Similar neural adaptation mechanisms underlying face gender and tilt aftereffects ☆

Chen Zhao [a], Peggy Sériès [a], Peter J.B. Hancock [b], James A. Bednar [a,*]

[a] Institute for Adaptive and Neural Computation, School of Informatics, The University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK
[b] Department of Psychology, University of Stirling, Stirling FK9 4LA, UK

ARTICLE INFO

ABSTRACT

Visual aftereffects have been found for a wide variety of stimuli, ranging from oriented lines to human faces, but previous results suggested that face aftereffects were qualitatively different from orientation (tilt) aftereffects. Using computational models, we predicted that these differences were due to the limited range of faces used in previous studies. Here we report psychophysical results verifying this prediction. We used the same paradigm to test tilt aftereffects (TAE) and face gender aftereffects (FAE) and found that they exhibited qualitatively similar aftereffect curves, when a sufficiently large range of test faces was used. Overall, the results suggest that similar adaptation mechanisms may underlie both high-level and low-level visual processing.

1. Introduction and background

The neural mechanisms for low-level processing of images have been studied in detail using animal models, focusing on early visual areas such as the primary visual cortex (V1). However, much less is known about the neural basis of high-level perception, particularly in humans. An important issue is whether and how we can apply lessons learned from low-level studies, such as how neurons in V1 respond to oriented line segments, to understanding high-level perception, such as human processing of faces.

Visual aftereffects, i.e., systematic changes in visual perception based on recent experience, are thought to reflect the underlying neural organization, and can thus potentially indicate similarities and differences in mechanisms (Thompson & Burr, 2009). Aftereffects have been found for a huge variety of stimuli, including both the tilt aftereffect (TAE) for oriented lines (Gibson & Radner, 1937; Mitchell & Muir, 1976) and the face aftereffect (FAE) for human faces (Leopold et al., 2001; Rhodes & Jeffery, 2006; Webster et al., 2004). Comparing these aftereffects may give insight into the similarities and differences between low-level and high-level visual processing.

As the orientation difference between test and adapting stimuli is varied, the TAE has a characteristic S-shaped curve for orienta-

tion perception change that is typical of low-level aftereffects: starting from no net effect when the adaptation stimulus has the same orientation as a test stimulus, increasing quickly in magnitude for slightly different stimuli, and eventually decreasing for larger differences in orientation. As an example, the result of a psychophysical experiment done by Mitchell and Muir (1976) is shown as a black curve in Fig. 1b. Each point in the figure stands for a condition where the subject was first tested using a vertical stimulus, then adapted to the indicated orientation, and was then tested again with the vertical stimulus. The aftereffect was then defined as the difference in perceived orientation of the vertical stimulus before and after adaptation. The amount of aftereffect reaches a peak around ±10° orientation difference, and soon decreases to zero or even changes sign thereafter (depending on the subject and condition).

Numerous computational and theoretical models (e.g. Bednar & Miikkulainen, 2000; Sériès, Stocker, & Simoncelli, 2009) have been proposed to explain the TAE curve and similar low-level effects, but nearly all are based on three main principles: (1) visual cortex neurons respond only to a limited range of stimulus values (e.g., V1 neurons have a limited tuning bandwidth for orientation), (2) neurons activated by a stimulus adapt, reducing their responsiveness by some means (whether by increased inhibition or by depletion of some resource; reviewed in Kohn (2007)), and (3) the perception at any instant is determined by the activity pattern across the population of neurons, such that the perceived value differs when some of the neurons are less responsive (see Sériès, Stocker, & Simoncelli, 2009).

Fig. 1a shows a concrete implementation of these principles using the motion-aftereffect model from Sériès, Stocker, and Simoncelli (2009) fit to the published TAE data, illustrating how

---
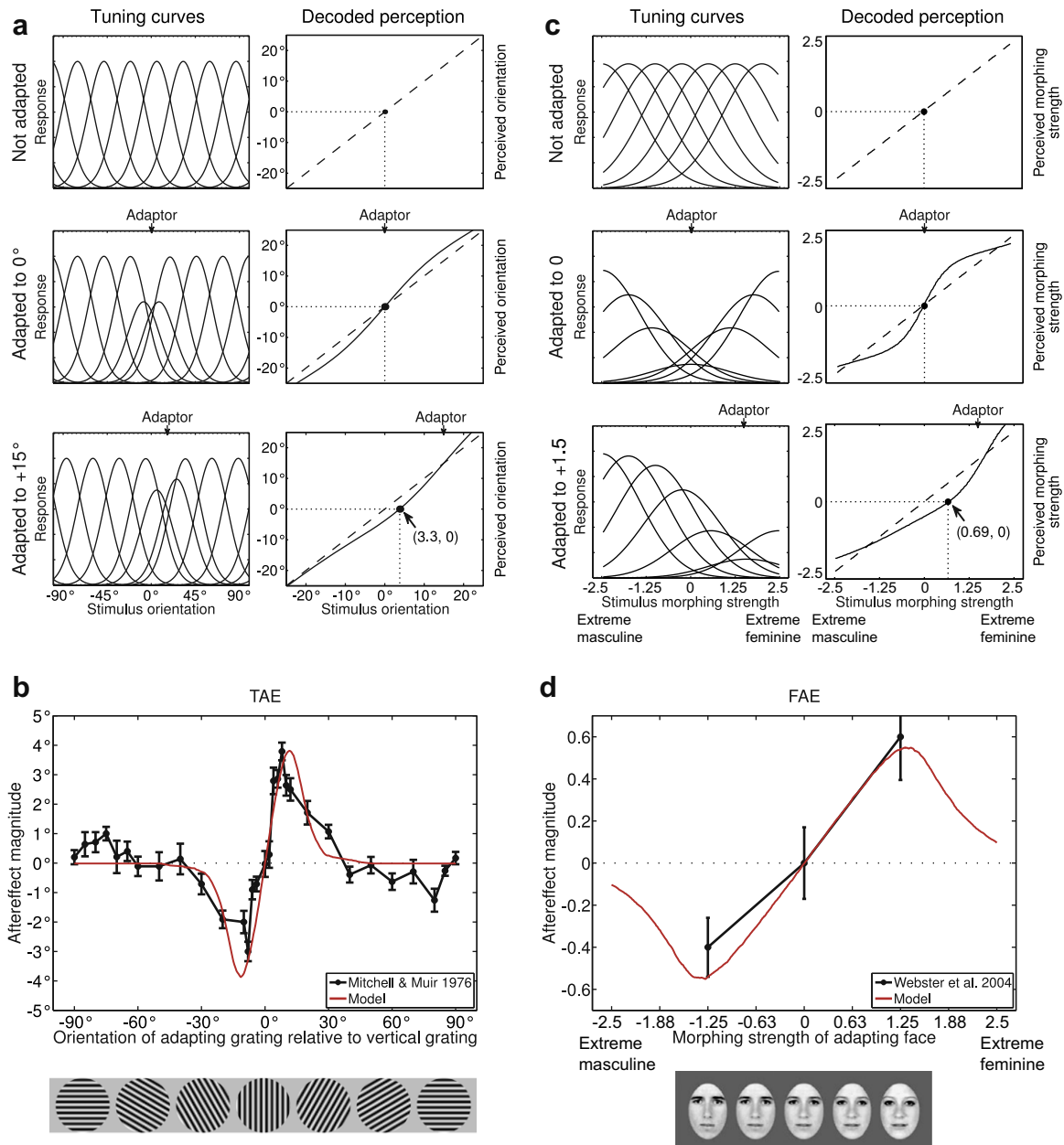
Fig. 1. General model for aftereffects based on adaptation of neurons tuned to specific feature values. (a, left) Gaussian tuning curves for nine neurons in a computational model of the TAE (adapted from Seriès, Stocker, and Simoncelli (2009)). Repeated presentation of one orientation (marked "Adaptor") reduces the responsiveness of neurons whose tuning curves overlap the adaptor (i.e., neurons that responded to the adaptor). The asymmetry for the 15° case is due to having a finite number of example neurons, such that any particular orientation like 15° does not fall midway between two of them. (a, right) Perceived orientation for each possible test pattern before (dashed line) and after (solid line) adaptation, calculated using a maximum-likelihood (ML) method. Filled circles show the new perceived "vertical" (0°). E.g., after adaptation to +15°, +3.3° is now decoded as 0°, yielding an aftereffect of +3.3°. (b) The red TAE curve summarizes these shifts in preferred orientation for each adaptor, yielding a prediction of the tilt aftereffect strength comparable to psychophysical results (e.g. from participant DEM of Mitchell and Muir (1976), as shown here). (c, left) The same model as in (a) and (b) applied to a facial gender aftereffect (FAE). (c, right) For each adaptation condition, the population response that is ML-decoded as androgynous (0) is shown with a filled circle. (d) The model FAE curve matches the sparse existing data (Webster et al., 2004), but strongly predicts that FAE values will decrease for larger morphing strength magnitudes (faces that are more feminine or more masculine than the examples shown).

neural adaptation leads to aftereffects with a realistic S-shaped curve (red curve in Fig. 1b). This simple model consists of a population of neurons, each with a tuning curve most selective for a particular orientation. When one orientation (the adaptor) is presented repeatedly, neurons responding to that orientation reduce their responsiveness (see the two examples in Fig. 1a left). The model then predicts how the perception of subsequent test patterns will differ as a result of these changes in responsiveness (see the two examples in Fig. 1a right). See Appendix A for the detailed methods for the modeling; other TAE models will differ in details, but generally follow the three principles above.

If high-level perception also follows these three principles, then one would expect face aftereffects to follow an S-shaped curve as well. To illustrate this idea, we again modified the Seriès, Stocker, and Simoncelli (2009) model, this time to simulate a facial gender aftereffect (FAE) by using a non-cyclic input dimension, broadly tuned neurons, and adaptation patterns ranging from masculine (morphing strength <0) to feminine (>0; Fig. 1c; see Appendix A for more details). As Fig. 1d illustrates, the model does predict an S-shaped curve for the FAE, with aftereffect magnitude decreasing for sufficiently gendered faces. Note that the tuning bandwidth for orientation, as a fraction of the stimulus range plotted (Fig. 1c), is

much narrower than for the faces (Fig. 1d), leading to a wider S-shaped curve in the FAE case. But because the apparent bandwidth for the FAE case depends on this stimulus range, it is not meaningful to compare this particular arbitrary portion of a non-cyclic dimension with the complete range covered for a cyclic dimension like orientation. For instance, if the face stimulus range were chosen to be much wider than depicted in Fig. 1d (i.e., beyond ±2.5 on the scale), the bandwidth for faces would look similar to that for orientation. The actual range of face-related dimensions is not known, and thus here we will focus only on the curve shape and not the numerical bandwidth.

Seemingly contrary to the model's clear predictions, existing psychophysical studies of human face gender, identity and distortion aftereffects (Jeffery et al., 2010; Leopold et al., 2001; Little, Debruine, & Jones, 2005; Webster et al., 2004) and physiological studies of the FAE in monkeys (Leopold, Bondar, & Giese, 2006) have instead found aftereffects that monotonically increase in magnitude with difference away from an average face. These results have led to an emerging consensus that faces are processed with "norm-based encoding" (Leopold et al., 2001; Leopold, Bondar, & Giese, 2006; Loffler et al., 2005; Rhodes & Jeffery, 2006; Robbins, McKone, & Edwards, 2007), with a neural representation based on differences from an average face, rather than following the three principles from V1 above. A very recent paper did find an S-shaped curve for face viewpoint (Chen et al., 2010), but as viewpoint is a general property shared across all types of physical stimuli, processing for higher level properties of faces like gender remains unclear.

In this paper, we describe a new method capable of measuring both the TAE and gender FAE under identical conditions, so that the aftereffects can be compared consistently. We then tested it on the TAE and on the gender FAE using a larger range of faces than tested previously, as suggested by the model (Fig. 1c and d), and find that the FAE for facial gender has precisely the "S" shape predicted from the models devised for V1. The results suggest that face processing uses adaptation mechanisms similar to orientation processing in V1, and cast doubt on some of the assumptions of the norm-based theory of face encoding. In the discussion, we explain our neural adaptation theory in more detail, and compared it with "norm-based" and "exemplar-based" approaches. We also compare our theory and results with recent studies on face distortion aftereffects and face viewpoint aftereffects. Details of our modeling approach can be found in Appendix A, but are not necessary for evaluating the experimental results.

2. Material and methods

2.1. Participants and apparatus

Two adult male participants were tested for the TAE—the experimenters CZ and JB. Six adult male participants—the experimenters CZ and JB plus naïve participants JA, CB, LW and ZK—were tested for the FAE. Their ages range from 23 to 39 years old. JA, CB and JB are Caucasian, with a lifetime of experience with Caucasian faces such as those used in the experimental stimuli. CZ, ZK, and LW are Chinese. CZ and ZK had lived in a Caucasian face environment for more than a year, as had LW for 4 months.

Stimuli in both TAE and FAE experiments were presented on a 17-in. 85 Hz CRT monitor viewed at roughly $16° \times 19°$ from roughly 45 cm.

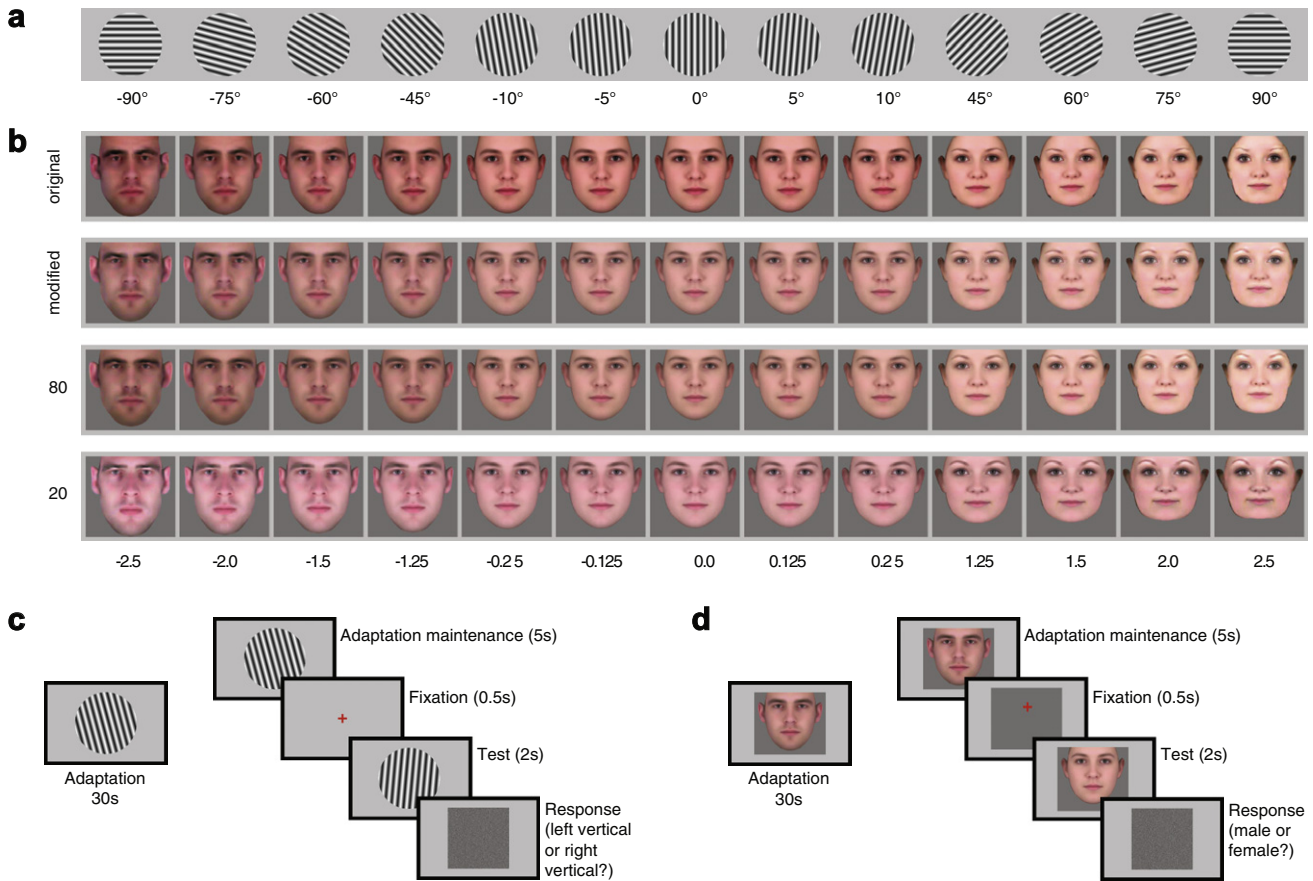### 2.2. Stimuli

The stimuli used in TAE experiments were sine gratings with orientations ranging from $-90°$ to $+90°$, at an approximate spatial frequency of 0.6 cycles per degree (see Fig. 2a for an example).

The stimuli used in FAE experiments were 501 composite faces generated using the methods described by Tiddeman, Burt, and Perrett (2001) from 150 young adult male and 150 young adult female Caucasian photographic faces collected by Ian Penton-Voak at Stirling University (Penton-Voak et al., 2006). For each photograph, Penton-Voak et al. manually marked key locations (179 points) around the main features and the outline of the face. Photographic quality average male and average female faces (denoted as -0.5 and $+0.5$ morphing strength) were created by triangulating all key locations, computing average shape vectors from these locations, warping all faces into that shape, and computing average facial images. We computed an initial 101 step morph sequence between the average male and female faces, the midpoint being the overall population average, denoted by a morphing strength of zero. We then morphed the average male face linearly away from the average female, giving an additional 150 images out to a morphing strength of -2.5, and the average female face in the opposite direction, to give 150 images out to a morphing strength of $+2.5$. Overall, therefore, we have a sequence of 501 images from -2.5 to $+2.5$, where one unit (1.0 morphing strength) is the average difference between male and female faces. A sample of the generated synthesized face continuum is shown in the top row in Fig. 2b. It can be seen that even the most extreme faces used here (e.g., -2.5 and $+2.5$) have very strong masculine or feminine features, but are still recognizable as human faces. The face stimuli used a darker gray background than the gratings, to provide high-contrast edges that make the faces salient. Further details about the face generation method were reported by Tiddeman, Burt, and Perrett (2001).

As controls, three additional different versions of the face stimuli were used for different participants, but all results are discussed together in Section 3 because no qualitative or quantitative differences in results were found between participants or stimulus sets. The first version of the face morph continuum is shown in Fig. 2b "original", used for participants CZ and JA. This face set, taken directly from the output of the face generator, was slight left-right asymmetric in face shape and the overall skin tone between male and female images were slightly different. To see if the results were affected by these artifacts, the "modified" face set in Fig. 2b was generated. For the "modified" face set, the key locations for the face morphing were averaged across the vertical midline, enforcing symmetry, and the overall trend of color changing with morphing strength was reduced by manually adjusting the color balance of the two male and female average images to more closely match that of the population average. As mentioned, the results from participants using the modified dataset were similar to the original, indicating that these features of the original dataset were not necessary for the results.

For both the original and modified datasets, participants were tested with faces drawn from the same morphing continuum as the adaptation stimuli. With faces generated along a smooth continuum, many of the low-level features will vary systematically along with gender, such as the shape of the eyes, face outline, and mouth, the brightness of skin textures, etc. To test whether low-level adaptation to the individual features could explain the observed FAE, the original face photographs were split into one set of 80% (120 male and 120 female) and another set with the remaining 20% (30 male and 30 female). A face continuum was then generated for each of those separate datasets by averaging and morphing as described above; the generated continuums are shown in Fig. 2b rows "80" and "20". The reason for choosing an 80%/20% split (rather than e.g., 50%/50%) is to make these two sets clearly different in identity, to reduce the interference of low-level factors. (An even more biased split (e.g., 95%/5%) would further increase the difference, but would leave too few faces in the smaller dataset to give a smooth face continuum.) Participants LW and ZK

Fig. 2. Face stimuli example and psychophysical aftereffect experiments paradigm. (a) Example stimuli for the TAE experiments, covering all orientations. (b) Example face stimuli used in each FAE experiment, covering a wide range of male and female faces. Each row shows 21 examples out of the 501 faces making up each continuum. The "original" face continuum (top row) was used for participants CZ and JA. The "modified" continuum (see text) was used for CB. The row marked "80" was generated like the modified continuum, but using only 80% of the faces. The remaining 20% were used for the continuum in the last row. Participants LW and ZK were adapted to the 20 continuum and tested on the 80. (c and d) Two-alternative forced-choice (2AFC) paradigm, identical for TAE and FAE except for the stimuli and the response criterion. Participants first adapted to a stimulus, then were tested on random ambiguous test patterns near their perceptual boundary (vertical or androgynous) so that an updated perceptual boundary could be estimated using a psychometric function on their responses.

were then adapted to faces from the 20% set, but tested on faces from the 80% set.

In addition to these three different face continuum datasets, a fourth condition was used to further investigate contributions from low-level effects, by varying the size between adaptation and test phases. The stimuli in the fourth FAE experiment were the same as the "modified" face set in Fig. 2b, except that the test faces were at 50% of the width and 50% of the height as the adaptation face. I.e., the overall size of the test face was 25% of the area covered by the adaptation face, disrupting many of the low-level features but keeping the face appearance and identity intact. The test and adaptation faces were aligned along the center point between two eyes, where the fixation point was placed. Otherwise, the experimental procedure was identical to the other conditions, as described below.

2.3. Experimental Procedure

A common method for measuring the TAE in previous experiments was to ask the participant to adjust a reference line at a distant location until it appeared parallel to a test line at the adapted location (Mitchell & Muir, 1976). Procedures of this type are not practical for measuring the FAE, because the FAE shows significant transfer across retinotopic locations (Afraz & Cavanagh, 2008.) Instead, we measured both the TAE and FAE by asking participants to make a two-alternative forced choice (2AFC) along a natural

perceptual boundary—either between lines tilted slightly left or right from the vertical, or between male and female faces. Aftereffects were defined as shifts in this subjective boundary after adaptation. The adaptation and test stimuli used in each experiment were drawn from the range of possible orientations (TAE, Fig. 2a) or the face continuums discussed above (FAE, Fig. 2b). Each test stimulus consisted of a randomly selected ambiguous test pattern chosen from near the perceptual boundary (vertical or androgynous) so that the test would be sensitive to changes in the boundary. The procedure is illustrated in Fig. 2c and d. Apart from the stimuli and the requested perceptual judgment, the procedures for the TAE and FAE tests were identical.

For each participant, the existing perceptual boundary was first measured in a baseline block of trials, and then a series of adaptation blocks measured the effects of adapting to different stimuli. For both baseline and adaptation blocks in both FAE and TAE experiments, the method of constant stimuli was implemented, where the test orientation or morphing strength was chosen randomly in each trial. Randomization prevents the participant from predicting the next stimulus, and therefore reduces errors of habituation and expectation. Baseline blocks were the same as adaptation except they had no adaptation and maintenance periods. In baseline blocks, test stimuli were chosen uniformly from a range of $-4°$ to $+4°$ around true vertical (orientation 0) or -0.3 to 0.3 morphing strengths around androgynous (morphing strength 0).

Once the unadapted (baseline) perceptual boundary was determined (see details in Appendix B), test stimuli for adaptation blocks were chosen from a range of −4° to +4° or −0.6 to 0.6 morphing strengths around the baseline boundary for that participant. For TAE experiments, stimulus ranges used in baseline and adaptation blocks were always identical. For FAE experiments, each participant had a different measured baseline ("androgynous") category boundary. To best suit the specific baseline for that participant, an appropriate test range was chosen individually—some were tested in the range of −0.1 to 0.6, others from −0.6 to 0.1. Of the six FAE participants, ZK, LW and JB used the same stimulus range for baseline and adaptation, while three others used different ranges. Note that in some cases, it has previously been shown that changing the range in this way could induce a shift in the point of subjective equality (PSE), i.e., the gender boundary we are testing (see Poulton, 1974; Laming, 1997). Such range effects could potentially bias the boundary points leftwards or rightwards (see crossing data points in Fig. 3b) in those participants, but they would not be expected to alter the overall shapes of the aftereffect curves, which are the focus of this study.

Apart from the narrow range around the boundary, two additional stimuli were added to each adaptation test block, significantly different from the perceptual boundary. For the TAE, they were −45° and +45°; for the FAE, they were −1.5 and 1.5 morphing strengths. These two stimuli were added to provide a brief respite and reassurance for the participants during these difficult experiments, being clearly recognizable as left or right (TAE tests) or male or female (FAE tests).

At the start of an adaptation block, the TAE or FAE adaptation stimulus was shown for 30 s (adaptation period). Then for each trial, adaptation was topped up for 5 s (maintenance period), followed by a 0.5-s fixation mark, a 2-s test stimulus, and then a noise pattern as a signal that a response was required. Participants were allowed to view the adaptation pattern freely during the adaptation period, and were instructed to move their eyes around to avoid afterimages and other distracting low-level effects. Once the noise pattern appeared, the participant indicated whether the test grating was left or right of the vertical (Fig. 2c), or whether the test face was male or female (Fig. 2d). Before an FAE block, the participants were instructed to judge the face holistically and use first impression only.

Each TAE or FAE adaptation block lasted about from 15 to 25 min, depending on the response time of the participant. Blocks were limited to one per day, to reduce transfer of adaptation across blocks.

3. Results

As seen in Fig. 3a, the TAE results are similar to those from classical TAE experiments (Mitchell & Muir, 1976), which suggests that this paradigm is comparable to earlier methods, while allowing testing with any type of stimulus.

As mentioned above, the specific face stimuli for each participant differed, but the FAE results are presented as one group because the FAE for every participant matched the TAE-like S-shaped curve predicted from the model (Fig. 3b). The effects were similar between participants CZ and JA using the "original" face set and participant CB using the "modified" face set. Similar curves were also found for ZK and LW, indicating that the results hold even for perceptually different adaptation stimuli (compare the −2.5 and +2.5 faces for the 80% and 20% datasets). These results presumably reflect the fact that the average faces (morphing strength 0) are similar for all of the datasets, as would be expected for any averages of large numbers of faces drawn from the same distribution. Finally, even when the stimulus size differed by a factor of four in area between adaptation and test (participant JB), a similar S-shaped curve was observed, though the aftereffect strength was lower in this case.

Fig. 3c shows that the presence and location of the peak and valley in the S-shaped curve were highly consistent across individuals, with aftereffects significantly stronger for morphing strengths ±1.5 than for ±2.5 (one-tailed paired Student t-test; $p < 0.0126$). Put together, all these results indicate that FAE magnitude decreases after reaching a maximum value in either direction of difference from the mean face. The control experiments also demonstrated that this aftereffect curve is qualitatively invariant to identity and size difference, as would be expected for high-level effects.

To evaluate how consistently the participants were able to perform the task, so that problems like fatigue would be evident, we measured average slopes of the fitted psychometric curves for each experimental block for all participants (2 for TAE and 6 for FAE). The slope of a psychometric curve is a measure of the participant's discriminability, i.e., ability to judge between two categories of stimuli. Fig. 4 shows that the discriminability for each participant has a small standard error of measurement (S.E.M.; plotted as error bars), indicating that it stayed largely constant over the course of the experiments, and thus was not seriously affected by fatigue or similar issues.

4. Discussion

Our experimental results for the FAE were fully consistent with the predictions of our model devised for low-level aftereffects like the TAE. To place these results in the context of other theories, let us consider previous ideas of how faces might be represented neurally in face space. According to face space theory, a particular face is a point in an underlying multidimensional space, where each dimension is a feature or attribute of the face (such as gender), and the center of the space along all such dimensions is the mean face (Valentine, 1991). Given this idea of face space, there has long been a debate about two main types of neural representations of the space: norm-based and exemplar-based. Norm-based theory proposes that the mean face has a special norm status, and that the neural representations of other faces are based on deviations from this norm (see Rhodes, Brennan, & Carey (1987) for the original work). More specifically, norm-based models typically employ two opponent pools of neurons selective for opposite stimuli in a dimension, with different attribute values represented as different balances between activations of these two pools (e.g., Rhodes & Jeffery, 2006). Multichannel exemplar-based theories instead employ a wide variety of neurons with different tuning curves, each selective for different faces in a dimension, and suggest that individual faces are coded by the neurons responding most strongly to that specific attribute value (Valentine, 1991; Valentine & Endo, 1992). See Webster and MacLeod (2011) for a review of these two coding methods and face adaptation.

Like our TAE model, our proposed FAE model is an example of a multichannel exemplar-based approach, albeit with broader tuning than typical of low-level multichannel models. The important feature of our model is not the tuning width, but the bell shape of the tuning curve that makes each neuron respond best to a specific range of stimuli. I.e., our results support the idea that face-selective neurons have tuning that is limited in extent, and that the neural population has neurons with preferences distributed across the range of face shapes, not just forming two separate pools flanking the face norm (as in norm-based opponent coding theories). In our model, neurons preferring each stimulus have equivalent properties, giving no special status to a "norm" face in the neural representation. Instead, the importance of a norm lies in its role of perceptual boundary in a specific perceptual task,
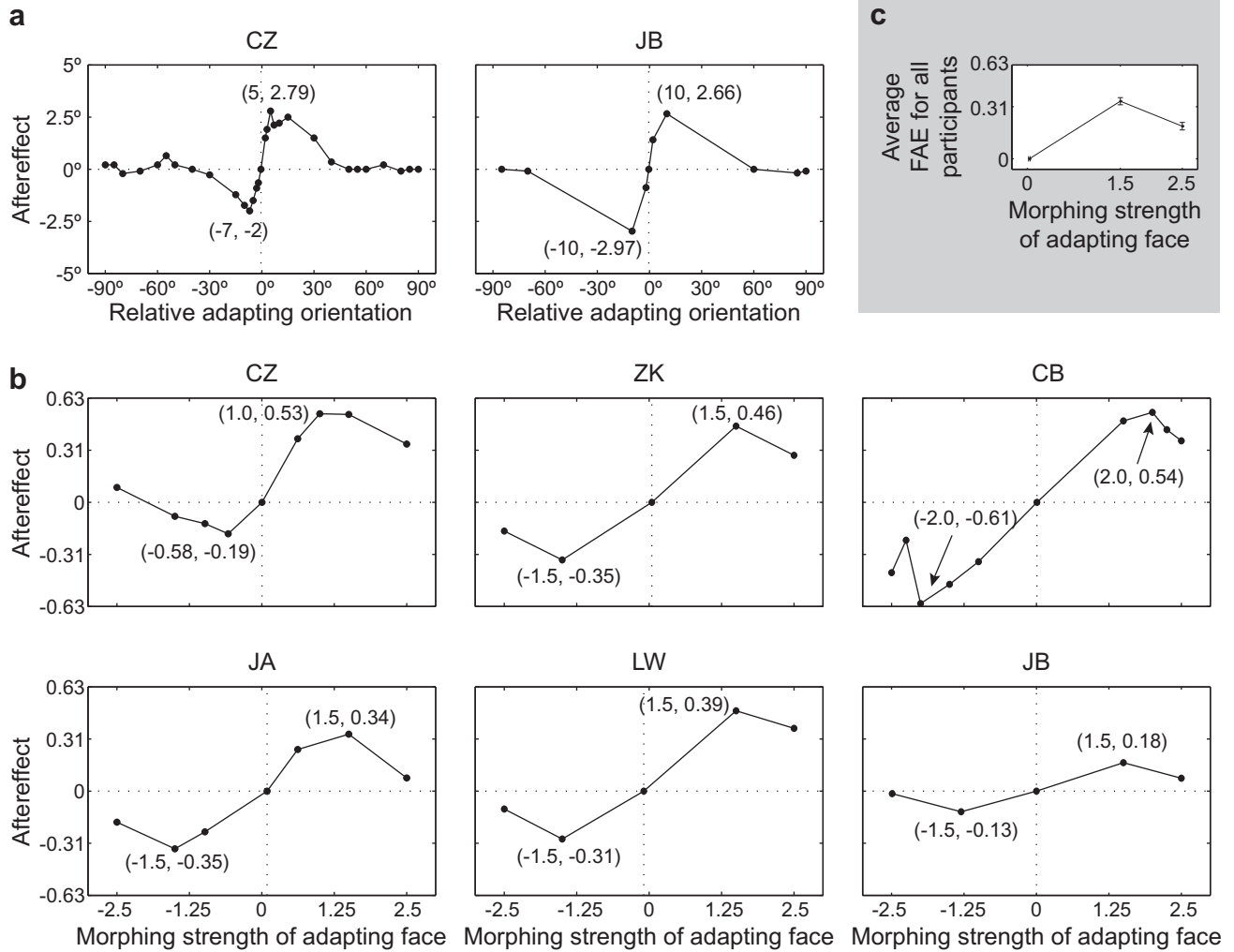
Fig. 3. Results of TAE and FAE experiments. (a and b) Aftereffects measured as shifts in the perceptual boundary after adaptation. TAE results for the two participants tested (a) are similar to previously published TAE results (e.g., Mitchell & Muir, (1976); Fig. 1b), verifying that the new paradigm tests similar mechanisms. FAE results for the six participants tested (b) suggested broad tuning, as expected, but were otherwise similar to the TAE results. In all cases the FAE magnitude at extreme morphing values decreased after reaching a maximum, as predicted from the models but contrary to previous reports. (c) The magnitude of the FAE averaged across all six participants and across both male and female test faces (for a total of 12 measurements), with error bars indicating the standard error of measurement. The FAE magnitude for morphing strengths with an absolute value of 2.5 is significantly lower than at an absolute value of 1.5 (one-tailed paired Student t-test; $p < 0.0126$), indicating that the "S" shape of the aftereffect was reliable across participants. The horizontal error bar at aftereffect zero shows the variation of the participant's perceptual boundary at baseline (i.e., no adaptation). It is too small to see clearly, indicating near-identical perceptual boundaries across participants.
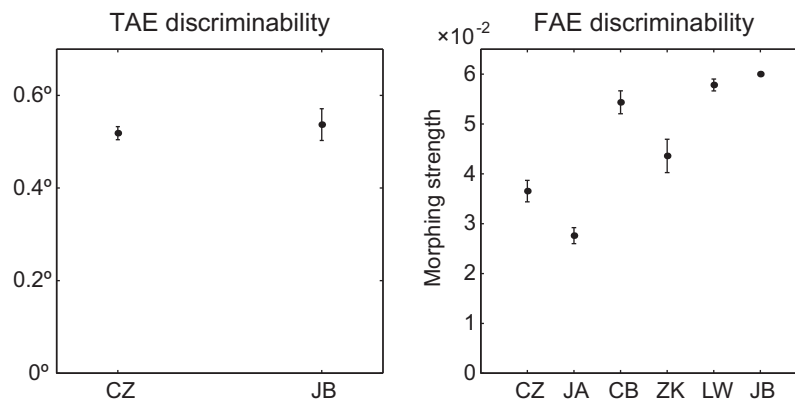
Fig. 4. Discriminability for TAE and FAE experiments. Participants' discriminability (slopes of the psychometric curves) for the TAE and FAE stimuli. Data points show each participant's average discriminability, with the standard error of measurement (S.E.M.) as error bars. These results show that the discriminability for each participant was largely constant in the experiments, and was not seriously affected by fatigue or similar issues.

whether for gender, distortion, identity, or other facial attributes. That is, the norm's importance is in decoding the perceptual boundary (before or after adaptation), rather than in building the underlying neural representation.

Many studies have instead argued that a norm-based representation is necessary to explain high-level face aftereffects. For example, Jeffery et al. (2010) recently argued that 4–6-year-old children and adults both exhibit norm-based face coding for face distortion, because they showed larger aftereffects for a large distortion value than for a smaller one. However, they tested only two widely separated data points, which does not provide enough information to distinguish between a linear increase and the S-shaped curve predicted by the multichannel models. More importantly, it is not clear how the artificial distortion used by Jeffery et al. (and by Burkhardt et al. (2010), Robbins, McKone, and Edwards (2007), and Susilo, McKone, and Edwards (2010)) relates to processing for real faces. Gender is a natural face-space dimension and a clear source of variance in real human populations (cf. Oruç, Guo, & Barton, 2011), whereas distortion leading to non-human-like faces is not. Although the dimensions of neural face space, if any, are not known, one may expect that neural representations would reflect previous visual experience. Given that such visual experience includes faces with varied levels of masculinity and femininity, whereas subjects are unlikely to have previously encountered a variety of unrealistic facial distortions in everyday experience, it is possible that tests with facial distortion may uncover different mechanisms than our experiment using a natural dimension like gender.

Other studies reporting evidence for norm-based representations focus on face identity (Leopold et al., 2001; Leopold, Bondar, & Giese, 2006; Rhodes & Jeffery, 2006; Webster & MacLin, 1999). Some of these, including Leopold et al. (2001) and Webster and MacLin (1999) (as well as Rhodes et al. (2011) for face gender), examine only a relatively limited range of values not far from the mean, and thus have not tested whether the aftereffect curve will decrease for higher values. Thus even though those studies are presented as supporting norm-based theories, it is possible that the underlying phenomena are compatible with an exemplar-based model like the one presented in this study. But other studies of face identity have used a large stimulus range, as well as other methods like adapting to the average face and adapting to an anti-face in the face space, and showed results suggesting monotonically increasing aftereffects, as predicted by norm-based opponent coding. Our current results are only for facial gender, and it is possible that facial identity and distortion are coded differently, so we cannot make a definitive claim rejecting norm-based coding for all kinds of face-related aftereffects. Addressing each of the published studies arguing for norm-based coding would represent a huge research program, beyond the scope of this paper, and so we can only suggest that those studies be reexamined to determine whether a model like the one used in this paper can account for their results.

Although opponent coding is often associated with norm-based coding, it is important to realize that opponency is also a feature of multichannel models. For example, low-level models for motion and color aftereffects can be considered "opponent", in that adaptation to one population (e.g. upward motion) biases the net perception towards the unadapted population (e.g. downward motion; Seriès, Stocker, & Simoncelli, 2009). But such motion and color models are also multichannel, because there are many such populations, with no inherently special population or boundary. The TAE model we described in this paper can also be interpreted as "opponent" in the sense that adapting to one population (e.g. 10°) will bias perception towards the nearby less-adapted populations. But because there are many populations for TAE models and no special orientations, TAE models are not usually described as being based on opponency. Color and motion direction happen to

have clearly defined opposites in a way that orientation does not, leading to a common interpretation of them as "opponent", but the same neural explanation could potentially account for all of them. Thus the degree of opponency is not necessarily an important distinguishing feature of models.

Other recent work has also found evidence for multichannel representations of relatively high-level processing of faces. Chen et al. (2010) very recently found a similar S-shaped curve for face viewpoint aftereffects, and Calder et al. (2008) give evidence that face eye-gaze aftereffects are better coded by a multichannel than a two-pool opponent approach. Both of these studies give support for the general idea of multichannel processing for visual stimuli, but both focus on dimensions that are closely related to orientation and direction, for which the S-shaped curve is already well established. We argue that the gender dimension used in our study is more clearly "high-level" than gaze or viewpoint direction, and thus a better test of the hypothesis that high-level processing uses mechanisms similar those established for low-level features like orientation.

Other recent work by Oruç and Barton (2010) found both similarities and differences between low-level and high-level adaptation. Their results for face contrast adaptation suggest that during adaptation, the responses of the adapted stimulus will be suppressed, but the responses of other stimuli away from the adapted unit will be suppressed even more. They suggested an adaptation scheme combining both "sharpening" and suppression, accounting for their experimental results and existing literature (Fig. 5C in Oruç and Barton (2010)). This model thus differs from ours (Fig. 1), where only the responses of the adapted unit are suppressed. Future work should thus test whether their proposed adaptation model would also lead to an S-shaped curve as in Fig. 1d.

Given that we argue that gender is a natural dimension of face variation, it is important to consider whether the face stimuli illustrated in Fig. 2b do represent "natural" faces. It is difficult to define naturalness objectively, but some features of the most extreme faces do appear unnatural to most observers, such as the relatively dark ears in the +2.5 feminine faces. Such artifacts are due to accentuating gender-related differences in the source images for the face generator, such as long hair more often surrounding female faces. It is possible that these less-natural elements contribute to a weaker aftereffect size after adapting to ±2.5 than after adapting to the ±1.5 faces. However, even for the ±2.5 faces, the overall effect is clearly of a human face, and most features are still quite natural. There is relatively little difference between the ±1.5 and ±2.5 face images, particularly for the masculine faces, compared to the large difference in aftereffect size. The difference in apparent naturalness between the extreme masculine and feminine faces was also not reflected in the aftereffect sizes, which were similar for masculine and feminine faces. Moreover, for participants CZ and CB, for whom additional points were measured between peak points and the ±2 extremes, the dip in aftereffect size was clear even for intermediate faces that most observers would consider very natural. Most importantly, actual human faces in the visual environment can be much more extreme in variation than any of the ±2.5 generated faces, and we would predict even smaller aftereffects from adapting to highly masculine or feminine real faces. To test this prediction, future work could use real human faces rated for masculinity and femininity, rather than the averaging-based generated faces used here that necessarily only cover a subset of the full range of human faces.

One may argue that the factor of face identity also varies along the gender continuum throughout the face sets described in this work. Given that the degree of masculinity or femininity is one clue (of many) to face identity, a gender continuum will effectively and necessarily also vary identity. It is thus possible that our results

reflect identity aftereffects to some degree, but such a confound would not affect our conclusions. Because the explicit behavioral task for the participants was to identity face gender, the measurements should reveal those aspects of the face images that relate to a gender judgment, whether or not they also change the identity.

As mentioned earlier, the stimulus faces were synthesized using a linear morphing method. Of course, each resulting face continuum may or may not be perceptually linear, i.e., with similar differences in face appearance at each pair of neighboring points on the continuum. Even so, as long as the continuum is perceptually monotonic (i.e., with faces consistently becoming more female as the morphing strength increases), finding an S-shaped curve will not be affected by such nonlinearities. Nonlinearities in a monotonic dimension can change the slope of specific parts of the S-shaped aftereffect curve, but could not flatten it as a whole or yield aftereffects that increase indefinitely. The required monotonicity can be verified visually for each dataset in Fig. 2b.

As we mentioned in the experimental procedure above, each experimental block typically took 15–20 min for the TAE and 15–25 min for the FAE. The time difference was due to a larger number of trials and slower average reaction times for the FAE stimuli. In pilot trials, the chosen duration was found to provide a good balance between getting sufficient data to measure the boundary precisely for that block, while avoiding participant fatigue that could compromise the quality of the results. As each participant was asked to conduct 5–10 blocks in total to get at least five adapter positions for the curves in Fig. 3b, each FAE curve represents 100–250 min on average spent performing this difficult task over 5–10 days. Thus it is not practical to repeat the entire task enough times to be able to estimate statistical significance or error bars for each participant using standard methods. Over the course of the 10 or so repetitions of the entire task (17–33 h over 50–100 days) that would be required for each participant in such tests, the participants would be likely to adapt to many aspects of the tests other than those being measured explicitly. Thus it would be difficult to interpret the results, even if time considerations and personal health issues did not preclude running such tests. We are currently working to develop an alternative, less time-consuming method for measuring these effects, using Bayesian methods to limit the range of test stimuli and thus approach the participant's boundary more rapidly (Watt & Andrews, 1981).

5. Conclusion and future work

Our results reveal previously undemonstrated similarity between low-level tilt aftereffects and high-level face gender aftereffects. We found that the FAE curve is not ever-increasing in opposite directions from the average face; instead, it reliably decreases once the adaptation faces are sufficiently dissimilar to the average face. Based on modeling and experimental work on the TAE, we thus suggest that the gender FAE reflects adaptation in neurons that are each tuned to a specific, though broad, range of masculine or feminine faces (cf. Zhao & Chubb, 2001). Our results are compatible with a recent study on the similarity of tuning function and aftereffect curves between TAE and another high-level aftereffect—face viewpoint adaptation (Chen et al., 2010). Our results are not consistent with conclusions from a recent study finding increasing aftereffects as faces were distorted further from the mean face (Jeffery et al., 2010). However, we do not yet know the full shape of the aftereffect curve in such cases, and moreover, such unnatural distortion of face appearance may reveal different mechanisms than natural variation like gender.

These results should prompt a reexamination of previous claims for norm-based encoding of faces. The same methods as in this paper can be applied to any other type of stimulus that can be generated along a continuum that has a natural boundary, in order to see whether the results apply more generally for high-level perception. To further investigate the role of "face norm", it will be informative to see if the S-shaped curve can be obtained by adapting and testing face continuums not passing through the mean face. Alternative experimental work can also employ cross-subject tests on both TAE and FAE, in order to quantify the relationship between their low-level and high-level tuning bandwidth. Alternative modeling work can test whether a more complicated adapting scheme can lead to similar S-shaped curve, in order to compare with other face adaptation experimental data such as Oruç and Barton (2010). It is also worthwhile to conduct future work investigating the relationship between simultaneous high-level and low-level adaptation, to determine if a fraction of the observed gender FAE is due to low-level rather than high-level adaptation (see Xu et al., 2008).

Appendix A. Computational modeling

In this section, we show the details of our TAE and FAE modeling methods. The results of the modeling motivated us to conduct experimental investigation to verify them, but the experimental results do not depend on the specific model used here.

Our simple models for the TAE and FAE were both adapted from the model for the motion direction aftereffect (DAE) described in Seriès, Stocker, and Simoncelli (2009). The TAE model used a population of N neurons with tuning curves

$$f(\theta) = \{f_1(\theta), f_2(\theta), \ldots, f_N(\theta)\} \tag{A.1}$$

describing the mean spike count of each neuron as a function of the stimulus direction θ. These N neurons were chosen to tile the space of all orientations uniformly and have unimodal tuning curves following the circular normal distribution (Seriès, Stocker, & Simoncelli, 2009):

$$f_i(\theta) = G_i \exp(\sigma^{-1}(\cos(\theta - \theta_i) - 1)) \tag{A.2}$$

where the gain G_i controls the response amplitude of neuron i, θ_i its preferred orientation, and σ the width of each tuning curve. The encoding model was specified as the probability of observing a particular population response $r(\theta)$ for a given stimulus θ (see Eq. (5.2) in Seriès, Stocker, and Simoncelli (2009)). For adaptation, the model reduces the gain more for the most active neurons, based on a normal function of the difference between the adaptor direction and the preferred direction of that neuron (see Eq. (5.3) in Seriès, Stocker, and Simoncelli (2009)). This TAE model was identical to the DAE model as published in Seriès, Stocker, and Simoncelli (2009), except for the specific parameter values chosen to fit the TAE data: tuning curve width $\sigma = 0.09\pi$ (was 0.18π in DAE), maximal suppression $\alpha_a = 100$ (was 50 in DAE) and spatial extent of the response suppression $\sigma_a = 0.06\pi$ (was 0.125π in DAE). Details of the suppression model parameters are shown in Eq. (5.3) of Seriès, Stocker, and Simoncelli (2009).

To decode the responses, we used the unaware maximum likelihood (ML) decoder described in Section 2.3 of Seriès, Stocker, and Simoncelli (2009). To match our experimental protocol, we measured how the perception of 0° or face with morphing strength 0 changes with adaptation, as illustrated in Fig. 1a and c.

To simulate the FAE, we modified the TAE model, primarily to account for the qualitatively different facial gender input dimension. Instead of using the circular normal specified in Eq. (7) above, suitable for cyclic quantities like orientation and direction, the FAE model used a non-cyclic normal distribution:

$$f_i(s) = \frac{G_i}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(s-s_i)^2}{2\sigma^2}\right) \tag{A.3}$$

where s denotes the stimulus morphing strength, and s_i the ith neuron's preferred morphing strength; G_i and σ have the same meaning as in Eq. (7). To match the FAE data, the FAE model (Fig. 1c) used broadly tuned neurons covering much of the input space, with $\sigma = 0.8$, $\alpha_a = 24$ and $\sigma_a = 0.64$ (where each is specified in units of morphing strength). Note that for the FAE, the stimulus value range ([−2.5, 2.5] morphing strength) was different from that of TAE ([−0.5π, 0.5π]), and therefore corresponding parameter values have different scaling. The FAE model was otherwise identical to the TAE model.

Appendix B. Experimental data analysis

The data from all trials in each block were collected and used to fit a sigmoidal psychometric function from which the current perceptual boundary could be estimated. The aftereffect for a given stimulus value was then the difference between the adapted perceptual boundary in this block and that of the baseline, leading to one data point in a TAE or FAE curve for a given adaptation orientation or morphing strength (Fig. 3a and b).

Each data point in Fig. 3a or b represents the results for one block of trials. Over the course of a block, each stimulus was presented multiple times in order to measure the reliability of the response. In a TAE block, 14 test points (12 for baseline) were used, and each point was repeated for 8 times, while in an FAE block, typically 18 test points (16 for baseline) were used, and each point was repeated for 8 times. As mentioned above, in FAE experiments, each participant's test ranges were slightly different, so there could be at most 20 or at least 17 test points in adaptation blocks. Therefore, in total 112 (96 for baseline) trials were conducted in a TAE block, and 136–160 (128 for baseline) trials were conducted in an FAE block. Overall, each block consisted of 96–160 trials. All the trials were conducted in random order within a block. Apart from the two extra test stimuli for adapted blocks, other aspects of test stimuli were the same for baseline and adapted blocks.

For each trial in an experimental block, the participant was asked to judge if the stimulus was left or right (TAE), or male or female (FAE). The number of times "right" or "female" was selected was counted for each stimulus. After all eight possible responses have been collected for a block, the response rates r_i were calculated for each stimulus i in a TAE block:

$$r_i = \frac{\text{number of times } right \text{ was selected for stimulus } i}{8} \tag{B.1}$$

or an FAE block:

$$r_i = \frac{\text{number of times } female \text{ was selected for stimulus } i}{8} \tag{B.2}$$

Then a sigmoidal psychometric function model f was fit to the response rate across all test stimuli:

$$f(s) = 0.5 + \frac{1}{\sqrt{\pi}} \int_0^{\frac{s-t}{\sqrt{2}b}} e^{-x^2} dx \tag{B.3}$$

where s is the stimulus test range (degrees for TAE and morphing strength for FAE), and t and b are the threshold and slope parameters of the psychometric curve. The measured perceptual boundary is the x-axis value (threshold) where the psychometric curve has a value of 0.5 on the y axis, representing a perception of vertical or
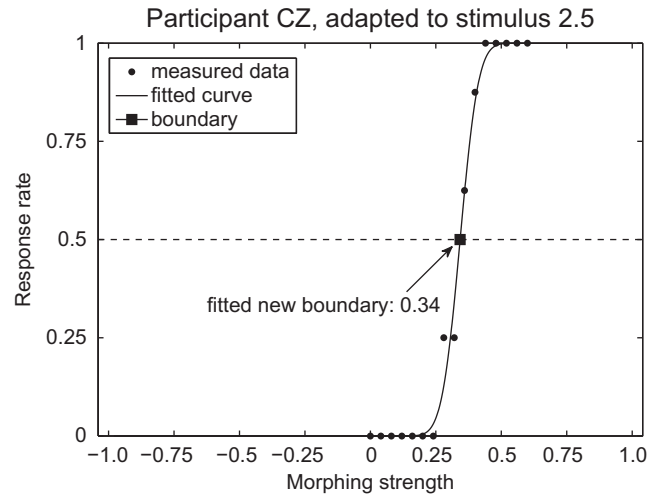
Fig. B.1. Fitted psychometric curve for one of CZ's adapted blocks (adapted to 2.5). The value of this psychometric curve at a y value of 0.5 represents the threshold between male and female perceptual judgments in this condition. The difference between this threshold and the value measured for the baseline represents the amount of aftereffect for this condition (adapted to 2.5).

androgynous. An example of the raw data and the fitted psychometric curve for an FAE block is shown in Fig. B.1. This is an adaptation block (participant CZ adapted to 2.5), and the measured new boundary is 0.27.

This psychometric curve model was fit by minimizing the residual sum of squares (RSS) between the collected rates r_i and model value f_i for each stimulus i:

$$RSS = \sum_{i=1}^{n} (f_i - r_i)^2 \tag{B.4}$$

where n denotes the total number of test stimuli used in a block. We used the residual standard deviation to quantify the quality of the fitting:

$$\hat{\sigma} = \sqrt{\frac{RSS}{n-p}} \tag{B.5}$$

where $p = 2$, denoting the number of parameters in the model. The average residual standard deviation $\hat{\sigma}$ for all the data points in Fig. 3a (TAE) is 0.0317 and for all the data points in Fig. 3b (FAE) is 0.0671.

References

Afraz, S.-R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Research, 48*(1), 42–54.

Bednar, J. A., & Miikkulainen, R. (2000). Tilt aftereffects in a self-organizing model of the primary visual cortex. *Neural Computation, 12*(7), 1721–1740.

Burkhardt, A., Blaha, L. M., Jurs, B. S., Rhodes, G., Jeffery, L., Wyatte, D., et al. (2010). Adaptation modulates the electrophysiological substrates of perceived facial distortion: Support for opponent coding. *Neuropsychologia, 48*(13), 3743–3756.

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.

Chen, J., Yang, H., Wang, A., & Fang, F. (2010). Perceptual consequences of face viewpoint adaptation: Face viewpoint aftereffect, changes of differential sensitivity to face view, and their relationship. *Journal of Vision, 10*(3), 12.1–12.11.

Gibson, J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology, 20*, 453–467.

Jeffery, L., Mckone, E., Haynes, R., & Firth, E. (2010). Four-to-six-year-old children use norm-based coding in face-space. *Journal of Vision, 10*(5), 18, 1–19.

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology, 97*(5), 3155–3164.

Laming, D. (1997). *The measurement of sensation. Oxford psychology series*. Oxford University Press.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature, 442*(7102), 572–575.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4*(1), 89–94.

Little, A. C., Debruine, L. M., & Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society of London Series B – Biological Sciences, 272*(1578), 2283–2287.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience, 8*(10), 1386–1390.

Mitchell, D. E., & Muir, D. W. (1976). Does the tilt after-effect occur in the oblique meridian? *Vision Research, 16*(6), 609–613.

Oruç, İ., & Barton, J. J. S. (2010). A novel face aftereffect based on recognition contrast thresholds. *Vision Research*, 1–10.

Oruç, İ., Guo, X. M., & Barton, J. J. S. (2011). Gender in facial representations: A contrast-based study of adaptation within and between the sexes. *PLoS ONE, 6*(1), e16251.

Penton-Voak, I., Pound, N., Little, A., & Perrett, D. (2006). Personality judgments from natural and composite facial images: More evidence for a kernel of truth" in social perception. *Social Cognition, 24*(5), 607–640.

Poulton, E. (1974). *Tracking skill and manual control*. New York: Academic Press.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19*(4), 473–497.

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.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research, 46*(18), 2977–2987.

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.

Seriès, P., Stocker, A., & Simoncelli, E. P. (2009). Is the homunculus aware of sensory adaptation? *Neural Computation, 21*(12), 3271–3304.

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*(3), 300–314.

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology, 19*(1), R11–R14.

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *Computer Graphics and Applications, IEEE, 21*(5), 42–50. doi:10.1109/38.946630.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology A, 43*(2), 161–204.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology A, 44*(4), 671–703.

Watt, R., & Andrews, D. (1981). Ape: Adaptive probit estimation of psychometric functions. *Current Psychological Reviews, 1*(2), 205–213.

Webster, M. A., & MacLin, O. (1999). Figural after-effects in the perception of faces. *Psychonomic Bulletin and Review, 6*(4), 647–653.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*(6982), 557–561.

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 366*(1571), 1702–1725.

Xu, H., Dayan, P., Lipkin, R. M., & Qian, N. (2008). Adaptation across the cortical hierarchy: Low-level curve adaptation affects high-level facial-expression judgments. *Journal of Neuroscience, 28*(13), 3374–3383.

Zhao, L., & Chubb, C. (2001). The size-tuning of the face-distortion after-effect. *Vision Research, 41*(23), 2979–2994.