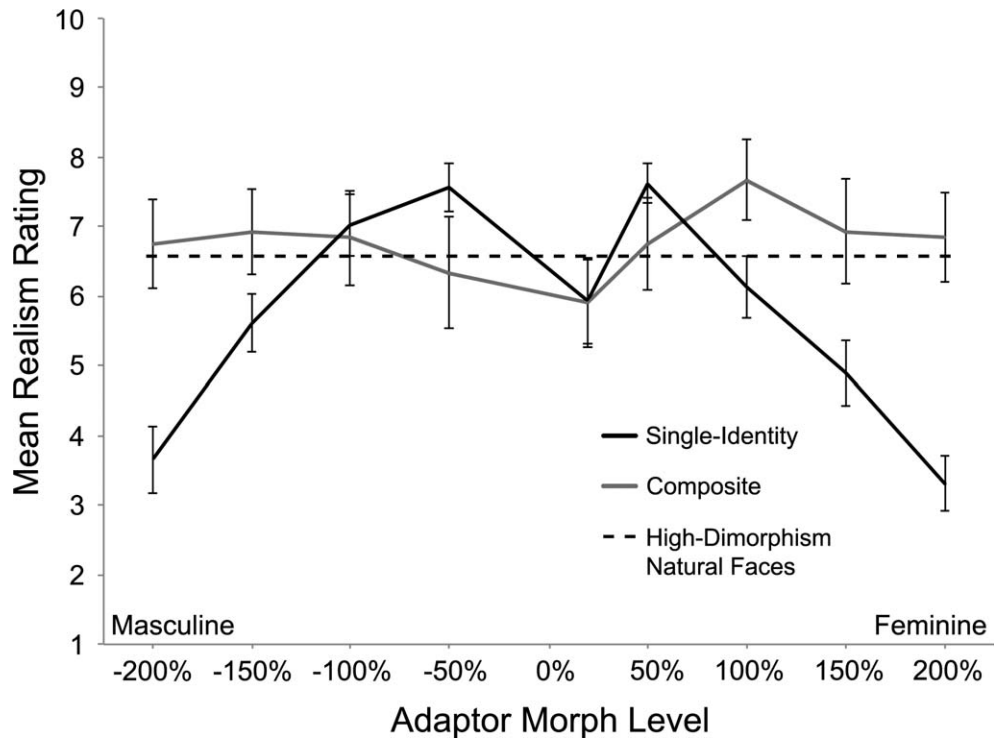


Figure 8. Mean realism rating as a function of adaptor morph level, for each continuum type. SEM bars are shown. The dashed line indicates the mean rating for the highly dimorphic natural male (–100%) and female (100%) parent faces used to make the single-identity continua.

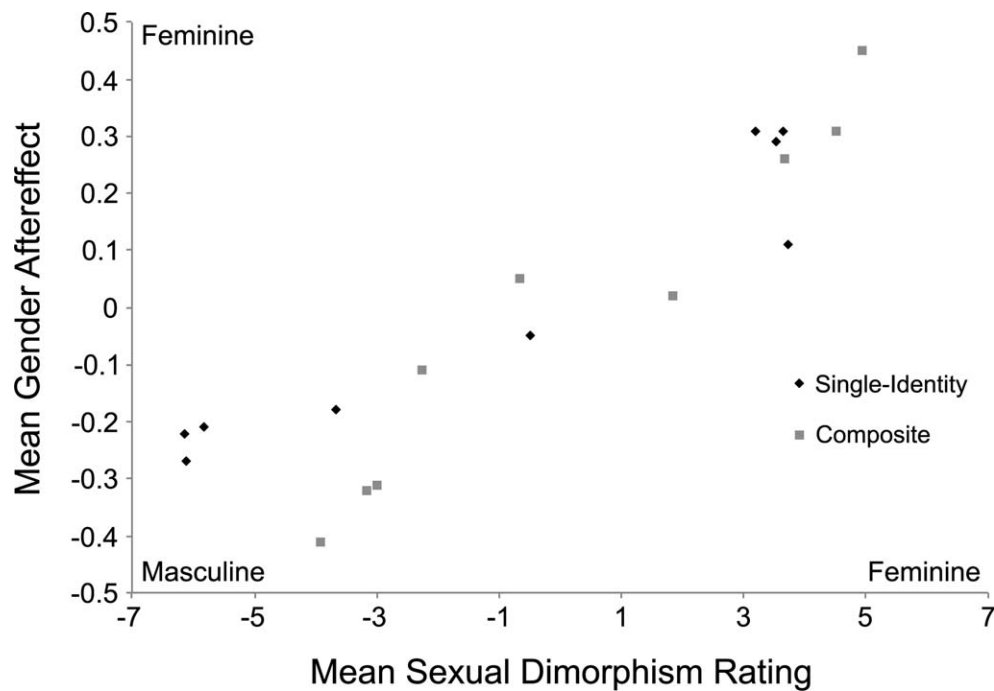


Figure 9. Scatterplot showing the strong, positive relationship between size of gender aftereffects and perceived sexual dimorphism of adaptors from single-identity and composite continua.

mean aftereffect size (averaged across participants) with the perceived sexual dimorphism of adaptors. We found a very strong positive correlation,  $r = 0.92$ ,  $n = 18$ ,  $p < 0.0001$  (Figure 9). The aftereffects did not correlate significantly with perceived realism of our adaptors,  $r = -0.09$ ,  $n = 18$ ,  $p < 0.72$ . (We note that realism was high, and the range relatively limited, in this study, so this result does not imply that aftereffects would never vary with realism). Overall, the size of face gender aftereffects is well explained by levels of perceived sexual dimorphism in the adapting stimuli. This strong positive relationship provides clear support for opponent coding of gender, with no evidence for the nonmonotonic pattern (increase followed by decrease) predicted by multichannel coding.

We note that our continua were created by caricaturing faces against an androgynous composite (created by averaging 24 male and 24 female faces). It was subsequently found that this composite looked slightly male, which means that all the images in the continuum may be slightly less masculine than they should be. However, this should not affect the size or direction of the correlation observed between perceived sexual dimorphism and size of aftereffects.

### Summary

With the contribution of low-level adaptation minimized, face gender aftereffects increased in magnitude as adaptors increased in perceived sexual dimorphism, supporting opponent coding. Importantly, our adaptors spanned a wider range of masculinity/femininity than was present in the stimuli of Zhao et al. (2011) and yet we found no decrease that would indicate multichannel coding. For single-identity continua, our 100% adaptors were selected from a previous study as having very high levels of sexual dimorphism within a large set of natural faces, and exaggerating the differences between them did not further increase their perceived sexual dimorphism. We were, however, able to increase the perceived dimorphism of female faces on our new composite continuum, and aftereffects continued to increase over that extended range.

## General discussion

Zhao et al. (2011) suggested that face gender is multichannel coded, based on their observation that adaptors with their strongest gender caricaturing (2.5) induced smaller gender aftereffects than adaptors with medium caricaturing (1.5). However, this interpretation assumes that the perceived sexual dimorphism of their adaptors increases monotonically with gender caricaturing, and particularly increases between 1.5 and 2.5.

When we tested this assumption in Study 1, we found that it held for their male adaptors, but did not hold for their female adaptors. Another problem we identified was that the realism of Zhao et al.'s adaptors was very low, and lower for 2.5 than 1.5 adaptors. Finally, low-level adaptation was not controlled for five of Zhao et al.'s six participants. Together, these results suggest that caution is needed in interpreting Zhao et al.'s drop in aftereffects from their 1.5 to 2.5 adaptors as evidence for multichannel coding of face gender.

In Study 2, we re-assessed how face gender is coded by measuring gender aftereffects for adaptors with a wider range of perceived sexual dimorphism than used by Zhao et al. (2011) and with low-level adaptation minimized. Our key finding was that face gender aftereffects increased over this full range. We found no evidence of any decrease in aftereffects for the most extreme adaptors. Rather we found a very strong positive relationship between adaptor sexual dimorphism and aftereffect size, as predicted by opponent coding.

Two other findings of interest emerged from our studies. First, male composites (average faces) were consistently rated as much less masculine than real male faces and caricaturing only raised their masculinity to natural levels. These results were replicated for five different composite face continua across Study 1 and Study 2. Second, female average composites had normal levels of femininity, but caricaturing either did not increase their femininity (Study 1) or increased it by a small amount (Study 2 composite continuum). These results suggest that it is surprisingly difficult to exaggerate facial sexual dimorphism beyond the levels found in strongly dimorphic natural faces.

### Why do our results differ from Zhao et al.'s (2011)?

Whereas we found increasing aftereffects across the full range of adaptor values for which sexual dimorphism increases, Zhao et al. (2011) found a decrease within this range. We see two plausible factors that could have contributed to their decrease. First, it could have been driven by the corresponding decrease in realism across the same stimulus values. Low realism could potentially contribute to the reduced aftereffects either by interfering with the perception of gender or by reducing any contribution of higher level face adaptation to the aftereffects. Second, it could have been driven by low-level contributions to the aftereffects. An important difference between our study and Zhao et al.'s is that we minimized the contribution of low-level adaptation (via size change), whereas they did not (except for one participant who showed very small aftereffects). The gender continuum produces a con-

tinuum of changes in various low-level properties like contour orientation and spatial frequency that differ between male and female faces. These properties are multichannel coded in V1. Thus aftereffects for these properties could have produced the evidence of multichannel coding.

Importantly, we can be confident that the difference between our results and Zhao et al.'s (2011) is not due to us not covering enough range. Multichannel coding predicts that aftereffects will peak, and then decline, as adaptors become more extreme. We found no evidence for any decline even though our adaptors spanned a wider range of perceived sexual dimorphism than those used by Zhao et al. (2011). The perceived sexual dimorphism of their adaptors did not exceed those of mixed-dimorphism natural faces, whereas our adaptors went up to (and in one case beyond) the level of high-dimorphism natural faces. Of course we cannot rule out the possibility of a peak for even more extreme adaptors.

It is an open question, though, whether it is possible to create such hypermale and hyperfemale faces. In both our studies the male averaged composite looked less masculine than natural male faces and caricaturing did not increase its masculinity beyond that of (mixed-dimorphism) natural faces. Caricaturing the new female composite in Study 2 did increase its femininity beyond high-dimorphism natural levels, but this was the only continuum out of those examined across our two studies where perceived sexual dimorphism increased beyond the level of natural faces. Also, the increase was relatively small (see +200% point in Figure 7) and, potentially, dimorphism for our +200% stimulus would not exceed some very feminine real world faces (i.e., the database of 196 faces from which we drew our natural faces would not necessarily include the most feminine females possible in the real world). Furthermore, our attempt to exaggerate perceived sexual dimorphism by caricaturing natural faces was not successful and resulted in decreased realism. It might be possible to preserve realism by using single identity composites made from multiple images of the same person (Jenkins, White, Van Montfort, & Burton, 2011) to reduce image differences (e.g., spatial frequency) between the parent faces and the androgynous composite (see Figure 5) (although this might also reduce masculinity of male faces). Overall, it appears difficult to increase perceived sexual dimorphism beyond the levels of highly sexually dimorphic natural faces using gender caricaturing.

### Challenges to the idea of opponent coding for gender

Our conclusion that face gender is opponent coded differs from that of another recent study conducted

using a very different adaptation paradigm (Storrs & Arnold, 2012). The authors asked participants to decide whether test faces were female, androgynous, or male, and measured category boundaries for each participant. Then, they examined how adaptation shifted those boundaries. A size change was used between adaptor and test images to minimize low-level, retinotopic adaptation. Opponent coding predicts renormalization, so that adaptation to a face at the boundary between the middle and upper categories should shift both upper and lower category boundaries in the same direction. In contrast, multichannel coding predicts repulsion away from the adapted point, so that only the lower category boundary should shift (away from the adaptor). The observed boundary shifts supported opponent coding for color and facial configuration, which are known to be opponent coded, thus validating their procedure. For face gender, the results were interpreted as supporting multichannel coding.

We suggest, however, that caution is needed in interpreting these results. First, face gender is perceived categorically, with a sharp boundary between male and female faces that would make androgyny a difficult and unnatural response category to use (Campanella, Chrysochoos, & Bruyer, 2001; Freeman, Rule, Adams, & Ambady, 2010). Consistent with this suggestion, Storrs and Arnold's (2012) participants appeared reluctant to categorize faces as androgynous (see their Figure 4b). Second, the crucial prediction differentiating multichannel from opponent coding is that there is no shift in the upper category boundary (between androgyny and female) after adapting to faces at that boundary. Because this is a negative result, it is vital to have good statistical power, and power was low, with only six participants. In fact there was a small, positive shift in this upper boundary, but no statistical test of significance was reported. The authors concluded that gender was multichannel coded based on finding that the lower boundary shifted significantly more than the upper boundary. Numerically, however, both category boundaries shifted in the same direction, which seems consistent with renormalization, i.e., opponent coding. They also found little change in the perceived gender of the adapting images, which is consistent with multichannel coding, but again this is a negative result and power was low.

### Challenges to the idea of opponent coding in general

Our aftereffect results suggest that structural attributes of faces related to gender are opponent coded, a conclusion that many have also drawn for identity (for a review see Rhodes & Leopold, 2011). Recently,

however, Ross, Deroche, and Palmeri (in press) have argued that certain data patterns traditionally taken as evidence for norm-based opponent coding models can be accounted for by exemplar-based models. Most relevant to the present study, they showed that exemplar-based (cf. multichannel?) models can simulate effects of adaptor position (extremity) on face identity aftereffects that have been interpreted as evidence for opponent coding. A limitation of their approach, however, is that they only used a very restricted range of adaptor levels (zero and  $-0.4$  identity strength, on a scale where undistorted faces have identity strength of unity), over which multichannel (cf. exemplar-based) and opponent (cf. two-pool norm-based) coding models make identical predictions (cf. Figure 1). Therefore, it is unsurprising that exemplar models could simulate the results. Exemplar models could also simulate repulsive identity aftereffects, but again this is not surprising because it is well accepted that the existence of such aftereffects, by itself, does not distinguish different models (see recent reviews by Rhodes & Leopold, 2011; Webster & MacLeod, 2011). More challenging is their finding that exemplar models with more than two dimensions could simulate larger identity aftereffects for opposite than nonopposite adapt-test pairs (matched on perceptual dissimilarity), a result that has been taken as strong evidence for opponent coding of identity (Rhodes & Jeffery, 2006). It remains to be seen how this result can be reconciled with evidence against multichannel coding of identity-related dimensions (Susilo et al., 2010).

### What is the origin of our gender aftereffects?

Our method minimized the contribution of low-level adaptation to our gender aftereffects. Their precise origin, however, remains an open question. It may include high-level face-selective coding mechanisms, and/or more generic (opponent) shape-coding mechanisms (e.g., Suzuki, 2005) that nevertheless contribute to the perception of face gender. It is also an open question whether our results derive from adaptation of some unitary “dimension” of gender (e.g., in face space). There are many physical sexually dimorphic properties in faces and it is not clear that they are ever integrated into a unitary visual gender dimension. If they are not, then it is possible that our aftereffects reflect adaptation of many distinct, sexually dimorphic properties.

Our face gender aftereffects seem likely to reflect adaptation of shape/structure, rather than color/brightness. Color/brightness information certainly differs between male and female faces (Russell, 2009; Van den Bergh & Frost, 1986), but variation in such information had little effect on perceived sexual

dimorphism in Study 1 (see also Tarr, Kersten, Cheng, & Rossion, 2001). Moreover, although it is possible that color adaptation contributed to the observed nonlinearities in the functions relating perceived sexual dimorphism and aftereffects, color adaptation cannot explain the overall pattern of aftereffects in Study 2 because color varied only between 0% and 100% levels, whereas aftereffects increased over the whole range. Finally, for color or brightness adaptation to produce the opposite gender aftereffects observed here (androgynous faces looking male after viewing female adaptors and female after viewing male adaptors), sexually dimorphic color/brightness differences would have to span the neutral (gray) points of color-opponent mechanisms. But they do not. Female faces are pinker than male faces, but male faces are not greenish, and female faces are lighter than male faces, but (Caucasian) male faces are not darker than midpoint of a black-white dimension. We suggest, therefore, that the aftereffects observed in Study 2, most likely reflect adaptation of structural attributes related to the perception of face gender.

A final question is whether our aftereffects are even visual in origin. Several recent studies have reported face gender aftereffects in the absence of face adaptors (Ghuman, McDaniel, & Martin, 2010; Javadi & Wee, 2012; but see Kloth, Schweinberger, & Kovacs, 2010). For example, adapting to male and female headless bodies can produce face gender aftereffects (Ghuman et al., 2010). These cross-modal aftereffects raise the possibility that face gender aftereffects might have a purely post-perceptual, semantic locus, affecting response decisions rather than visual coding. However, they could also reflect visual adaptation via top-down activation (and thus adaptation) of visual coding mechanisms. Imagined faces can certainly activate face-selective visual cortex (O’Craven & Kanwisher, 2000) and produce face aftereffects (Hills, Elward, & Lewis, 2010), consistent with such a top-down account. Irrespective of the source of cross-modal aftereffects, however, it seems implausible to suggest that the face aftereffects seen here, which are generated using *visual* adaptors, and which produce strong subjective changes in the appearance of faces, have no visual component.

### How can opponent coding of gender be reconciled with dissociable norms for males and females?

Male and female face norms can be independently adapted, indicating dissociability (Bestelmeyer et al., 2008; Jaquet & Rhodes, 2008). However, there is also partial transfer of adaptation between male and female faces, indicating some common coding (Jaquet & Rhodes, 2008). One proposal is that faces are



represented in a single face space containing sex-selective dimensions that apply primarily to one sex and also “common” (nonsex-selective) dimensions that apply to both male and female faces (Rhodes, Jaquet, et al., 2011). Common dimensions that are sexually dimorphic would contribute to perceived gender. Our results suggest that such dimensions are opponent coded, with the androgynous norm (neutral point) signaled by equal activation of the two opponent populations.

## Conclusion

So how is face gender coded? We have argued that previous evidence suggestive of multichannel coding is equivocal (Storrs & Arnold, 2012; Zhao et al., 2011) and/or reflects low-level adaptation (Zhao et al., 2011). Here, we found that face gender aftereffects increased with the perceived sexual dimorphism of adaptors, consistent with opponent coding of higher level, shape-coding dimensions related to the perception of face gender. These results are consistent with other evidence for opponent coding of structural properties of faces (for reviews see Rhodes & Leopold, 2011; Webster & MacLeod, 2011).

*Keywords:* face perception, face gender aftereffects, norm-based opponent coding, multichannel coding, caricaturing

## Acknowledgments

This research was supported by the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE110001021), an ARC Professorial Fellowship to Rhodes (DP0877379), an ARC Discovery Outstanding Researcher Award to Rhodes (DP130102300), and an ARC Queen Elizabeth II Fellowship to McKone (DP0984558). The data from Study 2 were reported in an honors thesis submitted to the University of Western Australia in 2012 by Stephen Pond. We are especially grateful to Zhao and his colleagues for their cooperation in providing us with the stimuli from their paper. We thank Mike Webster for helpful discussions about adaptation and coding mechanisms. Ethical approval was granted by the Human Research Ethics Committee of the University of Western Australia.

Commercial relationships: none.

Corresponding author: Gillian Rhodes.

Email: gillian.rhodes@uwa.edu.au.

Address: School of Psychology, University of Western Australia, Crawley, WA, Australia.

## Footnotes

<sup>1</sup>There are also skin color differences between male and female faces (e.g., Van den Berghe & Frost, 1986), but adaptation of these differences would not produce a multichannel coding pattern of aftereffects because color is opponent coded (Webster & Leonard, 2008).

<sup>2</sup>We use *sexual dimorphism* to refer to quantitatively varying differences between male and female faces (masculinity in male faces, femininity in female faces).

<sup>3</sup>More female than male faces were available in the database. There is no reason to expect that using 26 males faces (rather than 53 for females) would not produce a good male composite: As few as 16 faces produce a fairly stable composite that does not change much with additional faces (e.g., Langlois & Roggman, 1990). Also, our male and female faces appear equally good in image quality (see  $\pm 100$  ANU faces in Figure 2).

<sup>4</sup>ANOVAs restricted to Caucasian participants (so that faces were their own race) yielded the same significant effects as those reported below.

## References

- Bestelmeyer, P. E. G., Jones, B. C., DeBruine, L. M., Little, A. C., Perrett, D. I., Schneider, A., ... Conway, C. A. (2008). Sex-contingent face aftereffects depend on perceptual category rather than structural encoding. *Cognition*, *107*, 353–365.
- Blakemore, C., & Sutton, P. (1969). Size adaptation: A new aftereffect. *Science*, *166*(3902), 245–247.
- Burton, N., Jeffery, L., Skinner, A. L., Benton, C. P., & Rhodes, G. (2013). Nine year-old children use norm-based coding to visually represent facial expression. *Journal of Experimental Psychology: Human Perception & Performance*, *39*, 1261–1269.
- Calder, A. J., Jenkins, R., Cassel, A., & Clifford, C. W. G. (2008). Visual representation of eye gaze is coded by a nonopponent multichannel system. *Journal of Experimental Psychology: General*, *137*(2), 244–261.
- Calder, A. J., Rhodes, G., Johnson, M. H., & Haxby, J. V. (Eds.). (2011). *The Oxford handbook of face perception*. Oxford, UK: Oxford University Press.
- Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual Cognition*, *8*(2), 237–262.
- Clifford, C. W. G. (2002). Perceptual adaptation: Motion parallels orientation. *Trends in Cognitive Sciences*, *6*(3), 136–143.

- Clifford, C. W. G., Wenderoth, P., & Spehar, B. (2000). A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society of London, Series B*, 267, 1705–1710.
- Farkas, L. G. (1988). Age- and sex-related changes in facial proportions. In L. G. Farkas & I. R. Munro (Eds.), *Anthropometric proportions in medicine* (pp. 29–56). Springfield, IL: Charles C. Thomas.
- Fiorentini, C., Gray, L., Jeffery, L., Rhodes, G., & Pellicano, E. (2012). Reduced face identity aftereffects in relatives of children with autism. *Neuropsychologia*, 50, 2926–2932.
- Freeman, J. B., Rule, N. O., Adams, R. B., Jr., & Ambady, N. (2010). The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices. *Cerebral Cortex*, 20(6), 1314–1322.
- Ghuman, A. S., McDaniel, J. R., & Martin, A. (2010). Face adaptation without a face. *Current Biology*, 20(1), 32–36.
- Hills, P. J., Elward, R. L., & Lewis, M. B. (2010). Cross-modal face identity aftereffects and their relation to priming. *Journal of Experimental Psychology: Human Perception & Performance*, 36, 876–891.
- Jaquet, E., & Rhodes, G. (2008). Face aftereffects indicate dissociable, but not distinct, coding of male and female faces. *Journal of Experimental Psychology: Human Perception & Performance*, 34, 101–112.
- Javadi, A. H., & Wee, N. (2012). Cross-category adaptation: Objects produce gender adaptation in the perception of faces. *PLoS One*, 7, e46079.
- Jeffery, L., McKone, E., Haynes, R., Firth, E., Pellicano, E., & Rhodes, G. (2010). Four-to-six-year-old children use norm-based coding in face-space. *Journal of Vision*, 10(5):18, 1–19, <http://www.journalofvision.org/content/10/5/18>, doi:10.1167/10.5.18. [PubMed] [Article]
- Jeffery, L., Read, A., & Rhodes, G. (2013). Four year-olds use norm-based coding for face identity. *Cognition*, 127, 258–263.
- Jeffery, L., Rhodes, G., McKone, E., Pellicano, E., Crookes, K., & Taylor, E. (2011). Distinguishing norm-based from exemplar-based coding of identity in children: Evidence from face identity aftereffects. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1824–1840.
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121, 313–323.
- Kloth, N., & Schweinberger, S. R. (2010). Electrophysiological correlates of eye gaze adaptation. *Journal of Vision*, 10(12):17, 1–13, <http://www.journalofvision.org/content/10/12/17>, doi:10.1167/10.12.17. [PubMed] [Article]
- Kloth, N., Schweinberger, S. R., & Kovacs, G. (2010). Neural correlates of generic versus gender-specific face adaptation. *Journal of Cognitive Neuroscience*, 22(10), 2345–2356.
- Koehler, N., Simmons, L. W., Rhodes, G., & Peters, M. (2004). The relationship between sexual dimorphism in human faces and fluctuating asymmetry. *Proceedings of the Royal Society of London, Series B*, 271(Suppl.), S233–S236. doi:10.1098/rsbl.2003.0146
- Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, 1, 115–121.
- Lawson, R. P., Clifford, C. W. G., & Calder, A. J. (2011). A real head turner: Horizontal and vertical head directions are multichannel coded. *Journal of Vision*, 11(9):17, 1–17, <http://www.journalofvision.org/content/11/9/17>, doi:10.1167/11.9.17. [PubMed] [Article]
- O’Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, 12, 1013–1023.
- Rhodes, G., Jaquet, E., Jeffery, L., Evangelista, E., Keane, J., & Calder, A. J. (2011). Sex-specific norms code face identity. *Journal of Vision*, 11(1):1, 1–11, <http://www.journalofvision.org/content/11/1/1>, doi:10.1167/11.1.1. [PubMed] [Article]
- Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, 46(18), 2977–2987.
- Rhodes, G., & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 263–286). Oxford, UK: Oxford University Press.
- Rhodes, G., Simmons, L., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evolution & Human Behavior*, 26, 186–201.
- Robbins, R., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: Adapter position effects and neural models. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3), 570–592.
- Ross, D. A., Deroche, M., & Palmeri, T. J. (in press). Not just the norm: Exemplar-based models also

- predict face aftereffects. *Psychonomic Bulletin & Review*.
- Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, *38*, 1211–1219.
- Skinner, A. L., & Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychological Science*, *21*(9), 1248–1253.
- Storrs, K. R., & Arnold, D. H. (2012). Not all face aftereffects are equal. *Vision Research*, *64*(1), 7–16.
- Susilo, T., McKone, E., & Edwards, M. (2010). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research*, *50*, 300–314.
- Suzuki, S. (2005). High-level pattern coding revealed by brief shape aftereffects. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and aftereffects in high level vision* (pp. 135–172). Oxford, UK: Oxford University Press.
- Tarr, M. J., Kersten, D., Cheng, Y., & Rossion, B. (2001). It's Pat! Sexing faces using only red and green. *Journal of Vision*, *1*(3):337, <http://www.journalofvision.org/content/1/3/337>, doi:10.1167/1.3.337. [Abstract]
- Van den Berghe, P. L., & Frost, P. (1986). Skin color preference, sexual dimorphism and sexual selection: A case of gene culture co-evolution. *Ethnic & Racial Studies*, *9*, 87–113.
- Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(6982), 557–561.
- Webster, M. A., & Leonard, D. (2008). Adaptation and perceptual norms in color vision. *Journal of the Optical Society of America, A*, *25*, 2817–2825.
- Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society*, *366*, 1702–1725.
- Zhao, C., Serriès, P., Hancock, P. J. B., & Bednar, J. A. (2011). Similar neural adaptation mechanisms underlying face gender and tilt aftereffects. *Vision Research*, *51*(18), 2021–2030.