Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 974–989
Vision Research
www.elsevier.com/locate/visres

Adaptive face coding and discrimination around the average face

Gillian Rhodes [a,*], Laurence T. Maloney [b], Jenny Turner [a], Louise Ewing [a]

[a] The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
[b] New York University, 6 Washington Place, New York, NY 10003, USA

Received 7 August 2006; received in revised form 7 December 2006

Abstract

Adaptation paradigms highlight the dynamic nature of face coding and suggest that identity is coded relative to an average face that is tuned by experience. In low-level vision, adaptive coding can enhance sensitivity to differences around the adapted level. We investigated whether sensitivity to differences around the average face is similarly enhanced. Converging evidence from three paradigms showed no enhancement. Discrimination of small interocular spacing differences was not better for faces close to the average (Study 1). Nor was perceived similarity reduced for face pairs close to (spanning) the average (Study 2). On the contrary, these pairs were judged most similar. Maximum likelihood perceptual difference scaling (Studies 3 and 4) confirmed that sensitivity to differences was reduced, not enhanced, around the average. We conclude that adaptive face coding does not enhance discrimination around the average face.

Keywords: Face perception; Adaptation; Discrimination; Average face

1. Introduction

People are experts at discriminating between faces, despite the fact that faces all share a very similar internal structure (Maurer, Le Grand, & Mondloch, 2002). This ability rests on adaptive face coding mechanisms that are dynamically updated by the diet of faces experienced (for recent reviews see Clifford & Rhodes, 2005).

Many theorists have proposed that an average face, or prototype, is developed through experience, and could function as a norm against which identity is coded (e.g., Carey, 1996; Goldstein & Chance, 1980; Hebb, 1949; Hochberg, 1978; Loffler, Yourganov, Wilkinson, & Wilson, 2005; Rhodes, 1996; Rhodes, Brennan, & Carey, 1987; Valentine, 1991; Valentine & Bruce, 1986). Evidence for adaptive norm-based coding of faces comes from recent studies of face identity after-effects, in which adapting to a face biases perception towards the opposite identity (Anderson & Wilson, 2005; Leopold, O'Toole, Vetter, & Blanz, 2001; Leopold, Rhodes, Müller, & Jeffery, 2005; Rhodes & Jeffery, 2006; Tsao & Freiwald, 2006). For example, in Fig. 1, adapting to the identity opposite Dan (termed 'anti-Dan') temporarily shifts the average face towards that face, thus increasing the identity strength of Dan and reducing thresholds for identification of Dan. Adaptation to non-opposite, but equally dissimilar, identities does not facilitate identification of Dan (Rhodes & Jeffery, 2006). These results strongly suggest that the average face functions as a norm for coding identity.

Figural face after-effects also highlight the adaptive nature of face coding (MacLin and Webster, 2001; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Rhodes, et al., 2004; Watson and Clifford, 2003; Webster and MacLin, 1999; Webster, Werner, & Field, 2005). For example, after adapting to faces whose internal features have been "expanded", slightly expanded faces appear normal and after adapting to "contracted" faces, slightly contracted faces appear normal (e.g., Rhodes et al., 2003). In each case, undistorted faces appear distorted in the opposite way to the adapting faces. These after-effects demonstrate that what looks normal or average is dynamically calibrated by experience.

* Corresponding author. Fax: +61 8 6488 1006.
E-mail address: gill@psy.uwa.edu.au (G. Rhodes).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.12.010

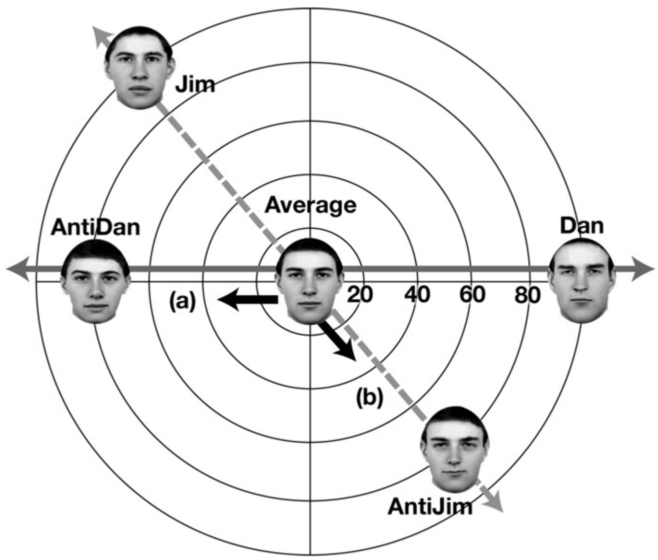

Fig. 1. An example of two identity continua in a simplified face space. The original faces have identity strengths of 100% and the average face has an identity strength of 0%. The computationally opposite anti-face is shown for each identity at −80% identity strength. The arrows represent the direction in which the average shifts after adapting to anti-Dan (a) and anti-Jim (b), respectively.

Both the figural and identity after-effects suggest that faces are coded relative to the average (Hurlbert, 2001; Rhodes et al., 2005). When the average is altered by adaptation, the appearance of subsequently viewed faces changes, because their deviation from the average changes. These changes affect identification (e.g., Leopold et al., 2001), perceived normality (e.g., Rhodes et al., 2003) and judgments of other socially relevant attributes including attractiveness (Rhodes et al., 2003), gender, race and expression (Webster, Kaping, Mizokami, & Duhamel, 2004). Conversely adapting to an average face has little effect on subsequent perception because it does not alter the psychological average (e.g., Leopold et al., 2001; Webster & MacLin, 1999).

Face after-effects are not solely due to adaptation of low-level coding mechanisms (Leopold et al., 2001; Rhodes et al., 2003; Watson & Clifford, 2003; Zhao & Chubb, 2001). They survive changes in orientation (+/−45° tilt from upright) (Rhodes et al., 2003; Watson & Clifford, 2003), retinal position (Leopold et al., 2001) and size (Zhao & Chubb, 2001) between the adapting and test stimuli. They are therefore mediated by higher-level visual mechanisms, and quite possibly by face-selective coding mechanisms (Rhodes et al., 2004).

Nevertheless face adaptation has similar perceptual consequences to adaptation of lower-level coding mechanisms. In both cases, perception is often biased towards the opposite attributes to those seen during adaptation. For example after adapting to downward motion, a stationary stimulus appears to move upwards (Mather & Harris, 1998) and after adaptation to a contracted face, an undistorted face appears expanded (e.g., Webster & MacLin, 1999). Furthermore, low-level and face after-effects follow a similar time-course, increasing logarithmically with adapting duration and decaying exponentially with test duration (Leopold et al., 2005).

These parallels raise the possibility that adaptation serves similar functions in low-level and high-level vision. An important function may be to optimize the use of a limited neural response range, by calibrating coding mechanisms to the visual environment (Barlow, 1990; Clifford, 2002; Clifford & Rhodes, 2005). This calibration can ensure good discrimination within that environment (e.g., Clifford, Ma Wyatt, Arnold, Smith, & Wenderoth, 2001; Phinney, Bowd, & Patterson, 1997; Regan & Beverley, 1985). For example, lightness adaptation in the retina ensures good discrimination of differences from the adapted level (Werblin, 1973). Adaptation of cortical coding mechanisms may have a similar function, with several studies reporting enhanced discrimination around the adapted level. This has been reported for direction of motion (Phinney et al., 1997, but see Hol & Treue, 2001), speed (Krekelberg, Van Wezel, & Albright, 2006), orientation (Clifford et al., 2001; Regan & Beverley, 1985, but see Barlow, MacLeod, & Van Meeteren, 1976) and contrast (Abbonizio, Langley, & Clifford, 2002; but see Barlow et al., 1976; Maattanen & Koenderink, 1991).

These findings suggest that adaptation in low-level vision may function to enhance discrimination of commonly experienced stimuli. Here, we consider whether face adaptation has a similar function. If it does, then discrimination should be enhanced around the adapted state, i.e., around the average face. To our knowledge only one study has tested this hypothesis directly (Wilson, Loffler, & Wilkinson, 2002). Wilson and colleagues compared thresholds for discriminating differences among two sets of synthetic, radial frequency faces, one close to, and one far from, the average face. Thresholds were significantly lower around the average face. Others have suggested that there is a discontinuity in the perception of face identity as a face crosses over to the, "other side of the mean" (Blanz, O'Toole, Vetter, & Wild, 2000, p. 885). Such a discontinuity could also reflect enhanced sensitivity to differences around the average, but caution is needed here because perceptual sensitivity was not directly assessed.

We report four studies, using three different paradigms, that test whether sensitivity is enhanced for differences around the average face. In all cases, we used digitized grey-scale images of faces, which look more like real faces than the simplified, synthetic faces used by Wilson et al. (2002). First, we measured discrimination of subtle changes in interocular spacing for faces at various distances from the average face (Study 1). Second, we compared similarity ratings for face pairs, at varying distances from the average (Study 2). Third, we derived super-threshold perceptual difference scales for faces at varying distances from the average (Studies 3 and 4). These scales were derived using maximum likelihood estimation (Maloney & Yang, 2003) and show how sensitivity to super-threshold differences changes with distance from the average face.

2. Study 1

We measured sensitivity to subtle changes in interocular spacing as a function of distance from the average face. Interocular spacing was chosen because the eyes are salient and important for face identification (Vinette, Gosselin, & Schyns, 2004). On each trial participants indicated which of two simultaneously presented faces, which differed only on interocular distance, had closer-set eyes. We hypothesized that discrimination would be best for pairs closest to the average. Distance from the average was manipulated by using target faces of different identity strengths. An original face has an identity strength of 100% and the average face has an identity strength of 0%. Each identity continuum was made by morphing a male target face towards an average male face (0%) and beyond to its anti-face (e.g., along the trajectory joining Dan to anti-Dan in Fig. 1).

We included three adapting conditions to provide explicit control over the participants' adapted state. We adapted participants to the average face on some trials (average adapt), expecting any peak in discrimination performance to be located at 0% (the average) on these trials. We also adapted participants to the target's anti-face (match adapt). This will shift the average to a slightly negative identity strength (≈−10%, Leopold et al., 2001), so that any peak in performance around the average should shift accordingly. Finally, we included a standard control condition in which participants adapted to a mismatching anti-face from a different identity continuum to the target face (Leopold et al., 2001). No systematic shift is expected in this condition because the average will shift in different directions depending on the anti-face used, resulting in no consistent shift along the target identity continuum (Fig. 1).

2.1. Method

2.1.1. Participants

Seventeen (15 female) undergraduates (aged 17–20, $M = 18.3$, $SD = 0.7$ years) received course credit for participating. All had normal or corrected to normal vision and were naïve to the adaptation paradigm and hypotheses.

2.1.2. Stimuli and apparatus

The stimuli were derived from four identity continua used in Leopold et al. (2005, face set 2). These were based on four black and white, front-view photographs of adult male faces (Ted, Rob, Jim, Dan) with neutral expressions chosen from a pool of 20 male photographs, and an average face constructed from these 20 faces using standard morphing procedures. Each continuum consisted of a series of morphed faces, ranging from the original face (100% identity strength) through the average face (0%) to an 80% anti-face (100% anti-faces showed some distortions). The images were created by morphing the original face toward and beyond the average face (Fig. 1) (see Rhodes, Sumich, & Byatt, 1999, for details). All images had the texture of the average face.

The anti-faces and the average face were used as adapting stimuli. The −40%, −30%, −20%, −10%, +10%, +20% and +30% images from each identity continuum were used to create the discrimination stimuli. For each image on each continuum the eyes (including the eyebrows) were shifted both toward (eyes-in), and away from (eyes-out), the midline by a distance of 2 pixels (Fig. 2). Pilot testing ($N = 11$; 9 female) revealed that this shift was discriminable on about 75% of trials for +20% and −20% identity strength images with a 500 ms exposure time. With performance above chance (50%), but below ceiling (100%), we should be able to see the predicted peak in performance around 0% strength in the average-adapt and no-adapt conditions.

The grey-scale textures of the average face were mapped onto the structure of all stimuli faces, to ensure uniform complexions, and all face stimuli were presented surrounded by grey oval masks that hid most of the hair (Fig. 2). Adapting stimuli were $9.9° \times 11.7°$ (within the oval mask), at a viewing distance of approximately 57 cm. Discrimination pairs consisted of one image from an identity continuum, presented next to an eyes-moved version of that image, with 3.1° visual angle between the inner edge of each stimulus. Every original image was paired once with the eyes-in version and once with the eyes-out version. All stimuli were presented on a 17-inch monitor (820×624 pixels resolution), using SuperLab software, and participants responded using a standard computer keyboard.

2.1.3. Procedure

There were three adapting conditions: (a) average adapt, in which the average face was used as the adapting stimulus; (b) anti-face match adapt, in which the anti-face corresponding to the identity used in the discrimination pair was used as the adapting stimulus; (c) anti-face mismatch adapt, in which the anti-face of a different identity was used as the adapting stimulus. For each discrimination pair identity, the three non-matching anti-faces were used an equal number of times as the adapting image.

Fig. 2. The average (0%) face (center) surrounded by its corresponding eyes-moved stimuli. Eyes-in is shown on the left, and eyes-out on the right. In both cases the eyes and eyebrows (together) have been moved 2 pixels.

Each trial consisted of a 150 ms central fixation cross (1.7° × 1.7° visual angle), followed by a 4000 ms central adapting stimulus, followed by a 150 ms inter-stimulus interval, followed by the discrimination pair for 500 ms. Participants were asked to indicate, "which face had the eyes closer together? The left or the right?", using labelled keyboard keys. The next trial began immediately following the participant's response. There were 384 test trials (4 identity continua × 8 images/continuum × 2 discrimination pair types × 2 left–right positions × 3 three adapting conditions) presented in random order with rest breaks every 96 trials. Participants began with 12 practice trials to illustrate the task.

2.2. Results and discussion

The mean proportion correct for each identity strength, in each adapting condition, was calculated and plotted as a function of identity strength (Fig. 3). There was no peak in discriminability around the average, or anywhere else, for any adapting condition. This was confirmed by a two-way repeated measures ANOVA on the proportion correct scores, with identity strength (−40, −30, −20, −10, 0, +10, +20, +30) and adapting condition (average, anti-face match, anti-face mismatch) as repeated measures factors. There was no significant main effect of either identity strength, $F(7,112) = 0.73$, $p = .65$, or adapting condition, $F(2,32) = 1.52$, $p = .23$, and no interaction, $F(14,224) = 1.15$, $p = .31$.

These results do not support the hypothesis that discrimination is facilitated around the average face. The absence of any peak in the discrimination function cannot be attributed to floor or ceiling effects, because performance was well above chance (50%) and well below ceiling (100%). In the absence of any peak, no interaction between identity strength and adapting condition would be expected and none was found. The absence of any peak also precludes us from testing whether the 0% is a good approximation of the psychological average face for our participants.

Our results contrast with those of Wilson et al. (2002), which showed better discrimination around the average face. There are many differences between their study and ours, but an important one may be that our discrimination pairs differed only on a single feature whereas theirs differed globally. Our participants may have focussed on the eye region in the discrimination pairs given the very brief presentation times used. Indeed some participants volunteered that they concentrated on the bridge of the nose, which may not have fully engaged face processing mechanisms. In this case, the discrimination test would be insensitive to adaptation of higher-level face coding mechanisms resulting from viewing the adapting face. Although viewing the adapting face would also adapt lower-level, retinotopically coded attributes (as well as spatiotopically coded attributes), these effects would not be picked up because of the changes in retinal (and spatial) position between the adapt and test faces.

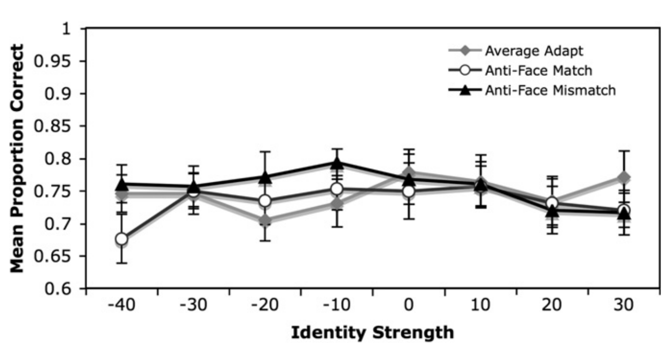

Fig. 3. Mean proportion correct (+SE) for each identity strength, in each adapting condition in Study 1.

In Study 2, therefore, we used faces that differed globally and were presented for longer. We also moved to a super-threshold discrimination task, which may be more relevant to normal face processing goals. Only in the case of identical twins, do we need to determine whether two faces are truly identical or not.

3. Study 2

In this study, participants rated the similarity of pairs of faces at varying distances from the average. Similarity ratings provide a measure of super-threshold discrimination (McKone, Martini, & Nakayama, 2001) and have been used successfully to model psychological face spaces and predict recognition performance (e.g., Lee, Byatt, & Rhodes, 2000; Rhodes, 1988). The two faces in each pair came from the same identity continuum but differed in identity strength, by either 40% (−50/−10, −20/+20, +10/+50) or 20% (−30/−10, −20/0, −10/+10, 0/+20, +10/+30) identity steps. The two step sizes were used to increase generality of the results. If sensitivity is enhanced for differences around the average face (0%), then similarity ratings should be *lowest* for pairs closest to the average. Note that low similarity corresponds to high discriminability, so that we now expect a *trough* rather than a *peak* (Study 1) around the average.

We included the three adapting conditions from Study 1, to provide explicit control over the adapted state of the participants. We also added a no-adapt condition, where pairs were rated in the absence of adapting faces, so that we could test our assumption that the average face (0%) approximates the psychological average for our participants. If it does, then performance in the no-adapt condition should match that in the average-adapt condition, where each pair is explicitly preceded by the average face (0%). In both cases, we expect a trough in similarity ratings around 0%. In the anti-face match and mismatch adapting conditions, face pairs were preceded by anti-faces from the same or different identity continua, respectively. Adapting to matching anti-faces should shift the average left along the identity continuum to a slightly negative (≈−10%) identity strength (Fig. 1). In the 40% condition, the −20/20 pairs should still span the (new) average, and so similarity ratings will remain lowest for this pair. However, we should see an asymmetry with lower ratings for −50/−10 pairs than +10/+50 pairs, because the average is now closer to the former than the latter. A similar asymmetry may also occur in the 20% condition, with lower ratings for negative than corresponding positive identity strength pairs (e.g., −20/0 lower than 0/+20). In addition, any trough in similarity ratings might shift left, from the −10/+10 pairs to the −20/0 pairs. As in Study 1, adapting to mismatching anti-faces will shift the average in a variety of directions, with no consistent effect on the location of any trough in the similarity ratings.

3.1. Method

3.1.1. Participants

Twenty-nine naïve participants (22 female) were recruited (aged 17–24, $M = 18.7$, $SD = 1.8$ years). All had normal or corrected to normal vision. Twenty-three received course credit for participation. None had participated in Study 1.

3.1.2. Stimuli and apparatus

The adapting stimuli were the same as in Study 1. The discrimination stimuli were taken from the four identity continua used in Study 1 (with no alteration to the eyes), with the addition of +50% and −50% identity strength images. Discrimination pairs comprised two images of the same identity varying by 20% or 40% in identity strength. For each of the four identities there were five discrimination pairs differing by 20% identity strength: −30/−10; −20/0; −10/+10; 0/+20 and +10/+30, and three discrimination pairs differing by 40%: −50/−10; −20/+20 and +10/+50, making a total of 32 discrimination pairs. Each face was 6.0° × 7.2°, at a viewing distance of approximately 57 cm. The discrimination stimuli were presented side-by-side in the center of the screen with a 2.0° visual angle between their inner edges. As in Study 1, stimuli were presented on a 17-inch monitor (820 × 624 pixels resolution), using SuperLab software.

3.1.3. Procedure

There were four blocks of trials: a familiarisation block, a no-adapt block, an adapt block in which the three adapting conditions were intermixed and a second no-adapt block.

3.1.3.1. Familiarisation block. Participants were presented with all 32 discrimination pairs (left/right position of highest identity strength counterbalanced) to familiarise them with the range of stimuli. Each pair was presented for 2000 ms with an inter-stimulus interval of 400 ms. Participants were instructed to pay close attention to these faces, but not to respond at this time.

3.1.3.2. No-adapt block. Participants were asked to rate the similarity of the discrimination pairs, and encouraged to use the whole range of the scale. Each pair was presented for 1000 ms followed by presentation of the question and scale: "How similar did those faces appear to you? 1 (less similar) to 9 (extremely similar)", which remained on the screen until the participant responded. Participants responded using the number keys on the keyboard and pressed the space-bar to begin the next trial. For each identity (4) each discrimination pair (8) was presented twice (with reversed left/right position of highest identity strength), making a total of 64 trials, which were presented in random order.

3.1.3.3. Adapt block. Trials were the same as in the no-adapt block except that an adapting face preceded each discrimination pair. The adapting sequence was the same as in Study 1. The only difference from Study 1 was that the discrimination pairs were presented for twice as long (1000 vs 500 ms). Each discrimination trial (64) was presented once in each adapting condition (3) making a total of 192 adaptation trials, presented in random order. This block began with five practice trials, chosen to represent the range of trials in the adapting phase.

3.2. Results and discussion

3.2.1. Forty percent identity steps

A preliminary analysis indicated that the pattern of similarity ratings did not differ for the two no-adapt blocks ($F < 1$, pair × adapting condition) so the data were combined. A two-way repeated measures ANOVA was conducted on mean similarity ratings with pair (−50/−10, −20/+20, +10/+50) and adapting condition (none, average, anti-face match, anti-face mismatch) as repeated measures variables. There were significant main effects of adapting condition, $F(3,84) = 5.50$, $p < .002$, and pair, $F(2,56) = 80.73$, $p < .0001$, and as expected, a significant interaction, $F(6,168) = 10.52$, $p < .0001$. To explore the interaction, we conducted separate one-way ANOVAs, with pair as a repeated measures factor, for each adapting condition. We used planned *t*-tests with Bonferroni correction for multiple comparisons ($N = 3$, critical $p = .017$) to compare −20/+20 pairs, which spanned the average, with the −50/−10 and +10/+50 pairs, and those pairs with each other. The latter were expected to differ only in the match adapt condition.

3.2.1.1. No-adapt condition. There was a significant main effect of pair, $F(2,56) = 159.77$, $p < .0001$ (Fig. 4a). However, the pattern was opposite to that predicted, with a peak rather than a trough at 0%. Similarity was significantly higher around the average (−10/+10) than for the other two pairs (−10/+10 vs −50/−10, $t = 16.03$, $p < .0001$; −10/+10 vs +10/+50, $t = 14.86$, $p < .0001$, all surviving Bonferroni correction). As expected, the −50/−10 and +10/+50 pairs did not differ, $t < 1$.

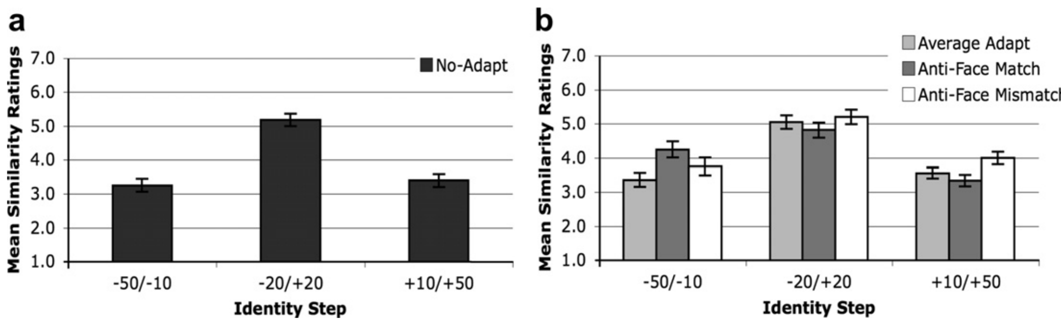

Fig. 4. Mean similarity ratings (+SE) for each 40% identity step pair in the no-adapt (a) and average, anti-face match, and anti-face mismatch adapting conditions (b) in Study 2.

3.2.1.2. Average-adapt condition. The results closely resembled those in the no-adapt condition (Fig. 4b vs 4a), as expected if 0% matches the psychological average. There was a significant main effect of pair, $F(2,56) = 54.31$, $p < .0001$, with higher ratings for −20/20 pairs than −50/−10, $t = 9.54$, $p < .0001$, or +10/+50 pairs, $t = 8.41$, $p < .0001$ and no difference between −50/−10 and +10/+50 pairs, $t = 1.13$, $p = .26$.

3.2.1.3. Direct comparison of no-adapt and average-adapt conditions. To test whether the pattern of similarity ratings differed in the no-adapt and average-adapt conditions we conducted a two-way (pair × adapting condition) repeated measures ANOVA, restricted to these two levels of adapting condition. Importantly there was no interaction, $F(2,56) = 1.17$, $p = .32$, and therefore no evidence that the pattern of similarity ratings differed in the no-adapt and average-adapt conditions. This result supports our assumption that the average face (0%) matches the psychological average for this population.

3.2.1.4. Anti-face match adapt condition. The predictions for this condition were made assuming that a trough in similarity ratings would be found around 0%. Although we observed a peak rather than a trough, shifts should still be seen in the predicted directions. Whatever happens at the average (peak or trough) should shift when the average shifts. Recall that adapting to the matching anti-face will shift the average to a weak negative identity strength (≈−10%). Therefore the peak observed above should remain at −20/+20, but ratings should increase for −50/−10 pairs relative to +10/+50 pairs because the average has shifted towards them. The results confirmed this pattern (Fig. 4b). There was a significant main effect of pair, $F(2,56) = 25.33$, $p < .0001$, with higher ratings for −20/+20 pairs than either −50/−10, $t = 2.71$, $p < .009$, or +10/+50 pairs, $t = 7.05$, $p < .0001$, and higher ratings for −50/−10 than +10/+50 pairs, $t = 4.34$, $p < .0001$.

3.2.1.5. Anti-face mismatch adapt condition. As expected, the pattern was the same as for the no-adapt and average-adapt conditions (Fig. 4b). There was a significant main effect of pair, $F(2,56) = 25.12$, $p < .0001$, with higher ratings for −20/+20 pairs than −50/−10, $t = 6.63$, $p < .0001$, or +10/+50 pairs, $t = 5.48$, $p < .0001$, and no difference between −50/−10 and +10/+50 pairs, $t = 1.14$, $p = .26$.

3.2.2. Twenty percent identity steps

A preliminary analysis indicated that the pattern of similarity ratings did not differ for the two no-adapt blocks ($F < 1$, pair × adapting condition), so their data were combined. A two-way repeated measures ANOVA was conducted on mean similarity ratings with pair (−30/−10, −20/0, −10/+10, 0/+20, +10/+30) and adapting condition (none, average, anti-face match, anti-face mismatch) as repeated measures variables. There were significant main effects of adapting condition, $F(3,84) = 4.82$, $p < .004$, and pair, $F(4,112) = 8.63$, $p < .0001$. Importantly, there was a significant interaction, as expected, $F(12,336) = 3.16$, $p < .0003$. To explore the interaction, we conducted separate one-way ANOVAs, with pair as a repeated measures factor, for each adapting condition. Planned *t*-tests with Bonferroni correction for multiple comparisons were used to compare the pair that spanned the average (−10/10 pair) with each of the other four pairs.

3.2.2.1. No-adapt condition. In the absence of adaptation, there was a significant main effect of pair, $F(4,112) = 6.91$, $p < .0001$ (Fig. 5a). As for 40% identity steps, the pattern was opposite to that predicted, with a peak rather than a trough at 0%. Pairs were significantly more similar around the average (−10/+10) than elsewhere (−30/−10: $t = 4.66$, $p < .0001$; −20/0: $t = 2.80$, $p < .006$; 0/+20: $t = 2.57$, $p < .02$; +10/+30: $t = 4.38$, $p < .0001$), with all but one comparison surviving Bonferroni correction ($N = 4$, critical $p = .0125$).

3.2.2.2. Average-adapt condition. As for 40% steps, the pattern closely matched that in the no-adapt condition (Fig. 5b vs 5a). Again, there was a significant main effect of pair, $F(4,112) = 3.83$, $p < .006$, with a peak rather than a trough at 0%. Similarity was significantly higher around the average than elsewhere (−30/−10: $t = 3.37$, $p < .001$; −20/0: $t = 2.51$, $p < .02$; 0/+20: $t = 2.44$, $p < .02$; +10/+30: $t = 3.37$, $p < .001$). The differences from both 10/30

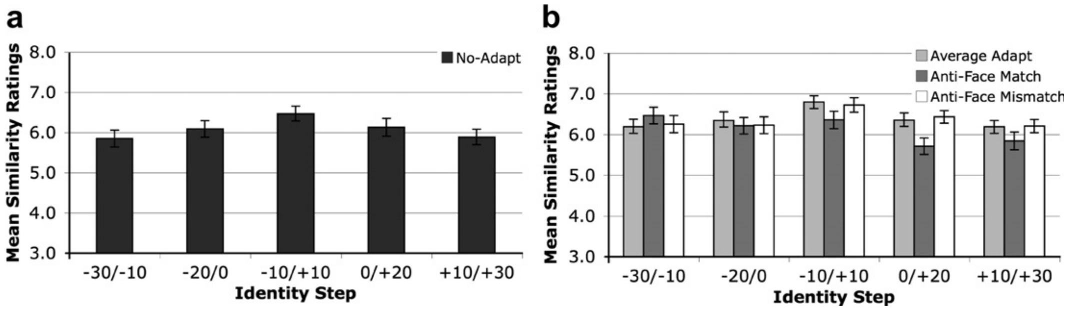

Fig. 5. Mean similarity ratings (+SE) for each 20% identity step pair in the no-adapt (a) and average, anti-face match, and anti-face mismatch adapting conditions (b) in Study 2.

steps survived Bonferroni correction. As in the 40% condition, the close match between the average-adapt and no-adapt conditions confirms that the average face (0%) matches the psychological average for this population.

3.2.2.3. Direct comparison of no-adapt and average-adapt conditions. As above, we used a two-way (pair × adapting condition) repeated measures ANOVA, restricted to these two levels of adapting condition, to test whether the similarity patterns differed for these conditions. As for the 40% step condition, there was no significant interaction, $F(4,112) = 0.11$, $p = .98$, supporting our assumption that the 0% image used here approximates the psychological average for our population.

3.2.2.4. Anti-face match adapt condition. We expected the average to shift left towards a weak identity strength (≈−10%). Therefore, the peak observed above should shift left, most likely to −20/0 pairs. There was a significant main effect of pair, $F(4,112) = 7.10$, $p < .0001$ (Fig. 5b), although the peak was at −30/−10, rather than −20/0, consistent with a larger shift than expected. However, the −30/−10 pairs were not significantly higher than the −10/+10 pairs, $t = 0.62$, $p = .54$. Nor were the −20/0 pairs (−20/0 vs −10/+10, $t = 0.65$, $p = .42$). Nevertheless, the similarity function was no longer symmetric about the −10/10 pairs, consistent with a shift in the average. The faces in both negative identity pairs were significantly more similar than the faces in the corresponding positive pairs (−30/−10 vs +10/+30, $t = 3.60$, $p < .0005$; −20/0 vs 0/+20, $t = 2.90$, $p < .005$), which remained significantly less similar than −10/+10 pairs (+10/+30 vs −10/+10: $t = 2.98$, $p < .004$; 0/+20 vs −10/+10: $t = 3.70$, $p < .0003$). All comparisons survived Bonferroni correction (this time for 6 comparisons, critical $p = .008$).

3.2.2.5. Anti-face mismatch adapt condition. Performance in this condition resembled that in the no-adapt and average-adapt conditions, as expected (Fig. 5b). There was a significant main effect of pair, $F(4,112) = 3.70$, $p < .008$, with a peak rather than a trough at 0%. Similarity was significantly higher around the average than elsewhere (−30/−10: $t = 2.93$, $p < .005$; −20/0: $t = 3.10$, $p < .003$; +10/+30: $t = 3.24$, $p < .002$), except for the 0/+20 pair, for which the difference was only marginally significant, $t = 1.83$, $p < .07$.

3.2.3. Summary

Super-threshold discrimination was not facilitated around the average face. On the contrary, face pairs close to the average were perceived as more, not less, similar than pairs further from the average. This finding was replicated with two different identity step sizes (20% and 40%) and under a variety of adapting conditions. With adaptation to the matching anti-face, negative identity strength face pairs were rated as more similar than the corresponding positive identity face pairs, consistent with the expected shift of the average towards the anti-face. With no adaptation and adaptation to the average face or the mismatching anti-face, similarity was greatest for pairs spanning the average (0%) face.

These results are not consistent with the discontinuity in perceived identity across the average in face space reported by Blanz and colleagues (2000). Our pairs differed by either 20% or 40% steps in identity strength, where theirs differed by 67% (−33/+33 vs +33/+100).[1] It is possible, therefore, that discrimination around the average may be enhanced for larger differences than those used here. However, this seems unlikely because our 40% differences are well above identification thresholds (Leopold et al., 2001; Rhodes & Jeffery, 2006).

Another possibility is that Blanz et al.'s (2000) results do not reflect a genuine perceptual discontinuity. Their task required participants to categorize face pairs as, "same person (1)", "similar person (2)" or a "dissimilar person (3)", which will certainly induce response discontinuities. However, these need not reflect perceptual discontinuities. In signal detection theory the observer's choice of sensory criterion (Green & Swets, 1966/1974) is affected by many non-sensory factors (prior odds, payoffs) but the accompanying changes in response are not thought to alter perception. Furthermore, they assumed that the alternative to a discontinuous relationship between perceptual distance and distance in face space was a linear relationship: " … if no

[1] They also used smaller differences, but in those cases the face pairs that did and did not span the average were not equated on identity strength.

discontinuity exists, we would expect the step sizes to be equal for all equidistant pairs.'' (Blanz and colleagues, 2000, p. 889). With this assumption, they can test for linearity and take any evidence of non-linearity as evidence for the presence of a jump discontinuity. However, these are not the only alternatives. In summary, it is unclear whether Blanz et al.'s dependent measure reflects distance in a perceptual face space. In Studies 3 and 4, therefore, we derived perceptual difference scales for face continua using an explicit model of difference measurement.

4. Study 3

In this study, participants indicated which of two pairs of faces, taken from one of four face/anti-face identity continua, contained the more similar faces. The responses were used to construct super-threshold perceptual difference scales for each participant, and for a ''supersubject'' based on their combined data. The scales were constructed using a maximum likelihood estimation scaling method, which has good reliability and distributional robustness (Maloney & Yang, 2003). If perceptual differences are greater around the average face, then the resulting functions of perceived against physical scale values should be steeper around the average. This ''crispening'' occurs in color space, where people are more sensitive to physical differences in the vicinity of the neutral ''white'' point than elsewhere (Whittle, 1992). Given the results of Study 2, however, we might rather expect to see ''anti-crispening'', with flattish functions around the average face. The results of Study 2 confirmed that the 0% face was a good approximation to the psychological average, so we dispensed with the adapting manipulation in this study.

4.1. Method

4.1.1. Participants

Sixteen (13 female) Caucasian adults (aged 21–38, $M = 25.2$, $SD = 5.1$ years) participated. Eight of the participants had been exposed to these identity continua in previous studies, and were labelled ''experts''. The rest had no prior exposure and were labelled ''novices''. These groups were distinguished just in case the difference in experience affected performance, although no particular difference was expected.

4.1.2. Stimuli

We constructed a face/anti-face identity continuum for each of the four male faces used in Studies 1 and 2 (Ted, Rob, Jim, Dan). Each contained 13 images with the following identity strengths (denoted s_i, $i = 1, 2, \cdots, 13$): $-90, -75, -60, -45, -30, -15, 0, 15, 30, 45, 60, 75, 90$ (Fig. 6). An identity strength of 0 corresponds to the average face. Identity strengths of 90 and -90 (not shown) approximate the original face (100) and its anti-face (-100), respectively. Full strength versions were not used because -100 anti-faces contained obvious distortions. The images on each continuum were created by warping the original face towards the average face (reducing identity strength from 90 to 0) and beyond (negative identity strengths). All images had the texture of the average face. The resulting images were sharpened and placed in an oval mask which hid most of the hair, but not the face outline. Images measured 6.2×8.8 cm and were viewed from approximately 60 cm.

4.1.3. Procedure

On each trial participants saw a quadruple of images from the same identity continuum, arranged as two pairs presented one above the other (Fig. 7). They were asked

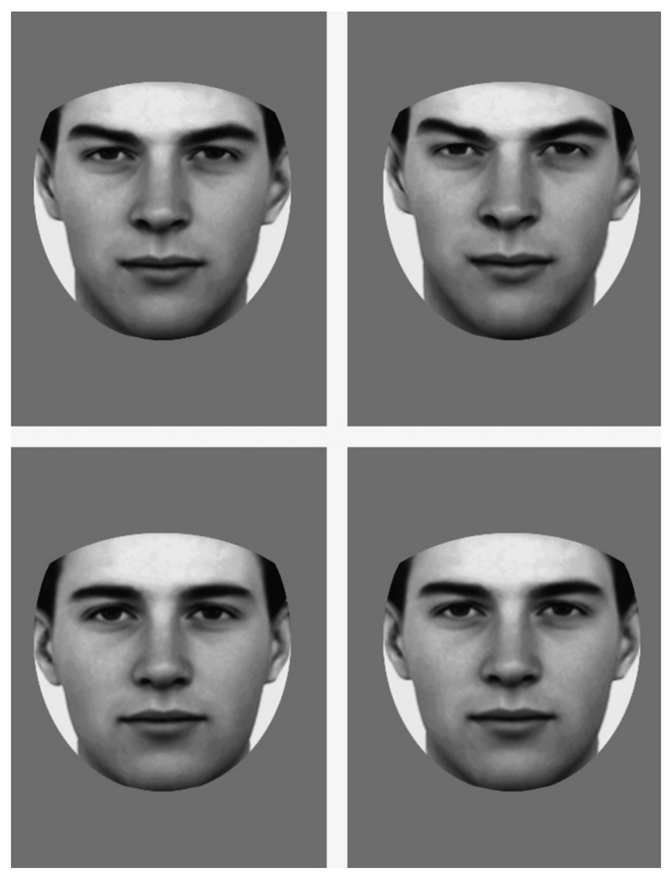

Fig. 7. A quadruple of images from Study 3.

Fig. 6. Example of a 13-face identity continuum used in Study 3. Identity strengths range from −90 (left end) to +90 (right end) in 15% steps. The average face (center) had 0% identity strength.

to indicate which pair looked more similar using labelled keyboard keys. The images remained visible until a response was made. The next trial was initiated by pressing the space-bar. Each participant completed 715 trials, consisting of all distinct non-overlapping quadruples for 13 images (Maloney & Yang, 2003). Trials were presented in random order, with the top–bottom arrangement of pairs randomized. Each participant saw faces from a single continuum and each continuum was rated by two experts and two novices. Each session began with six practice trials showing quadruples (from a different identity continuum to the one being tested) in which one pair was obviously more similar than the other. Each test session lasted approximately 90 min.

4.2. Results and discussion

In order to determine whether crispening occurred around the average, we fit the data using two procedures. We refer to the first procedure, taken from Maloney and Yang (2003), as the *non-parametric fit* and the second as the *parametric fit*. The non-parametric fit assigns numbers $\psi(s_i)$, $i = 1, \cdots, 13$ to the 13 faces with identity strengths s_i, $i = 1, \cdots, 13$ (listed above). It assigns numbers so that, given any two pairs of faces with identity strengths (s_1, s_2) and (s_3, s_4), respectively, the observer should judge the first pair to be more similar precisely when,

$$|\psi(s_1) - \psi(s_2)| < |\psi(s_3) - \psi(s_4)|. \tag{1}$$

We refer to the numerical differences $|\psi(s_1) - \psi(s_2)|$ and $|\psi(s_3) - \psi(s_4)|$ as the perceived *super-threshold differences* between the stimuli. Of course, when the super-threshold differences are almost equal, we do not expect the observer to consistently pick one or the other pair. Maloney and Yang (2003) treat the selection of the larger difference as a signal detection task where the observer is forced to choose between two differences contaminated by Gaussian noise. The standard deviation of the noise distribution is denoted σ and is estimated as part of the maximum likelihood fitting procedure. It is a measure of the observer's ability to discriminate between pairs. The non-parametric fitting procedure computes maximum likelihood estimates of the 13 parameters $\psi(s_i)$, $i = 1, \cdots, 13$ and the parameter σ. These estimates are on an interval scale and any common linear transformation of these parameters will give valid estimates. Accordingly, we normalize these values to fall between -1 and 1, where $\psi(s_1) = -1$ corresponds to the (-90%) anti-face and $\psi(s_{13}) = 1$ corresponds to the (90%) original face. The crucial information that we obtain from fitting the data is the possibly non-linear spacing of $\psi(s_2), \cdots, \psi(s_{12})$ between these limits.

The advantage of the non-parametric fit is that it requires no assumptions about the functional form of the mapping $\psi(s)$. A major disadvantage for our purposes is that it is difficult to judge the local slope of the $\psi(s)$ given only discrete estimates $\psi(s_i)$, $i = 1, \cdots, 13$. Our main interest is to determine whether this slope is largest near the average face, consistent with an expanded representation of the face continuum near the average face.

Accordingly, we also fit the data to a parametric family of functions that we refer to as a double power function with three free parameters C, r_-, r_+ in addition to σ. The fitting procedure is very similar to the non-parametric fit except that the values $\psi(s_i)$, $i = 1, \cdots, 13$ must now satisfy the following equations,

$$\begin{aligned} \psi(s) &= \left(\tfrac{s-C}{M-c}\right)^{r_+} \quad & s > C \\ &= -\left(\tfrac{C-s}{C-m}\right)^{r_-} \quad & s \leqslant C. \end{aligned} \tag{2}$$

where m, M denote the minimum and the maximum of the identity strength scale, respectively. The function $\psi(s)$ consists of two power functions joined at a breakpoint C on the identity strength continuum. The power function coefficient for values of s greater than C was denoted r_+ and that for values of s less than C was denoted r_-. Note that the power function on the lower side is "upside-down" because of the minus sign in Eq. (2), second line. We illustrate some of the functional shapes available with this form in Fig. 8. In Fig. 8a, we show the linear fit with $r_+ = r_- = 1$ and $C = (M + m)/2$, the middle of the identity strength scale, i.e., the identity strength of the average face. If observers judge the faces equally spaced in identity strength as equally spaced in super-threshold difference, the parametric estimates will look like those in Fig. 8a.

In Fig. 8b, we plot a function whose slope reaches a maximum at the average face, consistent with an exaggeration of perceived super-threshold difference around the average face ("crispening"). Note that C is located at 0 in this case. Fig. 8c shows the opposite pattern: the slope is

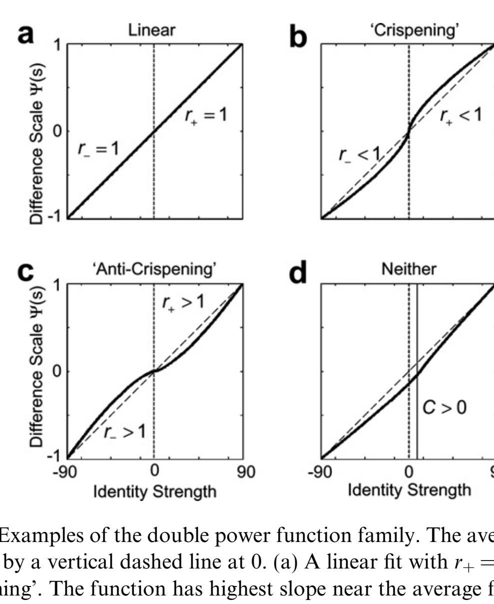

Fig. 8. Examples of the double power function family. The average face is marked by a vertical dashed line at 0. (a) A linear fit with $r_+ = r_- = 1$. (b) 'Crispening'. The function has highest slope near the average face: $r_+ < 1$, $r_- < 1$ and $C = 0$. (c) 'Anti-crispening'. The function has lowest slope near the average face: $r_+ > 1$, $r_- > 1$ and $C = 0$. (d) An alternative fit. The double power function need not be symmetric about the average face. Here $r_+ < 1$, $r_- < 1$ and $C = 15\%$. The slope increases across the continuum from anti-face to original face with the highest slope at the original face not at the average face.

lowest near the average face indicating that the face continuum is contracted near the average. The parametric family contains a range of other possible forms, one of which is shown in Fig. 8d: here the slope increases steadily from the anti-face to the original face and there is again no indication of an exaggerated spacing around the average.

The advantage of the double power family is that we can use it to test hypotheses about slope near the average face. The disadvantage of choosing a parametric family such as the double power function is that we have no reason to think that it captures the true functional form of $\psi(s)$. Consequently, we will report both the parametric and non-parametric fits in summary plots so that the reader can judge whether the conclusions we draw from the parametric fit are consistent with the non-parametric.

4.2.1. Individual fits

We report the fitted parameter values for the double power functions for each participant in Table 1. Two participants were treated as outliers, because of extreme values on r_+ (P10) and sigma (P14), respectively. All analyses were conducted on the remaining 14 participants. As expected, experts and novices did not differ on any of the parameters, all $t(12)$'s < 1.46, ns., so familiarity with the faces will not be considered further. The average breakpoint C occurred at 10.55% identity strength, slightly, but significantly, above the average face (0%), $t(13) = 2.99$, $p < .02$. The coefficient of the lower power function ($M = 0.86$, $SD = -.17$) was significantly less than one, $t(13) = 3.01$, $p < .02$, indicating a significant deviation from linearity. However, it did not differ significantly from the coefficient for the upper power function ($M = 0.96$, $SD = 0.29$), $t(13) = 1.32$, ns., which did not differ significantly from one, $t(13) = 0.49$, ns.

We plot the parametric fit functions for individual subjects in Fig. 9. The key outcome for us concerns the slope at the average face. If subjects perceived greater super-threshold differences among faces near the average (‘‘crispening’’), then we would expect that the fitted function $\psi(s)$ would resemble Fig. 8b. For the majority of subjects this is not the case. Due to the deviation of the break point C away from identity strength 0, many of the individual fits resemble Fig. 8d with the slope of the curve increasing steadily from the anti-face to the original face. These results provide little support for the claim that the slope reaches a maximum at the average face.

4.2.2. Group fit

We next report the analysis with all the individual data modelled as a single ‘‘supersubject’’ (Fig. 10). The fitted double power function is shown as a curve and the corresponding non-parametric fit values are plotted as points. Inspection of Fig. 10 shows that the fitted power functions ($r_+ = 0.94$, $r_- = 0.92$) did not achieve a slope maximum near the average face (0%) or near the breakpoint (=11.96%). The estimated breakpoints and power coefficients closely matched those for the individual fits, justifying this group analysis. Importantly, no evidence for ‘‘crispening’’ was found. Nor was there clear evidence for ‘‘anti-crispening’’. Rather the slope increased steadily from the anti-face to the original face with the steepest slopes for positive identity strengths.

The fitted values of the breakpoint C deviated from the average face and this in turn complicated the interpretation of the remaining parameters since they now signal informa-

Table 1
Parameter values for participants in Study 3

Participant no.	Expert/novice	Face seen	r_+	r_-	Breakpoint C	Sigma error
1	Expert	Ted	0.76	0.95	26.55	0.10
2	Expert	Ted	0.47	0.83	29.55	0.27
3	Expert	Jim	1.00	0.79	−13.50	0.17
4	Expert	Jim	0.92	0.97	−4.50	0.31
5	Expert	Rob	0.92	0.88	12.90	0.11
6	Expert	Rob	0.90	1.07	18.00	0.11
7	Expert	Dan	0.89	0.79	13.05	0.17
8	Expert	Dan	1.08	0.62	12.00	0.33
9	Novice	Jim	1.60	1.25	1.05	0.19
10	Novice	Jim	2.45	0.82	−30.00	0.20
11	Novice	Dan	0.82	0.98	−5.85	0.16
12	Novice	Dan	0.71	0.68	10.35	0.13
13	Novice	Ted	0.76	0.70	26.40	0.26
14	Novice	Ted	0.66	0.53	27.00	0.81
15	Novice	Rob	1.24	0.73	19.80	0.13
16	Novice	Rob	1.41	0.88	1.95	0.13
M[a]			0.96	0.86	10.55	0.19
SD[a]			0.29	0.17	13.19	0.08
N[a]			14	14	14	14

r_+ and r_- are the coefficients of the upper and lower power functions, defined relative to the breakpoint C. Sigma error is a measure of the observer's ability to discriminate between the face pairs.

[a] Outliers (P10, P14) excluded.

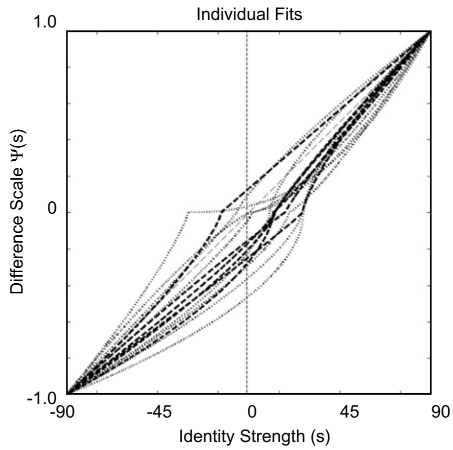

Fig. 9. Perceptual difference scaling results for individual subjects in Study 3. The fits of the double power parametric model are shown for 14 subjects (excluding the two expert subjects classified as outliers—see text). The fits for expert subjects are plotted as dashed lines, those for novice subjects as dotted. The dotted vertical line indicates the location of the average face at the center of the continuum.

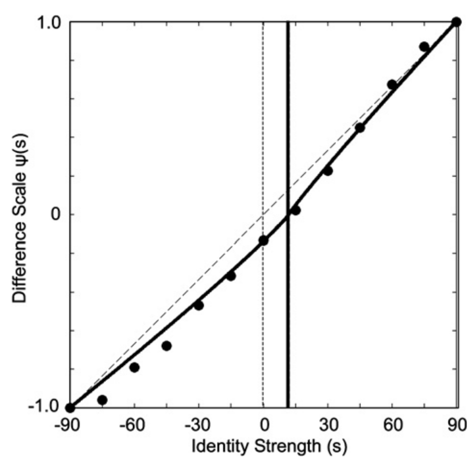

Fig. 10. Perceptual difference scaling results for a single "supersubject" combining data from all of the individual subjects in Fig. 9. Double power functions were fit either side of an empirically derived breakpoint (solid vertical line). The results of fitting the "non-parametric" models are shown as points. The dotted vertical line indicates the location of the average face.

tion about the slope of the function $\psi(s)$ near C away from the average face. One evident result is that the slope of $\psi(s)$ over the range of anti-faces was consistently less than over the range of positive identity strengths (mixtures of the original face and the average face). The anti-faces were perceived as consistently less different from one another than were the remaining faces. This could be due either to an artefact of the morphing procedure used to create anti-faces, or to some real difference between faces and anti-faces, to which participants are sensitive. For example, if anatomical constraints result in skewed distributions along face dimensions, then anti-faces might look a little unusual. In the next study, we will look at a different portion of the face continuum (from the average face to the original face and beyond) excluding the anti-faces and design the fitting procedure so that one parameter unambiguously signals whether the face continuum is expanded or compressed around the average face.

5. Study 4

In this study, we sought converging evidence on sensitivity to differences around the average face using the same super-threshold difference scaling method as in Study 3 for identity continua with the average face (0%) as the lower endpoint. The upper endpoint was a 150% identity strength version of the original face, i.e., a 50% caricature of that face. If perceptual differences are greater around the average face, then the slopes of the fitted functions should be steepest around the average, at the lower end of the continuum. The difference scaling task was the same as in Study 3, and we fit two power functions as before to two segments of the scale, from the average face to a breakpoint C and from the breakpoint C to the right end of the scale corresponding to an identity strength of 150%. The breakpoint itself is estimated from the data as in Study 3. Note that the average face is now at the left end of the scale and we are interested in the slope of the scale at that point.

5.1. Method

5.1.1. Participants

Sixteen (12 female) Caucasian adults (aged 21–49, $M = 25.7$, $SD = 4.9$ years), participated. Eight were familiar with the identities shown (experts) and eight were not (novices). Thirteen (8 experts, 5 novices) had participated in Study 3, but judged a different identity continuum in this study.

5.1.2. Stimuli

We constructed 11-step identity continua for each of the four male faces used in Study 3. Each continuum contained images with the following identity strengths (s): 0, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150 (Fig. 11). As before, an identity strength of 0 corresponds to the average face. Identity strengths less than 100 correspond to anticaricatures, and identity strengths greater than 100 correspond to caricatures, of the original face. Caricatures were made by exaggerating all shape differences between the original face and the average face using Gryphon's Morph. Anticaricatures were made by reducing these differences, effectively warping the shape of the original face towards the average shape. All images had the texture of the original face. The resulting images were sharpened and placed in an oval mask which hid all of the hair, but not the face outline. Images measured 6.2×8.8 cm and were viewed from approximately 60 cm.

5.1.3. Procedure

The procedure was the same as Study 3 except that participants completed 660 trials, consisting of two presenta-

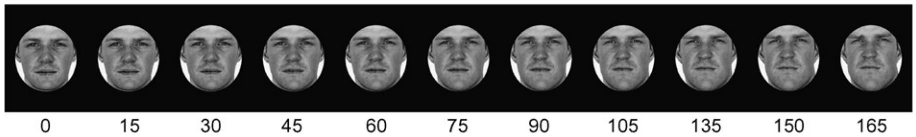

Fig. 11. Example of an 11-face continuum used in Study 4. Identity strengths are shown below the images. 0 = average face, percentages less than 100% = anticaricatures of the target, percentages greater than 100% = caricatures of the target.

tions of all distinct non-overlapping quadruples ($N = 330$) for 11 images (Maloney & Yang, 2003). As before, each participant judged one identity continuum and each identity continuum was judged by two experts and two novices. Each test session lasted approximately 70 min.

5.2. Results and discussion

A function with three free parameters C, r_+, r_0 was fit to each participant's data using the methods of Maloney and Yang (2003). The function is most easily defined in stages. First, the right half of the scale in Study 3 (from the average face to the original face) corresponds to the left half of the scale in Study 4. The results of Study 3 were complicated by subjects' choices of break points away from the average face. This made it difficult to interpret the results directly in terms of the parameter values as discussed above. In this experiment, we will force the first power function to run from the average face upwards toward the original face. We denote the parameter of this power function by r_+ to remind the reader that it characterizes roughly the same region of the face continuum as did r_+ in the previous study. Now, if there is perceptual expansion of the face continuum (perceived super-threshold differences are greater) around the average face we expect that $r_+ < 1$, otherwise $r_+ \geqslant 1$.

As in Study 3, the value C divides the range into an upper and lower part. The lower part runs from the average face with minimum identity strength m to the breakpoint C and corresponds to the upper power function in Study 3 where the average face fell in the center of the range of identity strengths. We add a second power function that runs from the breakpoint to the maximum identity strength $M = 150$. This part of the face continuum was not presented in Study 3 and we denote the parameter of this power function by r_0. In equation form,

$$\begin{aligned} \psi(s) &= 0.5\left[\tfrac{s-m}{C-m}\right]^{r_+} & s < C \\ &= 0.5\left[\tfrac{s-C}{M-C}\right]^{r_0} + 0.5 & s \geqslant C, \end{aligned} \tag{3}$$

where s is, as above, the identity strength and m, M are the minimum and maximum values of the identity strength scale. The value of the parameter r_+ gives an unambiguous test of the hypothesis: a parameter value of $r_+ < 1$ would be consistent with expansion of the face continuum around the average face, a value $r_+ > 1$ would be consistent with contraction. We emphasize that the second power function used in the model with coefficient r_0 and the breakpoint C are not relevant to testing whether perceived super-threshold differences are greater around the average. However, they will allow us to explore the shape of the function $\psi(s)$ in the region near and beyond the original face.

5.2.1. Individual fits

The parameter values for the fitted power functions are shown for each participant in Table 2. Two participants with extreme values on sigma (P4) and the breakpoint (P13) were excluded. All analyses were conducted on the remaining 14 participants. We plot the parametric fits in Fig. 12 in the same format as Fig. 9. As can be seen the left-most parts of the function near the average face tend to be flat or convex, not concave, inconsistent with expansion of the region of the face continuum around the average face. We replot the values of r_+ from Table 1 in Fig. 13. We note that they are typically greater than one, consistent not with expansion of the face continuum around the average face but with compression. The results provided no evidence

Table 2
Parameter values for participants in Study 4

Participant no.	Expert/ novice	Face seen	r_+	r_0	Breakpoint C	Sigma error
1	Expert	Dan	1.13	0.81	75.00	0.19
2	Expert	Dan	1.52	0.36	90.00	0.20
3	Expert	Jim	1.57	1.29	68.85	0.17
4	Expert	Jim	0.41	0.84	85.94	0.99
5	Expert	Rob	1.01	0.53	59.55	0.24
6	Expert	Rob	1.50	1.10	70.05	0.18
7	Expert	Ted	1.76	0.31	75.00	0.35
8	Expert	Ted	1.71	0.53	89.83	0.26
M[a]			1.46	0.70	75.47	0.23
SD[a]			0.28	0.37	11.14	0.06
N[a]			7	7	7	7
9	Novice	Dan	1.06	0.50	60.15	0.47
10	Novice	Dan	1.09	0.51	74.25	0.22
11	Novice	Jim	0.98	0.72	86.40	0.39
12	Novice	Jim	0.90	0.83	88.95	0.43
13	Novice	Rob	2.17	1.62	15.75	0.22
14	Novice	Rob	1.30	0.73	85.20	0.22
15	Novice	Ted	1.25	1.11	66.60	0.34
16	Novice	Ted	1.20	0.36	103.35	0.22
M[b]			1.11	0.68	80.70	0.33
SD[b]			0.15	0.25	14.70	0.11
N[b]			7	7	7	7

r_+ and r_0 are the coefficients of the lower and upper power functions, defined relative to the breakpoint C. The coefficient r_+ corresponds to the same part of the face continuum as did r_+ in Study 3 and Table 1. Values $r_+ < 1$ indicate an expansion of the face continuum around the average face. As can be seen, values of r_+ are typically greater than 1. Sigma error is a measure of the observer's ability to discriminate between the face pairs.

[a] Outlier P4 excluded.

[b] Outlier P13 excluded.

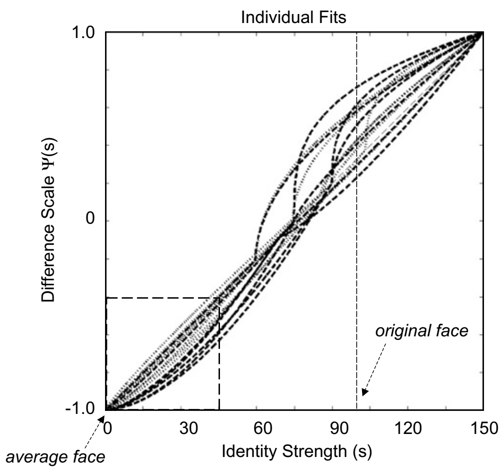

Fig. 12. Perceptual difference scaling results for individual subjects in Study 4. The fits of the double power parametric model are shown for 14 subjects (excluding one expert and one novice subject classified as outliers—see text). The fits for expert subjects are plotted as dashed lines, those for novice subjects as dotted. The dotted vertical line indicates the location of the original face. The average face corresponds to 0 on the horizontal scale. The dashed square on the lower-left of the plot region marks a region of interest. If the region around the average face were perceptually expanded, the parts of the curves enclosed in the dashed square should be concave, above the 45° line. Instead the majority are convex, below the 45° line.

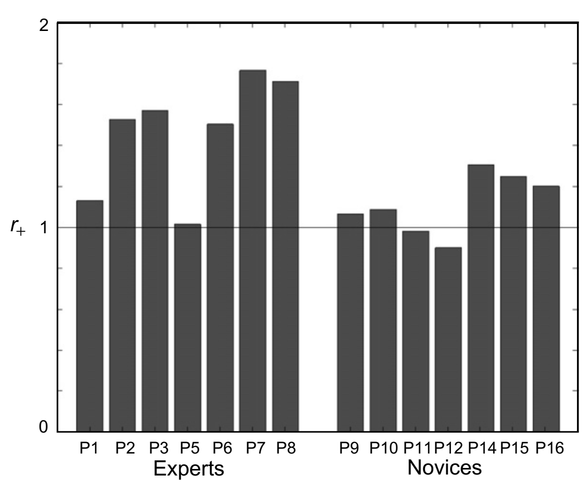

Fig. 13. The values of r_+ for individual subjects. These values (taken from Table 2) are plotted for experts and novices separately. Values for one expert and one novice (classified as outliers) are excluded (see text). Expansion of the face continuum around the average face requires that $r_+ < 1$. The values are, in contrast, typically greater than 1.

for "crispening" around the average face. The lower power coefficients r_+ differed significantly for experts ($M = 1.46$, $SD = 0.28$) and novices ($M = 1.11$, $SD = 0.15$), $t(12) = 2.88$, $p < .02$, but in both cases were greater than 1.0, significantly for experts, $t(6) = 4.29$, $p < .01$, and marginally for non-experts, $t(6) = 2.01$, $p < .10$.

The breakpoints C did not differ significantly for experts ($M = 75.47$, $SD = 11.14$) and novices ($M = 80.70$, $SD = 14.70$), $t(12) = 0.75$, $p = .47$. Overall they were significantly below ($M = 78.15$, $SD = 12.75$) the original face (100), $t(13) = 6.42$, $p < .0001$, indicating that the upper power functions spanned the original face. The power coefficients r_0 for these upper functions ($M = 0.69$, $SD = 0.31$) were significantly less than 1.0, $t(13) = 3.74$, $p < .01$, indicating that the steepest slope was immediately to the right of the break point, i.e., somewhat to the left of the original face, qualitatively consistent with what we found in Study 3. These power coefficients did not differ significantly for experts ($M = 0.70$, $SD = 0.37$) and novices ($M = 0.68$, $SD = 0.25$), $t(12) = 0$, $p = .89$. Sigma error was marginally lower for experts ($M = 0.23$, $SD = 0.06$) than novices ($M = 0.33$, $SD = 0.11$), $t(12) = 2.09$, $p < .06$, indicating that participants who knew the identities were better able to discriminate the pairs.

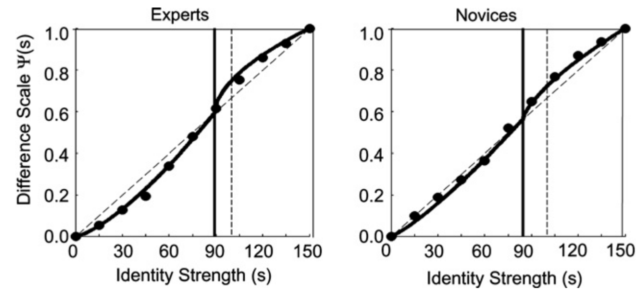

Fig. 14. Perceptual difference scaling results for a single "expert supersubject" and a single "novice supersubject" in Study 4. Double power functions were fit either side of an empirically derived breakpoint (solid vertical line). "Non-parametric" scale values are shown as points. The dotted vertical line indicates the location of the original target face.

5.2.2. Group fit

We also modelled all the individual data from experts and from novices, respectively, as single expert and novice "supersubjects" (Fig. 14). For experts the parameters of the power functions were: $r_0 = 0.56$, $r_+ = 1.39$, $C = 89.45\%$, and for novices they were: $r_0 = 0.70$, $r_+ = 1.22$, breakpoint = 84.78%. Inspection of Fig. 14 clearly shows the pattern described above for individual participants, with relatively flat functions around the average face (0) and steepest slopes just left of the original face (100). These effects are strongest in the expert data, but can also be seen in the novice data. They suggest that participants are more sensitive to differences in the vicinity of normal faces than differences in the vicinity of the average face. As above, the non-parametric fit is shown as discrete points on Fig. 14. They fall very close to the power functions, confirming their goodness of fit.

6. General discussion

Converging evidence from three different paradigms failed to support enhanced discrimination around the average. Discrimination of small interocular differences did not vary as a function of distance from the average (Study 1). Rated similarity was not reduced close to the average, whether or not pairs spanned the average (Study 2). On the contrary, similarity was highest for face pairs close to the average. Nor were perceptual differences amplified around the average (Studies 3 and 4), with larger perceptual differences around the original faces than the average.

Wilson et al. (2002) found reduced discrimination thresholds around the average for synthetic radial frequency faces. The results of Study 1 suggest that this may not be the case for more realistic face images. Alternatively, the discrepancy in results could reflect the many procedural differences between their study and ours. One difference is that we always measured discrimination along axes originating from the average face, whereas they did not. However, it seems unlikely that this difference explains the discrepancy because axis direction did not affect discrimination (at least for faces close to the average) in Wilson et al. (2002). It is possible that we would have replicated their results had we measured thresholds, although again this seems unlikely given that performance in our 2AFC interocular spacing task was effectively at 75%-threshold. We might also have found reduced thresholds around the average had we used a more global difference than interocular spacing, although our super-threshold discrimination results for pairs that differed globally provides little reason to think so. Clearly, more work is needed to determine what factors affect face discrimination thresholds.

Normal face perception requires sensitivity to numerous differences that signal important social information, such as identity, emotional expression, direction of attention, age, sex, race, kinship, attractiveness, etc. Although some of these differences may be subtle, most are not close to discrimination thresholds. Therefore, our super-threshold discrimination results (Studies 2–4) may be more relevant to understanding the function of adaptive face coding mechanisms. Those results were clear. Sensitivity was worse, not better, for differences close to the mean. These results offer no evidence that face adaptation functions to highlight the differences between the faces we encounter most often or most recently.

Although face adaptation does not selectively increase sensitivity to differences close to the average, it may nevertheless help us identify the faces of those around us. Watson and colleagues (2006) found that adapting to a morphed average of faces from a given race enhances identification of faces of that race compared with faces of another race. Therefore, face adaptation may function to match the dynamic coding range to the current diet of faces, without selectively enhancing the perception of differences close to the central tendency of that population.

Adaptive norm-based coding of identity is thought to ensure that the most distinctive information, which is most diagnostic of identity, is encoded (Leopold et al., 2001; Rhodes & Jeffery, 2006). Both single cell recordings in monkeys and fMRI studies in humans show that neural activity in face-sensitive cortex increases with distance from the average (i.e., distinctiveness) in face space, consistent with this model (Leopold, Bondar, & Giese, 2006; Loffler et al., 2005). By matching the dynamic response range to the current input, adaptation may ensure that the central tendency of the current diet of faces is used to code identity. It would also ensure that neural activity is minimized, with least responding to the most common inputs.[2]

Many studies have shown that typical faces, which lie close to the average, are harder to recognize than more distinctive faces (Valentine, 1991; Valentine & Bruce, 1986). Valentine has attributed this effect to the greater local density of distracters close to the average. Our results suggest that it could also reflect reduced sensitivity to perceptual differences close to the average. It is perhaps possible that reduced sensitivity around the average is itself the result of high density around the average. However, the only geometric model we know of which relates perceived similarity to local density proposes the opposite relation, namely that two points in a dense region of space will look less similar than two equidistant points in a less dense region (Krumhansl, 1978).

Categorical perception of familiar identities is well established (Beale & Keil, 1995), with better discrimination of face pairs, taken from opposite sides of an identity boundary (normally the midpoint of a faceA–faceB continuum), than for equidistant pairs from the same side of the boundary. Although not always found, categorical perception has been reported for unfamiliar identities (Campanella, Hanoteau, Seron, Joassin, & Bruyer, 2003). We would, therefore, expect face pairs that span the average to look less similar than other equidistant pairs. However, we found the opposite in Study 2, with greater similarity for pairs spanning the average. Our result raises the possibility that reduced sensitivity to differences around the average can reverse the enhanced discrimination across the category boundary normally seen with categorical perception. Future studies could test this hypothesis by comparing categorical perception for unfamiliar face pairs taken from identity continua that do and do not cross the average.

We found no evidence that adaptive face coding enhances discrimination around the average face. Yet for simpler stimulus attributes, such as orientation, contrast, and direction of motion, discrimination thresholds are sometimes slightly reduced around the adapted (average) state (e.g., Abbonizio et al., 2002; Clifford et al., 2001; Phinney et al., 1997). An important difference may be that in those domains stimuli are more uniformly distributed than are faces in face space. If face attributes are normally distributed on the dimensions of face space (Valentine, 1991), then the greater density of faces around the average could offset any improvement in discrimination resulting from adaptation. This observation raises the possibility that enhanced discrimination occurs for synthetic radial

[2] Although distinctive faces, en masse, are more common than typical faces, there are more faces close to the average face than to any other point in face space, assuming a multivariate normal distribution (Burton & Vokey, 1998).

frequency faces (Wilson et al., 2002) because these form a distinct face space, with relatively few exemplars and less crowding around the average to mask enhanced discrimination there.

In summary, we found no evidence that adaptive face coding increases sensitivity to differences around the average face. However, adapting to an average face can enhance identification of faces of the same race compared with faces from another race (Watson, Rhodes, & Clifford, 2006). Taken together, these results suggest that adaptive face coding may function to calibrate a limited dynamic range to prevailing stimuli, rather than to enhance discrimination close to the center of that range (cf Barlow, 1990). Other important functions of face adaptation may be to conserve neural responses (Leopold et al., 2006), to enable computationally efficient coding of identity (Leopold et al., 2001, 2006; Rhodes & Jeffery, 2006) and to direct our attention to faces with unexpected or unusual properties (Webster et al., 2005).

Acknowledgments

This work was supported by the Australian Research Council and the National Institutes of Health (EY08266). We thank Linda Jeffery for making the stimuli for Studies 3 and 4 and for helpful comments on the manuscript.

References

Abbonizio, G., Langley, K., & Clifford, C. W. G. (2002). Contrast adaptation may enhance contrast discrimination. *Spatial Vision, 16*, 45–58.

Anderson, N. D., & Wilson, H. R. (2005). The nature of synthetic face adaptation. *Vision Research, 45*, 1815–1828.

Barlow, H. B. (1990). A theory about the functional role and synaptic mechanism of visual after-effects. In C. Blakemore (Ed.), *Vision: Coding and efficiency*. Cambridge: Cambridge University Press.

Barlow, H. B., MacLeod, D. I. A., & Van Meeteren, A. (1976). Adaptation to gratings: no compensatory advantages found. *Vision Research, 16*, 1043–1045.

Beale, J. M., & Keil, C. F. (1995). Categorical effects in the perception of faces. *Cognition, 57*, 217–239.

Blanz, V., O'Toole, A. J., Vetter, T., & Wild, H. A. (2000). On the other side of the mean: the perception of dissimilarity in human faces. *Perception, 29*, 885–891.

Burton, A. M., & Vokey, J. R. (1998). The face-space typicality paradox: understanding the face-space metaphor. *Quarterly Journal of Experimental Psychology, 51A*, 475–483.

Campanella, S., Hanoteau, C., Seron, X., Joassin, F., & Bruyer, R. (2003). Categorical perception of unfamiliar facial identities, the face-space metaphor, and the morphing technique. *Visual Cognition, 10*(2), 129–156.

Carey, S. (1996). Perceptual classification and expertise. In R. Gelman & T. K. Au (Eds.), *Perceptual and cognitive development: Handbook of perception and cognition* (2nd ed., pp. 49–69). San Diego, CA: Academic Press.

Clifford, C. W. G. (2002). Perceptual adaptation: motion parallels orientation. *Trends in Cognitive Sciences, 6*, 136–143.

Clifford, C. W. G., Ma Wyatt, A., Arnold, D. H., Smith, S. T., & Wenderoth, P. (2001). Orthogonal adaptation improves orientation discrimination. *Vision Research, 41*, 151–159.

Clifford, C. W. G., & Rhodes, G. (Eds.). (2005). *Fitting the mind to the world: Adaptation and after-effects in high-level vision*. Oxford: Oxford University Press.

Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *Journal of Psychology, 105*, 47–59.

Green, D. M., & Swets, J. A. (1966/1974). *Signal detection and psychophysics*. New York: Wiley.

Hebb, D. O. (1949). *The organisation of behaviour*. New York: Wiley.

Hochberg, J. E. (1978). *Perception* (2nd ed.). Englewood Cliffs, New Jersey: Prentice-Hall.

Hol, K., & Treue, S. (2001). Different populations of neurons contribute to the detection and discrimination of visual motion. *Vision Research, 41*, 685–689.

Hurlbert, A. (2001). Trading faces. *Nature Neuroscience, 4*, 3–5.

Krekelberg, B., Van Wezel, R. J. A., & Albright, T. D. (2006). Adaptation in macaque MT reduces perceived speed and improves speed discrimination. *Journal of Neurophysiology, 95*, 255–270.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. *Psychological Review, 85*, 445–463.

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness and identification: testing the face-space framework. *Psychological Science, 11*, 379–385.

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

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

Leopold, D. A., Rhodes, G., Müller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society of London, Series B, 272*, 897–904.

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

Maattanen, L. M., & Koenderink, J. J. (1991). Contrast adaptation and contrast gain control. *Experimental Brain Research, 87*, 205–212.

MacLin, O. H., & Webster, M. A. (2001). Influence of adaptation on the perception of distortions in natural images. *Journal of Electronic Imaging, 10*, 100–109.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision, 3*, 573–585.

Mather, G., & Harris, J. (1998). Theoretical models of the motion after-effect. In G. Mather, F. Verstraten, & S. Anstis (Eds.) (pp. 157–185). Cambridge, Massachusetts: MIT Press.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processes. *Trends in Cognitive Sciences, 6*, 255–260.

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception & Performance, 27*, 573–599.

Phinney, R. E., Bowd, C., & Patterson, R. (1997). Direction-selective coding of stereoscopic (cyclopean) motion. *Vision Research, 37*, 865–869.

Regan, D., & Beverley, K. I. (1985). Postadaptation orientation discrimination. *Journal of the Optical Society of America A, 2*, 147–155.

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. *Perception, 17*, 48–63.

Rhodes, G. (1996). *Superportraits: Caricatures and recognition*. Hove: The Psychology Press.

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

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

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: face adaptation and attractiveness aftereffects. *Psychological Science, 14*, 558–566.

Rhodes, G., Jeffery, L., Watson, T., Jaquet, E., Winkler, C., & Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face coding mechanisms. *Current Biology, 14*, 2119–2123.

Rhodes, G., Robbins, R., Jaquet, E., McKone, E., Jeffery, L., & Clifford, C. W. G. (2005). Adaptation and face perception: How aftereffects implicate norm-based coding of faces. In C. W. G. Clifford & G. Rhodes (Eds.). *Fitting the mind to the world: Adaptation and aftereffects in high-level vision*. Oxford: Oxford University Press.

Rhodes, G., Sumich, A., & Byatt, G. (1999). Are average facial configurations attractive only because of their symmetry? *Psychological Science, 10*, 52–58.

Tsao, D. Y., & Freiwald, W. A. (2006). What's so special about the average face? *Trends in Cognitive Sciences, 10*, 391–393.

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

Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception, 15*, 525–535.

Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. *Cognitive Science, 28*, 289–301.

Watson, T. L., & Clifford, C. W. G. (2003). Pulling faces: an investigation of the face-distortion aftereffect. *Perception, 32*, 1109–1116.

Watson, T. L., Rhodes, G., & Clifford, C. W. G. (2006). Norm-based coding of faces adapts to help us identify those around us. Poster. *Vision Sciences Society, Florida*, 2006.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural face categories. *Nature, 428*, 557–560.

Webster, M. A., Werner, J. S., & Field, D. J. (2005). Adaptation and the phenomenology of perception. In C. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and aftereffects in high-level vision*. Oxford: Oxford University Press.

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review, 6*, 647–653.

Werblin, F. S. (1973). The control of sensitivity in the retina. *Scientific American, 228*, 70–79.

Whittle, P. (1992). Brightness, discriminability and "Crispening Effect". *Vision Research, 32*, 1493–1507.

Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research, 42*, 2909–2923.

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