



# What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation

Tirta Susilo\*, Elinor McKone, Mark Edwards

Department of Psychology, Australian National University, Canberra, ACT 0200, Australia

## ARTICLE INFO

### Article history:

Received 26 August 2009

Received in revised form 17 November 2009

### Keywords:

Face perception

Face space

Norm-based coding

Face discrimination

Adaptation aftereffect

Tuning curves

## ABSTRACT

Recent evidence has shown that face space represents facial identity information using two-pool opponent coding. Here we ask whether the shape of the monotonic neural response functions underlying such coding is linear (i.e. face space codes all equal-sized physical changes with equal sensitivity) or nonlinear (e.g. face space shows greater coding sensitivity around the average face). Using adaptation aftereffects and pairwise discrimination tasks, our results for face attributes of eye height and mouth height demonstrate linear shape; including for bizarre faces far outside the normal range. We discuss how linear coding explains some results in the previous literature, including failures to find that adaptation enhances face discrimination, and suggest possible reasons why face space can maintain detailed coding of values far outside the normal range. We also discuss specific nonlinear coding models needed to explain other findings, and conclude face space appears to use a mixture of linear and nonlinear representations.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

The present study concerns the shape of the neural response curves underlying two-pool opponent coding in face space. In face space, an individual face is coded in terms of a smaller number of underlying dimensions, and the centre of the space represents the average face (Valentine, 1991). Two versions of face space have been traditionally proposed: an exemplar-based version in which individual faces are coded by their absolute dimensional values (Valentine, 1991; Valentine & Endo, 1992), and a norm-based version in which the average face has special status as the norm and individual faces are coded in terms of their deviation from this norm (Rhodes, Brennan, & Carey, 1987; Valentine, 1991). Recently, it has been recognised that norm-based coding corresponds to two-pool opponent coding at a neural level while exemplar-based coding corresponds to multichannel coding, as illustrated in Fig. 1 (Calder, Jenkins, Cassel, & Clifford, 2008; Rhodes & Jeffery, 2006; Robbins, McKone, & Edwards, 2007).

Testing the predictions of these neural models has led to conclusive evidence in favour of two-pool opponent coding.<sup>1</sup> In human adaptation studies, adapting to the average face does not

shift perception of nonaverage faces (Leopold, O'Toole, Vetter, & Blanz, 2001; Webster & MacLin, 1999), perceptual aftereffects are larger for identities opposite in face space than for equally dissimilar nonopposite identities (Leopold, O'Toole, Vetter, & Blanz, 2001; Rhodes & Jeffery, 2006), and aftereffects become larger for adaptors with more extreme distortion levels (Robbins et al., 2007). In concert, monkey studies have found that most face-selective neurons in temporal cortex show monotonic tuning curves: the average value of a facial attribute produces an intermediate level of response, and attribute values in one direction average (e.g., eyes shifted up) increase response while attribute values in the other direction (e.g., eyes shifted down) decrease response (Freiwald, Tsao, & Livingstone, 2009; Leopold, Bondar, & Giese, 2006). Current knowledge regarding the two-pool opponent coding leaves open the important question of the shape of the opponent response functions. Here we ask whether these are *linear* or *nonlinear* (S-shaped), as illustrated in Fig. 2.

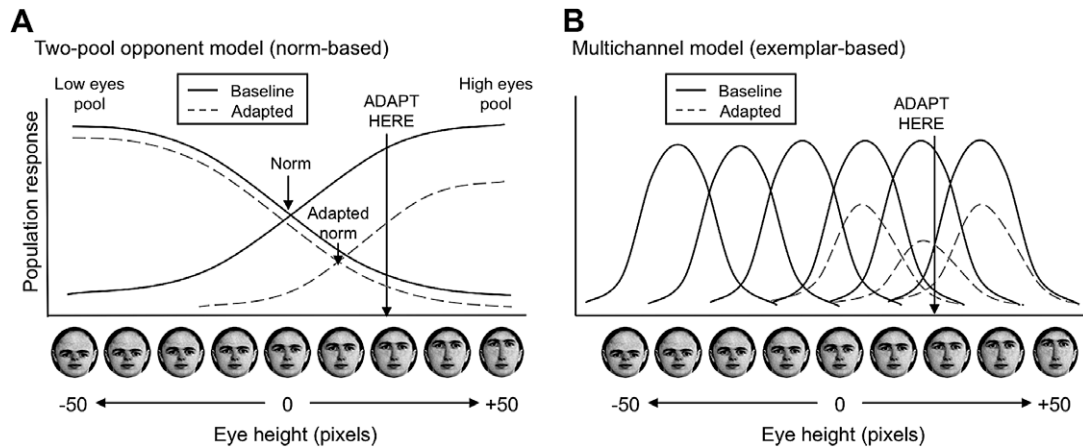
In the *linear*<sup>2</sup> model (Fig. 2A), a facial attribute such as eye height is coded by two opposing response functions that are each straight lines, together forming an "X" shape. The key property of this model is that a unit increment in attribute value (e.g., an increase in eye height of 1 pixel) will cause a constant change in the relative activations of the two pools, regardless of whether the starting value of the attribute is close to the norm (e.g., a 1-pixel change from 0 pixel

\* Corresponding author. Address: Department of Psychology, Australian National University, Canberra, ACT 0200, Australia. Fax: +61 2 6125 0499.

E-mail address: [tirta.susilo@anu.edu.au](mailto:tirta.susilo@anu.edu.au) (T. Susilo).

<sup>1</sup> This is true only for coding of facial identity. Multichannel coding has been revealed for eye gaze (Calder et al., 2008; Jenkins, Beaver, & Calder, 2006), face view in 3D (Fang & He, 2005), and body view in 3D (Lawson, Clifford, & Calder, 2009).

<sup>2</sup> Throughout this article, the term *linear* is used in its strict mathematical sense of describing a function which is a *straight line*.



**Fig. 1.** Illustrations of two-pool opponent (norm-based) coding and multichannel (exemplar-based) coding, using the facial attribute of eye height as an example. (A) Two-pool opponent coding, in which physical eye height values are coded by the relative activation of two neural populations that are oppositely tuned: each shows monotonically increasing or decreasing firing rate in response to attribute values ranging from one extreme end of the range (e.g., eyes down; negative) to the other (eyes-up; positive). (B) Multichannel model in which eye heights are coded by several neural populations with bell-shaped tuning.

deviation to +1 pixel deviation), or whether it is further from the norm (e.g., a 1-pixel change from +50 pixels to +51 pixels).<sup>3</sup> This property carries the corollary that perceptual sensitivity to changes of a fixed magnitude should be equally good regardless of whether these changes are close to or further from the average.

In the S-shaped model (Fig. 2B), a facial attribute is coded by two opposing response functions that are each steep for values close to the norm – and steepest at the crossover point of the two response functions – and shallower for values further from the norm in either direction. In this model, a unit change in attribute value that begins closer to the average face will cause a greater change in the relative activations of the two pools than will a unit change in attribute value that begins further from the average face. The corollary is that perceptual sensitivity will be best for changes around the average value, and poorer around values further from the average.

No authors to date have explicitly put forward the linear and S-shaped models in theoretical contrast to one another. The literature contains schematic illustrations of both models, in explaining the differences between two-pool opponent and multichannel coding mechanisms (Calder et al., 2008; Ng, Boynton, & Fine, 2008; Rhodes & Jeffery, 2006; Robbins et al., 2007; Tsao & Freiwald, 2006). However, none of these authors stated a rationale for drawing the response curves as either linear or nonlinear, and no authors have explicitly drawn out the predictions of the linear and S-shaped models and tested them experimentally. Several studies have offered relevant empirical data and considered theoretical implications for face space coding, but the collective results are mixed and complex to interpret (Dakin & Omgie, 2009; Freiwald et al., 2009; Leopold et al., 2006; Ng et al., 2008; Rhodes, Maloney, Turner, & Ewing, 2007; Tanaka & Corneille, 2007; Wilson, Loffler, & Wilkinson, 2002). They will be discussed in the Section 6.

Here we sought to discriminate between the linear and S-shaped models using two psychophysical methods: adaptation aftereffects and pairwise discrimination tasks. In low-level vision, adaptation aftereffects have been dubbed the “psychophysicist’s microelectrode” (Frisby, 1979) because of its extensive use in

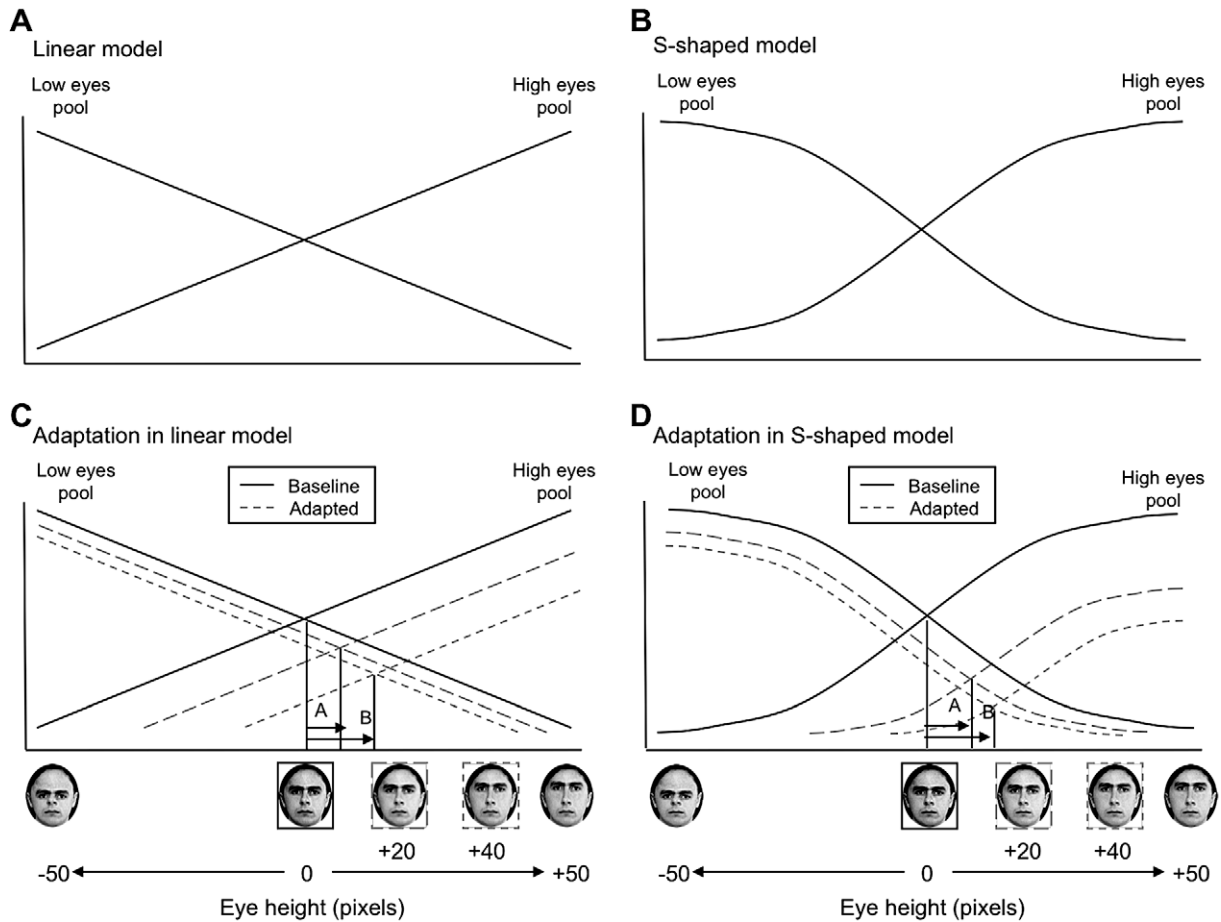
determining the tuning properties of neurons. More recent studies have shown that the mapping between neural tuning and perceptual aftereffects extends to mid- and high-level vision. Specifically, the tuning shape of V4 neurons obtained with single-cell recording in monkeys (Muller, Wilke, & Leopold, 2009) is consistent with that deduced from shape-specific aftereffects in humans (Suzuki, 2005).

In our adaptation task, we measured the aftereffect magnitude (the shift in the physical attribute value that observers perceived as most normal following adaptation) as a function of multiple adaptor positions varying in distance from the norm. As the physical difference between the adaptor and the norm is increased, the linear model predicts a linear increase in aftereffect magnitudes (Fig. 2C), whereas the S-shaped model predicts a less-than-linear increase in aftereffect magnitudes (Fig. 2D). Fig. 3A shows the predicted results of the two models.

In the pairwise discrimination task, we measured observers’ ability to discriminate two faces differing by a constant physical amount (e.g., five pixels of eye height) at multiple starting positions in the attribute continuum (e.g., –25 pixels, +10 pixels). The linear model predicts that perceptual sensitivity to changes of a fixed magnitude should not depend on the starting position, so that a plot of discrimination accuracy against starting position should produce a flat line. In contrast, the S-shaped model predicts greatest perceptual sensitivity to changes around the average value, so that a plot of discrimination accuracy against starting position will produce a function that peaks around zero and falls off for values away from zero in either positive or negative directions. These predicted results are shown in Fig. 3B.

Another aim of the present study was to examine the *range* of values of a given attribute over which a tuning model held. This was important for two reasons. First, the distinction between the linear and the S-shaped models can only be ascertained after testing exhaustively for all locations in the continuum, because the S-shaped model could look linear if one did not test values sufficiently far from the average (see Fig. 2A and B). Second, evidence from adaptation studies suggests that face space codes even very *extreme* values, far outside the normal range seen in everyday natural faces. Bizarre and unnatural distortions of face shape have been shown to produce strong aftereffects on perception of the average face (Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Robbins et al., 2007; Watson & Clifford, 2003; Webster & MacLin, 1999; Yamashita, Hardy, DeValois, & Webster, 2005; Zhao & Chubb, 2001), arguing for at least some sensitivity of face space coding for values beyond the normal range. The S-shaped model

<sup>3</sup> There is a technical assumption here: the readout from the two pools is determined by the *difference* of their outputs (i.e., Pool 1 – Pool 2), rather than the *ratio* (i.e., Pool 1/Pool 2). Only difference-readout produces sensitivity that is symmetric around the average; ratio-readout predicts dramatically stronger sensitivity to changes on one side of the average (e.g., eyes down) than on the other side (e.g., eyes up).

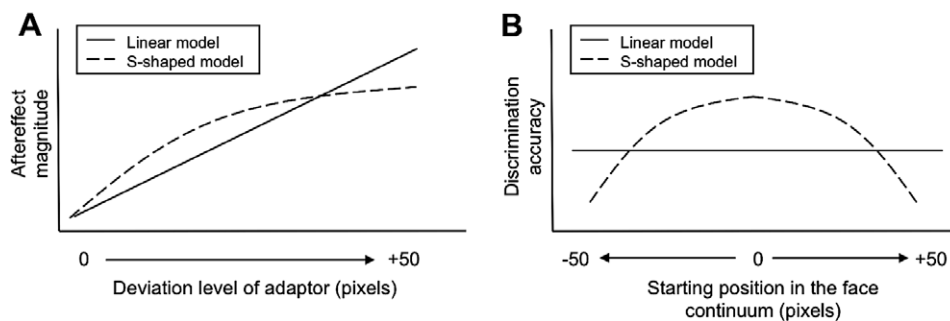


**Fig. 2.** The two tuning models we contrasted in the present study. (A) The *linear model*. (B) The *S-shaped model*. (C) The linear model predicts that, as the distance of an adaptor is shifted twice as far from the original norm (physically average face), the crossover point of the two pools (i.e., the new perceived norm) will shift by twice as much (i.e., B equals twice A); thus, adaptors that are evenly spaced away from zero will produce a linear increase in aftereffect magnitude. (D) The S-shaped model predicts that, as the distance of the adaptor is shifted twice as far from the original norm, the crossover point of the two pools will shift by less than twice as much (i.e., B is less than twice A); thus, adaptors that are evenly spaced away from zero will produce a less-than-linear increase in aftereffect magnitude (technically, a function that is monotonically increasing and concave down). Note all predictions assume that adaptation reduces firing rate in proportion to the initial activation level of that pool by that stimulus (Maddess, McCourt, Blakeslee, & Cunningham, 1988; Movshon & Lennie, 1979).

predicts that discrimination sensitivity should be weaker for bizarre than for near-average faces, while the linear model predicts that discrimination sensitivity would be equal for bizarre and near-average faces.

The facial attributes we manipulated were eye height and mouth height. They were selected because each has a clear physi-

cal range: the eyes cannot be higher than the hairline or lower than the nose, and the mouth cannot be higher than the nose or lower than the chin. These facts meant we were able to clearly define, and cover, the minima and the maxima of the two attribute continua. We first report aftereffect magnitudes as a function of multiple adaptor positions, for eye height (Experiment 1), and then



**Fig. 3.** Predictions of the linear and S-shaped models in our experiments. (A) Adaptation aftereffects experiments. Plotting aftereffect magnitudes as a function of adaptor position (deviation of the adaptor from the norm), the linear model predicts a linear increase while the S-shaped model predicts a nonlinear increase that is monotonic and concave down (see Fig. 2 for reasoning). (B) Pairwise discrimination experiments. Plotting discrimination accuracy for a fixed-magnitude difference (e.g., five pixels) as a function of starting position in the attribute continuum, the linear model predicts a flat line, and the S-shaped model predicts a peak around zero (see main text for reasoning).

mouth height (Experiment 2). We then report pairwise discrimination ability as a function of starting position from the average, for eye height (Experiment 3), and mouth height (Experiment 4).

## 2. Experiment 1: adaptor-position effects (eye height)

Experiment 1 measured aftereffect magnitude following adaptation to seven different adaptor positions spanning the eye height continuum from average to extremely high. We tested four observers and analysed their results individually (because pilot testing showed considerable variation in observers' subjective judgement of what is normal, and in their sensitivity to adaptation). Each 1-h session tested baseline perception of eye height as well as perception following adaptation to one adaptor position. The time of 7 h per observer limited us to testing eyes-up adaptors (a previous study found no difference between aftereffect patterns for eyes-up and eyes-down adaptors; Robbins et al., 2007). Test faces varied across 29 eye height values (both up and down relative to the average) and resulting psychometric curves for "too high versus too low" decision were used to determine the eye height which the observer perceived to be average. To avoid potential low-level after-effects, we used images of different people as adaptor and test faces, as well as smaller image size of adaptors (Anderson & Wilson, 2005; Rhodes et al., 2004; Robbins et al., 2007; Watson & Clifford, 2006; Zhao & Chubb, 2001).

### 2.1. Method

All experiments were approved by the Human Research Ethics Committee of the Australian National University.

#### 2.1.1. Observers

Four observers (two female, age 23–30) were psychophysically experienced, right-handed, with normal or corrected-to-normal vision. Observers TG, TA and OM were naïve, paid \$70 for 7 h of testing, and Caucasian (same race as the face stimuli); TS (the first author) was not naïve and is East Asian.<sup>4</sup>

#### 2.1.2. Design

Each observer was tested in seven separate sessions, one for each of the seven adaptor positions (eyes shifted up by +5, +12, +20, +27, +35, +42, and +50 pixels) in random order. Each session tested a baseline phase first, followed by an adapted phase. There was a gap of at least 24 h between any two sessions to avoid carry-over of adaptation from the previous session (Robbins et al., 2007). The task was to decide whether the eyes were 'too high' or 'too low', based on comparison with their imagined average eye height for real-world faces.

#### 2.1.3. Stimuli

Stimuli were created from nine front view neutral expression Caucasian male individuals (seven from the Stirling PICS database (<http://pics.psych.stir.ac.uk/>) and two from the Harvard Face Database, F. Tong and K. Nakayama). One individual with the hairline location clearly visible was used as a background head; his internal features were ablated and replaced with the internal features of each of the eight other individuals in turn (copied and pasted as a block with original spacing maintained between eyes, nose and mouth) (McKone, Aitkin, & Edwards, 2005; Robbins et al., 2007). Out of the eight resulting individuals (shown in Fig. 4A in their

"zero"-deviation-level version), four were used as adaptors (for all subjects), and four as test faces (again for all subjects). The "zero" version of each face had been found in other testing in our lab to be the version which observers typically perceived as having the most normal eye height.

Deviation levels for each face were made by moving the eyes-up (+) or down (–) using Adobe Photoshop CS2 (note that shifting the eyes also changed the apparent length of the nose). For each of the four adaptor individuals, seven altered versions were created with the eyes shifted up by +5, +12, +20, +27, +35, +42, and +50 pixels. For each of the four test individuals, 29 versions were used with eyes shifted up and down by 0, ±1, ±2, ±3, ±4, ±5, ±6, ±7, ±8, ±9, ±10, ±12, ±14, ±18, and ±24 pixels (Fig. 4B). All manipulations were made on an image sized 370 (vertical) × 310 (horizontal) pixels. One pixel corresponded to 0.29% of the full head height (i.e., top of head to chin).

Stimuli were presented using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993) on a CRT screen iMac computer (36 cm screen, 1024 × 768 resolution). Observers viewed the screen at 40 cm with a chin rest. For presentation, adaptor faces were resized to 227 × 190 pixels (viewing angle 7.9° vertical by 5.7° horizontal), and test faces to 298 × 250 pixels (10° by 7.9°).

#### 2.1.4. Procedure

Each session started with the baseline phase. Each trial presented a test face in the centre of the screen for 250 ms, followed by a blank screen with the question "Were the eyes too high or too low?" There was no fixation point. Observers responded 'too high' with button 'z', or 'too low' with keypad '3'; this response arrangement was reversed for half the observers. The next trial followed 400 ms after response. Each block of 116 trials presented, in random order, each eye height level once for each of the four test faces (29 eye height levels × 4 test-face identities). There were 3 blocks (total of 348 trials), with a 1-min break between blocks. Collapsing across blocks and the test-face identity, each observer's %-too-high score at each eye height level was based on 12 trials.

The adapted phase procedure was identical except that each adapted trial started with the adaptor presented in the centre of the screen for 4000 ms, then a 400 ms blank screen before the test face. To ensure observers could easily distinguish adaptors and test faces, adaptors were outlined by a black rectangle. A practice block containing 20 trials (randomly-chosen eye heights and test individuals) preceded each testing session.

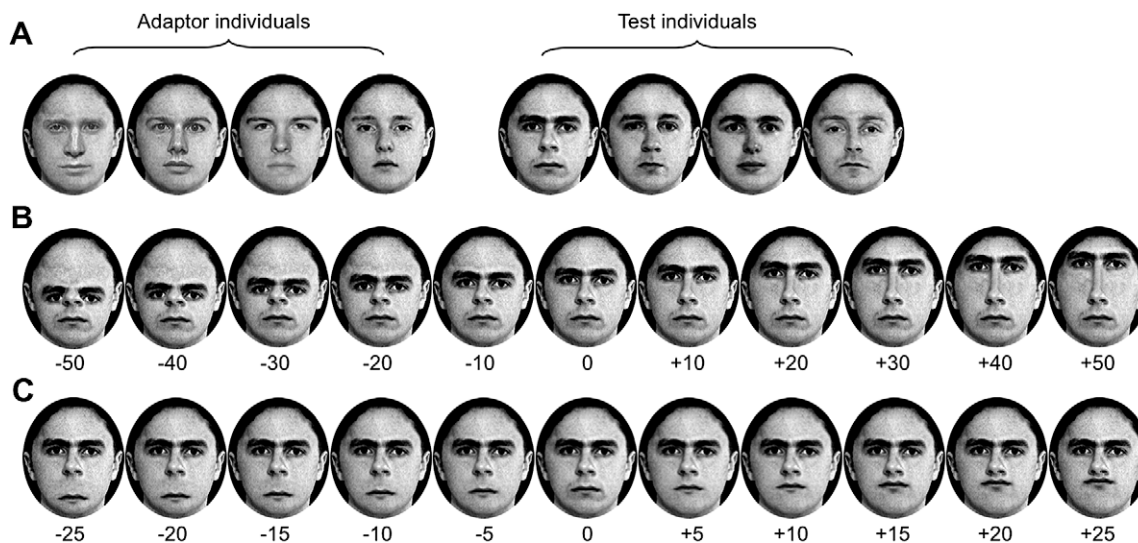
#### 2.1.5. Data analysis

Psychometric curves (Fig. 5A) were fitted using the logistic function in psignifit version 2.5.6 (<http://bootstrap-software.org/psignifit>) (Wichmann & Hill, 2001a). All fits produced  $R^2 > .93$ . The point of subjective equality (PSE) – which reflects the eye height perceived as being the most average – was given by the eye height level corresponding to 50% 'too high' responses. Standard error for the PSE was estimated by the 68% confidence interval following the bootstrap procedure with 2000 Monte Carlo replications (Wichmann & Hill, 2001b). Aftereffect magnitude was calculated by taking the difference between baseline and adapted PSEs within each session (Fig. 5B). A positive score indicated that the shift was in the direction corresponding to an adaptation aftereffect (the PSE should move toward the adaptor, because adaptation to an eyes-up face will make a somewhat eyes-up face now seem normal). Standard error for the aftereffect magnitude was computed as: the square root of [(the squared baseline PSE standard error) plus (the squared adapted PSE standard error)].

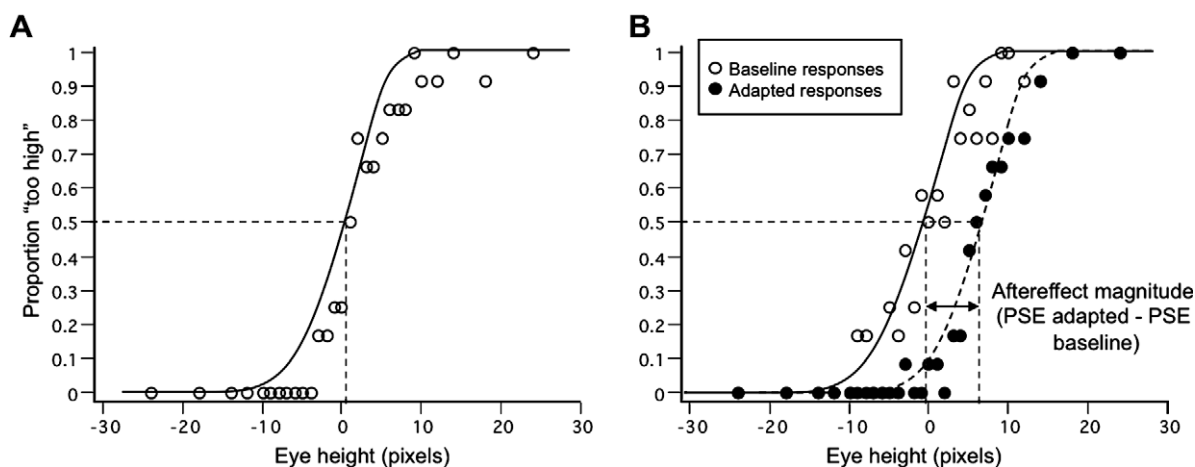
To gain the clearest indication of the relationship between aftereffect magnitude and adaptor position relative to the baseline norm, we plotted aftereffect magnitude as a function of *perceptual*

<sup>4</sup> Importantly, TS demonstrates good processing of Caucasians faces – obtaining better than average performance ( $z = 1.61$ ) on the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), and average performance ( $z = -0.59$ ) as well as normal inversion effect ( $z = -0.03$ ) on the Cambridge Face Perception Test (Duchaine, Germine, & Nakayama, 2007).





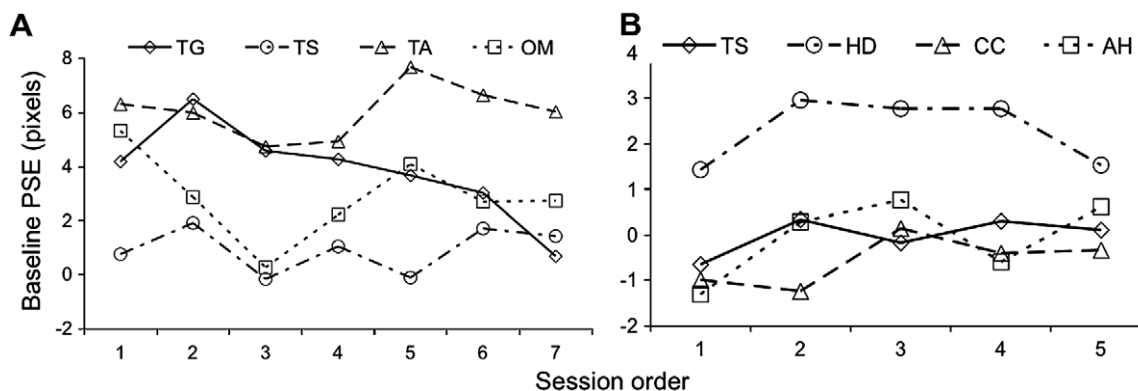
**Fig. 4.** Stimuli used in the present study. (A) The zero-deviation versions of the four individuals used as adaptors in Experiments 1 and 2 and the four different individuals used as test faces in all Experiments. (B) Examples of eye height deviations used in Experiments 1 and 3 (applied to one of the test individuals): examples include the most extreme values we tested (–50 pixels and +50 pixels). (C) Examples of mouth height deviations used in Experiments 2 and 4.



**Fig. 5.** (A) Example of a psychometric curve fitted to the data of a single observer from a one phase (unadapted) in a single session. (B) Another example data set showing how the aftereffect magnitude was computed as the difference between the PSEs of the baseline and adapted conditions.

adaptor position. We used, for each observer and each session, the difference between the physical adaptor value and the corresponding baseline PSE for that session (e.g., if the physical adaptor was +50 pixels and the observer had a baseline PSE of +4 pixels in that session, then the perceptual adaptor value was +46 pixels).

This measure provided the most accurate reflection of how far the adaptor was from each observers' current norm. The need to do this was indicated by a plot of baseline PSE against session number (Fig. 6A), which shows noticeable variation in baseline PSE across sessions (possibly reflecting changes in decision criteria, or in the



**Fig. 6.** Each observer's baseline PSE plotted for each successive session of testing, for (A) eye height in Experiment 1 and (B) mouth height in Experiment 2.

variety of eye heights to which observers were exposed in their natural environments between testing sessions).

## 2.2. Results and discussion

The linear model predicts the functions relating aftereffect magnitude to adaptor position should be straight lines, while the S-shaped model predicts the functions should be steepest close to the norm and then less steep further from the norm (Fig. 3A). Results supported the linear model.

Fig. 7 illustrates that linear fits to the data pass through the error bars of all data points in all cases. The linear fits produced  $R^2$  of .944 (TS), .984 (TG), .923 (TA) and .914 (OM). Trend analysis revealed this linear trend was significant in all cases ( $ps < .001$  for TS and TG;  $ps = .001$  for TA and OM). We also tested for higher-order trends, noting that the curve shape predicted by the S-shaped model would produce one or more nonlinear components (e.g., concave-down quadratic, reflecting a downwards bend relative to linear). No higher-order trend was found to approach significance for any observer (all  $Fs < 1$ ).

We also checked there was no carryover of adaptation across sessions. Because all adaptors were eyes-up (positive), carryover of adaptation would have produced a general trend in which baseline PSE increased (became more positive) across sessions. Fig. 6A shows this did not occur. Three observers had random changes in baseline PSE across successive sessions, with correlations between session number and baseline PSE of .152 (TS), .261 (TA) and  $-.212$  (OM). One observer showed a consistent trend ( $r = -.805$ , TG) but in the reverse direction to that predicted by carryover.

Results of Experiment 1 supported the linear model, in which face space codes eye height variations with equal sensitivity regardless of distance from the average eye height. Moreover, results of Experiment 1 showed that the linear response continued for eye heights far outside the normal range (i.e., even between +25 pixel and +50 pixel conditions). This argues that the eye height trajectory through face space codes a unit small difference between two bizarre extreme eye heights (i.e., eyes approaching the hair-line) as accurately as it codes a unit small difference between two normal eye heights (i.e., average and slightly-above-average).

## 3. Experiment 2: adaptor-position effects (mouth height)

In Experiment 2, we tested the generalisability of the linear model to shifts of mouth position. We also tested generalisability to downward shifts in adaptors, rather than upward shifts as in Experiment 1. The mouth was chosen because evidence suggests that, if any features of the face were encoded differently by face space than the eyes, then the mouth would be the most likely one. The mouth is typically less attended than the eye region in eye movements studies (Luria & Strauss, 1978; Walker-Smith, Gale, & Findlay, 1977), inversion effects can be smaller for mouth changes than eye changes (Barton, 2003; Barton, Keenan, & Bass, 2001), and prosopagnosics can show weaker performance deficits for the mouth than for the eyes (Orban de Xivry, Ramon, Lefèvre, & Rossion, 2008). Experiment 2 used five adaptor positions rather than the seven of Experiment 1, because the physical face region available for vertical movement is smaller for the mouth than for the eyes. However, the more extreme adaptor positions were again of bizarre appearance and clearly outside the normal range (Fig. 4C).

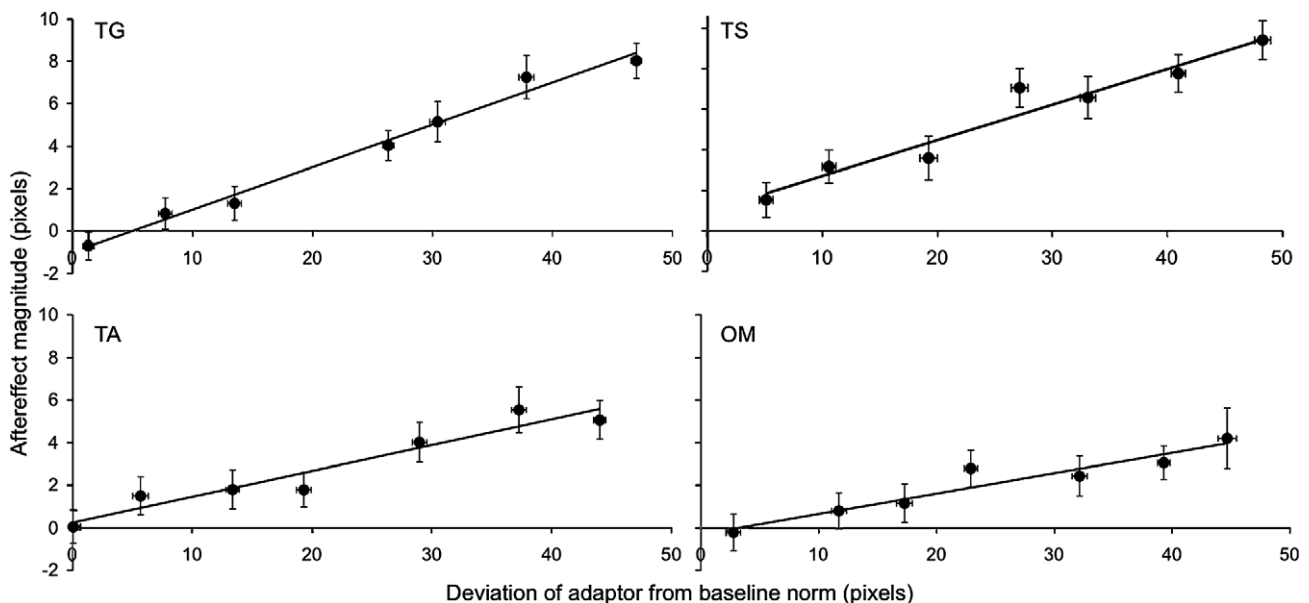
### 3.1. Method

#### 3.1.1. Observers

There were three naïve, experienced psychophysical observers (age 18–40; all male; all Caucasian), plus the first author. All were right-handed and had normal or corrected-to-normal vision. Each observer was tested for five 1-h sessions. Naïve observers were paid \$50.

#### 3.1.2. Design, stimuli, procedure and data analysis

Design, stimuli, procedure and data analysis were identical to Experiment 1. As illustrated in Fig. 4C, the four adaptor individuals (Fig. 4A) each had their mouth moved down to five adaptor positions ( $-5$ ,  $-10$ ,  $-15$ ,  $-20$ , and  $-25$  pixels), and the four test individuals (Fig. 4A) each had mouth moved up and down at 21 levels ( $0$ ,  $\pm 1$ ,  $\pm 2$ ,  $\pm 3$ ,  $\pm 4$ ,  $\pm 5$ ,  $\pm 6$ ,  $\pm 8$ ,  $\pm 10$ ,  $\pm 12$ , and  $\pm 15$  pixels); note shifting the mouth also changed apparent chin size and upper lip length. Each observer was tested in five separate sessions in random order. The task was to decide whether the mouth were 'too



**Fig. 7.** Results of Experiment 1: eye height aftereffect magnitudes as a function of perceptual adaptor position (i.e., difference between physical adaptor deviation level and baseline PSE for the session). Error bars in each direction show  $\pm 1$  SEM.

high' or 'too low', based on comparison with their imagined average mouth height of real-world faces. Each phase (baseline or adapted) contained 252 trials, structured as three blocks of 84 (21 mouth height levels  $\times$  4 test-face identities, in random order).

### 3.2. Results and discussion

Results (Fig. 8) again supported the linear model. The functions relating mouth adaptor position to size of aftereffect were straight lines.

Linear fits in Fig. 8 passed through the error bars of all data points in all cases, and produced  $R^2$  of .966 (TS), .972 (HD), .976 (CC) and .934 (AH). Trend analysis revealed the linear trend was significant for all four observers ( $ps < .001$ ). No higher-order trends were found (all  $Fs < 1$ ). Fig. 6B shows there was no carryover of adaptation across testing sessions. None of the correlations between baseline PSE and testing session for the four observers showed a negative trend (.574 (TS), .003 (HD), .615 (CC), and .533 (AH)), as would be predicted by carryover (because the adaptors were all mouth-down).

Experiments 1 and 2 tested the predictions of the linear and S-shaped tuning models via the effects of adaptor position on aftereffect magnitude. Results of the two experiments are in agreement, with two major findings. First, the linear model of opponent tuning curves underlies the neural coding of both eye height and mouth height within faces. For both attributes, there is no difference in coding sensitivity for faces that are closer to the average of the attribute trajectory than for faces that are further away from the average. Second, this linear tuning continues even for attribute values far outside the normal range of values in natural faces, for both eye height and mouth height.

### 4. Experiment 3: discrimination across the face continuum (eye height)

Experiments 3 and 4 were designed to examine the predictions of the linear and the S-shaped models with a different experimen-

tal method. The experiments also used a multiple-subject design, potentially allowing for converging evidence from more observers. We tested both eye height (Experiment 3) and mouth height (Experiment 4), each with both upwards and downwards shifts.

In Experiment 3, we tested discrimination of two faces differing by a fixed amount of eye height starting at different positions across the eye height continuum. Two faces differing only in eye height were presented sequentially, and observers indicated which of the two had higher (or lower) eyes. Pilot testing ( $N = 5$ ) revealed a 5-pixel eye height difference to be effective in avoiding both floor and ceiling effects, producing discrimination accuracy averaging approximately 70% for upright faces. Across the eye height continuum, we selected 20 starting positions ranging from bizarre eyes-down faces ( $-47.5$ ) to bizarre eyes-up faces ( $+47.5$ ), from which the eyes were moved up and down by 2.5 pixels. Discrimination performance at "starting position  $-47.5$ ", for example, refers to observers' ability to discriminate between the  $-50$  and  $-45$  pixel stimuli.

We tested discrimination of upright faces as our condition of primary interest. When discrimination accuracy is plotted against starting position, the linear model predicts a flat line (i.e., identical discrimination accuracy across all regions of the continuum) whereas the S-shaped model predicts a peak of discrimination accuracy near the average eye height compared to more extreme positive or negative starting eye heights (Fig. 3B). We also tested inverted faces as a control, to ensure our design was tapping the operation of the *face* system. Logically, observers could make sequential-presentation decisions based on attending to local distances (e.g., eye-nostril or eye-hairline) while ignoring the rest of the face. In practice, we thought this was unlikely given the substantial evidence that, for upright faces in experimental designs similar to the present study (i.e., sequential presentation and with only 1 session of trials per observer), observers process spatial relations between face features in the context of the whole face (for recent review, see McKone & Yovel, 2009). If our experimental design properly engages the face system, then we would expect to find better discrimination performance for upright than inverted faces.

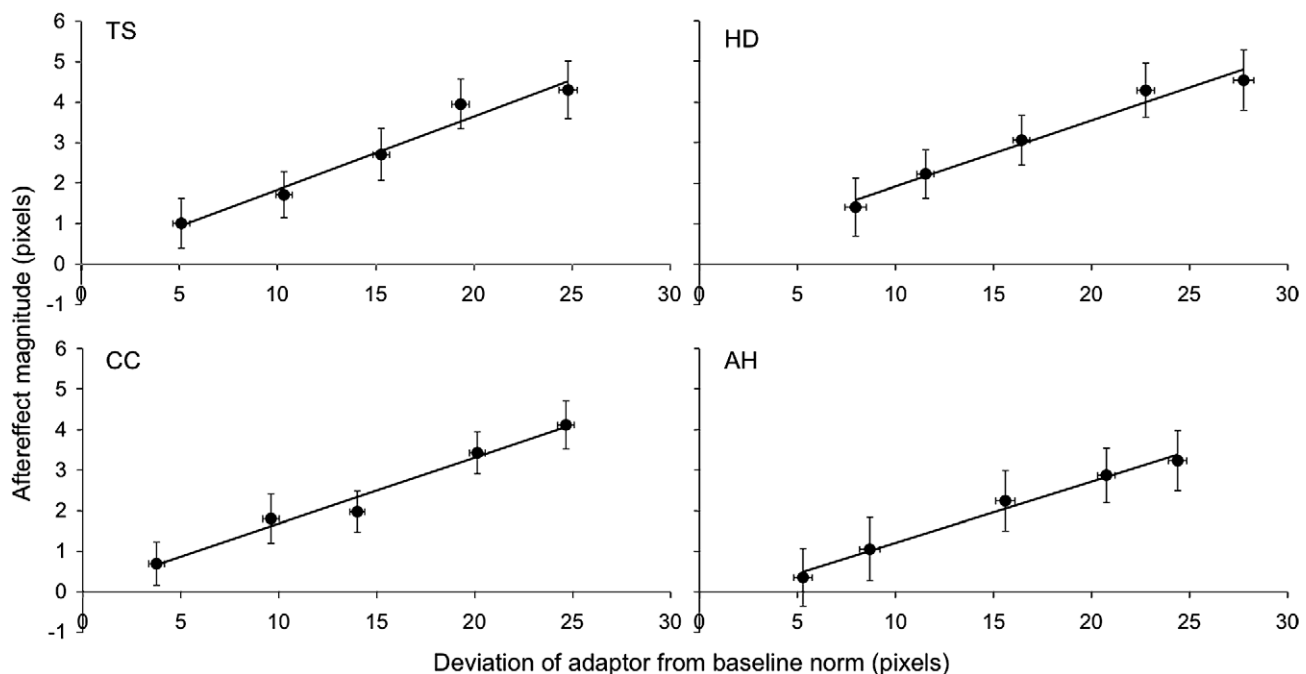


Fig. 8. Results of Experiment 2: mouth height aftereffect magnitudes as a function of perceptual adaptor position (i.e., difference between physical adaptor deviation level and baseline PSE for the session). Error bars in each direction show  $\pm 1$  SEM.

## 4.1. Method

### 4.1.1. Observers

Eighteen Australian National University students (age 18–26, 8 male) received course credit or \$10 payment. All observers were Caucasian, naïve to the purpose of the experiment, had normal or corrected-to-normal vision, and had not participated in Experiments 1 or 2. Each observer was tested for about 40 min.

### 4.1.2. Design

Starting position (20 positions) and orientation (upright, inverted) were manipulated within-subjects. Trials were blocked by orientation; half of the observers did the upright condition first, the other half did inverted first. The dependent measure was discrimination accuracy (% correct).

### 4.1.3. Stimuli

Stimuli were derived from the four test-face identities in Fig. 4A. For each of 20 starting positions of eye height ( $\pm 2.5$ ,  $\pm 7.5$ ,  $\pm 12.5$ ,  $\pm 17.5$ ,  $\pm 22.5$ ,  $\pm 27.5$ ,  $\pm 32.5$ ,  $\pm 37.5$ ,  $\pm 42.5$ , and  $\pm 47.5$  pixels), stimuli were pairs of a given individual with eye height differing by 5 pixels and centred around the starting position (e.g., for  $-17.5$ , the test images were  $-20$  and  $-15$  pixels). The inverted face images were identical to the upright-face images, except rotated  $180^\circ$ . Software and equipment were as for Experiment 1. Faces were presented at the centre of the screen, sized 298 by 250 pixels ( $10^\circ$  by  $7.9^\circ$ , viewed at 40 cm).

### 4.1.4. Procedure

On each trial, the first test face appeared for 250 ms followed after a 400 ms ISI by the second test face for 250 ms. This was followed by a blank screen and a question asking which of the two faces had higher eyes (or lower eyes, for half the observers). Pilot tests confirmed 400 ms ISI avoided any perception of induced motion of the eyes. Observers were told to respond on every trial and guess when uncertain. Keypads '1' and '2' were used to indicate their choice of the first or second face, respectively. There were 20 practice trials.

For each orientation, there were two identical blocks of testing. Within a block, for each starting position, each pair of test faces was presented sequentially once in each of the two possible orders (e.g.,  $-25$  followed by  $-20$ , and  $-20$  followed by  $-25$ ); the resulting 160 trials (20 starting positions  $\times$  4 test-face identities  $\times$  2 presentation orders) were presented in random order. The number of trials per orientation was 320 (20 starting positions  $\times$  4 test-face identities  $\times$  2 presentation orders  $\times$  2 blocks). There was a 1-min break between blocks. Collapsing across the four individual faces, each observer's accuracy for each position was based on 16 trials.

## 4.2. Results and discussion

Fig. 9A shows mean discrimination accuracy for eye height.

We first analysed upright faces. Results supported the linear model: discrimination accuracy was constant across the continuum for upright faces. This finding was examined statistically in two ways. First, a one-way repeated measures ANOVA found no difference in discrimination accuracy across the 20 starting positions,  $F(19, 323) < 1$ ,  $MSE = .010$ ,  $p > .7$ . Second, given that ANOVA is insensitive to the order of conditions on the  $x$ -axis, we also conducted trend analysis to provide a more powerful test for any non-linear components. To further maximise power, we collapsed data across negative and positive directions (noting that coding the upright data as a 10 position  $\times$  2 direction design revealed no interaction involving direction,  $F < 1$ ). We performed trend analysis on the resulting 10 starting position values, all now coded as positive (i.e.,  $+2.5 \dots +47.5$ ). Any improved discriminability close to zero

(i.e., the S-shaped model) would have predicted either a downwards linear trend across the positions, and/or a downwards trend reflected in more complex-shaped trend components. Results showed no linear trend ( $p > .5$ ), and no significant higher-order trends (up to 9th order, as there are 10 positions;  $ps > .08$ ), in support of the linear model. To ensure our upright-face task had tapped the face system, we conducted a 2 (orientation)  $\times$  20 (starting position) ANOVA. The main effect of orientation was significant ( $F(1, 17) = 12.379$ ,  $MSE = 1.726$ ,  $p < .01$ ), reflecting superior performance for upright faces ( $M = 67.1\%$ ) compared to inverted ( $M = 57.3\%$ ). No other effects or interactions were significant ( $ps > .6$ ).

Results of Experiment 3 confirmed that coding of eye height – including eye height well outside the normal range – follows the linear model. This replicates our adaptation results in Experiment 1. The approximately 10-percentage-point difference in accuracy between upright and inverted faces is typical for the size of inversion effects in sequential-presentation tasks (McKone & Yovel, 2009)<sup>5</sup>, confirming that discrimination for upright faces did indeed tap face perception, rather than low-level strategies.

## 5. Experiment 4: discrimination across the face continuum (mouth height)

Experiment 4 was similar to Experiment 3, but used mouth height. Pilot testing ( $N = 4$ ) revealed that a 5-pixel difference between test-face pairs was again effective in avoiding floor and ceiling effects. Across the mouth height continuum, we selected 10 starting positions ranging from bizarre mouth down ( $-25$  pixels) to bizarre mouth up ( $+25$  pixels). A smaller number of starting positions was used due to the more restricted range of movement possible for the mouth than the eyes.

### 5.1. Method

#### 5.1.1. Observers

Eighteen new Australian National University students (age 20–39, 9 male) received course credit or \$5 payment. All were Caucasian, naïve to the purpose of the experiment, and had normal or corrected-to-normal vision. Each observer was tested for about 20 min.

#### 5.1.2. Design, stimuli, and procedure

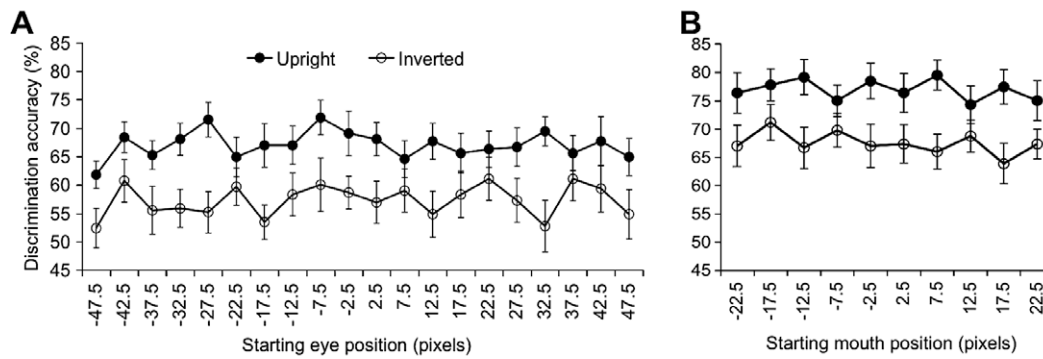
The experiment was a 10 (starting position:  $\pm 2.5$ ,  $\pm 7.5$ ,  $\pm 12.5$ ,  $\pm 17.5$ ,  $\pm 22.5$ ) by 2 (orientation: upright, inverted) within-subjects design. Stimuli and procedure were as for Experiment 3 except that the mouth was manipulated, rather than the eyes. The number of trials per observer in a given orientation was 160, divided in two blocks of 80 (10 starting positions  $\times$  4 face individuals  $\times$  2 presentation orders).

### 5.2. Results and discussion

Fig. 9B shows mean discrimination accuracy for mouth height. For upright faces, results indicate equal discrimination accuracy across all 10 positions of the continuum as predicted by the linear model ( $F(9, 153) < 1$ ,  $MSE = .011$ ,  $p > .8$ ). Further, collapsing positive and negative directions of shift (a 5 position  $\times$  2 direction ANOVA revealed no interaction with direction,  $F < 1$ ), trend analysis on the resulting 5 positions coded as positive ( $+2.5 \dots +22.5$ ) revealed no downwards linear trend ( $p > .7$ ) and no higher-order trends ( $ps > .5$ ). A 10 (position)  $\times$  2 (orientation) repeated mea-

<sup>5</sup> Face inversion effects are even larger on recognition memory tasks (around 15–25%).





**Fig. 9.** Accuracy for discriminating a fixed stimulus difference of five pixels in feature height as a function of starting position in the attribute continuum, averaged across observers, for (A) eye height in Experiment 3, and (B) mouth height in Experiment 4. Error bars show  $\pm 1$  SEM.

tures ANOVA revealed a strong main effect of orientation ( $F(1, 17) = 43.623$ ,  $MSE = 0.803$ ,  $p < .001$ ), with no other effects significant ( $ps > .6$ ). There was again a large approximately 10-percentage-point advantage in discrimination for upright faces ( $M = 76.9\%$ ) compared to inverted faces ( $M = 67.5\%$ ).

Results of Experiment 4 supported the linear tuning model for mouth height, in agreement with results of Experiment 2. Discrimination accuracy again remained as good for extreme attribute values (e.g., mouth nearly touching the nose) as for values close to the average, confirming that the linear model applies even for mouth values far outside the normal range. The large inversion effect confirmed that discrimination performance for upright faces was tapping the face system rather than low-level representations.

## 6. General discussion

The primary aim of the present study was to investigate the shape of the opponent response curves underlying face space coding of identity-related information. Our results showed clear evidence in support of the linear model, obtained across two experimental methods (adaptation aftereffects and pairwise discrimination) and two face manipulation types (eye height and mouth height). Moreover, because our manipulations covered the entire range that was physically possible in the face stimulus – such as from the mouth nearly touching the nose to the mouth nearly touching the chin – our results have shown that the linear model holds even for values of attribute positions far outside the normal range. That is, not only that coding along a trajectory in face space is as detailed for attribute values towards the edge of the normal range as for values around the average, but also that the same detail of coding is maintained for attribute values that would never be encountered in natural face images.

### 6.1. Mouth height and two-pool opponent coding

Our findings also address the more basic question of whether face space uses two-pool opponent (norm-based) or multichannel (exemplar-based) coding. The increase in aftereffect magnitude as mouth position of the adaptor was shifted further from the average (Experiment 2) adds the attribute of *mouth height* to the list of identity-related variables coded in a two-pool fashion (see Fig. 1 for logic). This type of coding has now been found for eye height (human subjects, Robbins et al., 2007; present Experiment 1), mouth height (humans, present Experiment 2), various other feature position and feature shape manipulations including interocular distance and eye size (single-cell recording in monkeys, Freiwald et al., 2009), horizontal or vertical expansion–contraction

(humans, Webster & MacLin, 1999), and the face anti-face identity manipulation (humans and monkeys, Leopold et al., 2006; Leopold et al., 2001; Rhodes & Jeffery, 2006). Thus, it is now very clear that, while face view and eye gaze appear to be coded in a multichannel fashion (Calder et al., 2008; Fang & He, 2005; Jenkins et al., 2006), face shape, as involved in representing facial identity, is coded in a two-pool opponent fashion.

This result is of theoretical importance because the lowest-level precursor of shape information – the coding of line orientation in V1 – is represented in a multichannel fashion, with narrowly tuned bell-shaped response functions (cf., Fig. 1B). This indicates that, at some point in the ventral processing stream for faces, there must be a switch in coding style for shape information. Currently, it is unknown where this switch occurs. Potentially, it might not occur until the stage of representation of faces *per se* in high-level vision. Alternatively, given evidence of two-pool opponent coding for basic shape attribute of horizontal versus vertical elongation (Regan & Hamstra, 1992; Suzuki, 2005), it is possible the switch has already occurred by the stage of mid-level representations that feed into high-level face space (Connor, Brincat, & Pasupathy, 2007).

### 6.2. Does linearity hold for all face shape attributes? Theoretical issues

Our main finding concerns the specific shape of the opponent tuning curves, which is linear for the two facial attributes we tested, namely eye height and mouth height. How generalisable is this finding? Do all possible trajectories through face space use linear coding? Or are mouth height and eye height somehow special in using linearity, while other trajectories (e.g., an anti-face identity trajectory in the procedure of Leopold et al., 2001) use nonlinear coding? We discuss relevant empirical findings in the next section. Here we raise two theoretical issues regarding whether eye height and mouth height might be “special”.

First, eye height and mouth height are so-called *spacing* or *second-order relational* attributes (e.g. Barton et al., 2001; Le Grand, Mondloch, Maurer, & Brent, 2001; Leder & Bruce, 2000), and there has been a long tradition of proposing that the spacing between face features is independent of the shape of the features themselves, and that spacing plays some special role in face perception, at least with respect to holistic/configural processing (Maurer, Le Grand, & Mondloch, 2002). Given that in the present study we did not test any pure feature *shape* changes (e.g., *size* of mouth), we cannot rule out the possibility that our linear pattern applies to spacing information but not feature information. However, we suspect this is unlikely. A recent review (McKone & Yovel, 2009) found that dissociations between features and spacing are *not* the typical pattern, and that they occur only when the feature changes can be discriminated based on information available

from outside the face system (e.g., when there is a large colour change, or the feature is presented alone outside facial context). Also note that the basic idea of independent spacing versus feature differences is fraught with conceptual difficulty. For example, our manipulations of eye height and mouth height – which would typically be labelled ‘spacing’ changes in this literature – also resulted in changes to local feature shape (e.g., nose length and chin size).

A second reason for potentially considering eye height and mouth height as special is it that it can be tempting to think of these as “simple” attributes of faces, in the sense that they can be defined mathematically merely by the  $(x, y, z)$  head coordinates of the centre of the feature. This view would imply that a trajectory corresponding to facial identity strength (e.g., Leopold et al., 2001) would be more “complex”, in that stimuli along the trajectory vary in information from across the entire facial region. If “complex” trajectories like identity were formed from the underlying projections onto cardinal axes comprising “simple” variables like eye height, then full face identity would be coded linearly only if *all* the underlying simple attributes from which that identity of a face is constructed were coded linearly.<sup>6</sup> An alternative view, however, is that eye height and mouth height are *not* “simple”. Mathematical simplicity does not guarantee simplicity to the perceptual system of humans; that idea would essentially reduce to the claim that attributes like eye height and mouth height are the true basis dimensions of face space and, currently, the basis dimensions of face space are unknown. Indeed, some computational models (Sirovich & Kirby, 1987; Turk & Pentland, 1991) produce eigenfaces as basis dimensions, each of which appears highly “complex” (i.e., includes information from across the entire face, and of different types such as position, brightness, texture and curvature). If the true cardinal axes of face space were eigenfaces, then the apparently simple variable of eye height would in fact be constructed from complex contributions of multiple different basis vectors.

### 6.3. Does linearity hold for all face shape attributes? Empirical evidence

Several studies have produced data relevant to whether linearity holds for other face shape attributes. None have used our adaptor-position method (Experiments 1 and 2). Six studies have tested *discrimination* at various points along a trajectory (similar to Experiments 3 and 4), or equivalent measures such as similarity rating or perceptual difference scaling. Relevant data are also available from two studies testing whether *adaptation improves discrimination*, and from two studies reporting *single unit data* in monkeys.<sup>7</sup> Our discussion of these studies will lead to the following conclusions: (a) there is considerable heterogeneity in findings regarding linear versus nonlinear tuning; (b) this heterogeneity appears to be genuine rather than due to any methodological flaws; (c) the origin of the heterogeneity cannot be explained in terms of several factors we consider, and currently remains mysterious; and (d) neural response models can assist in providing at least local coherence of several findings.

<sup>6</sup> This idea assumes that face space uses a Euclidean distance metric. This has been demonstrated with simplified synthetic faces (Wilson et al., 2002), but has not been investigated using natural faces.

<sup>7</sup> Some readers might think recognition memory studies are relevant: the classic finding that memory is better for distinctive than typical faces (Light, Kayra-Stuart, & Hollander, 1979; Valentine & Bruce, 1986) might be taken to suggest that faces further away from the centre of face space are coded with more sensitivity than are faces closer to the center. However, density of exemplars in face space is higher closer to the average and lower further from the average (Catz, Kampf, Nachson, and Babkoff, 2009; Valentine, 1991), and memory confounds underlying sensitivity of coding with effects arising from the fact that, at memory retrieval, there are fewer close neighbours with which a distinctive face can be confused than there are for a typical face.

#### 6.3.1. Discrimination close to and away from the average face

Discrimination studies have produced heterogeneous results, consistent with three different tuning models. Two findings agree with ours in showing linearity: equally good discrimination of stimulus changes centred around a face closer to the average and around a face further from the average. One is Experiment 1 of Rhodes et al. (2007), where the manipulation was of interocular distance around starting faces with different identity strength. The other is Ng et al. (2008), where the manipulation was of gender around starting faces at different positions on the male–female gender continuum. In contrast, other studies revealed nonlinearity and, moreover, different *types* of nonlinearity. Two findings follow the S-shaped model: better discrimination around faces closer to the average than around faces further from the average. This has been reported by Wilson et al. (2002), who used synthetic radial frequency faces with a face-cube manipulation (i.e., faces shifted in orthogonal directions from a starting face), and Tanaka and Cornille (2007), who used morphed faces between a typical and a distinctive natural face. In the third pattern of results, two studies have found *worse* discrimination for faces close to the average than for some other locations further from the average. Rhodes et al.’s (2007) Experiments 2–4 found this pattern using the identity-strength manipulation. Worse discrimination was obtained around the average face than around faces of up to 30% identity strength (Experiment 2), with discrimination then continuing to improve until a maximum at approximately 90% identity strength, and gradually worsening again for caricatures up to 150% (Experiments 3 and 4). A similar pattern was also reported by Dakin and Omigie (2009). They found considerable variability across observers and faces as to the exact location of maximum discriminability which the authors attributed to the distinctiveness of the endpoint face. These studies suggest a different shape of tuning curve than either the linear or S-shaped models. We refer to this new model as “separated-S”, in that the response functions are S-shaped but cross at the zero attribute value in a shallow region of the S, rather than at the steepest region (Fig. 10).<sup>8</sup>

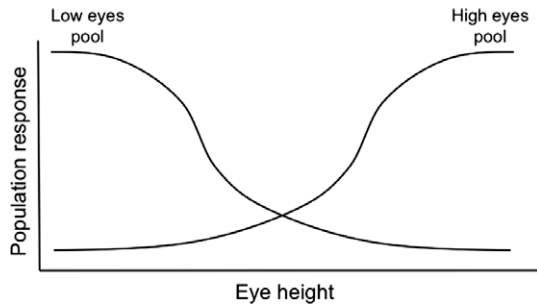
#### 6.3.2. Heterogeneity is not attributable to methodological flaws

The heterogeneity of discrimination findings does not appear to be attributable to methodological problems. The only methodological critique raised by Rhodes et al. (2007) was that Wilson et al. (2002) used synthetic faces that were unnatural in appearance. However, all other studies have used realistic-appearance faces and have still found all three patterns. Further, where data are available from an independent technique, this has always agreed with the results in the same study for the discrimination task. In our own experiments, our adaptor-position findings (linear) agreed with our discrimination findings (linear). In the next subsection, results of the effects of adaptation on discrimination also agree with the baseline (unadapted) discrimination findings.

#### 6.3.3. Changes in discriminability following adaptation

Another approach to determining response curve shape is to examine discriminability before (i.e. baseline) and after adaptation, as has been conducted by Rhodes et al. (2007) and Ng et al. (2008). Both studies took an idea from low-level vision, namely that adaptation may enhance discriminability in the range of stimuli most commonly encountered (Barlow, 1990; Simoncelli & Olshausen,

<sup>8</sup> The model we illustrate here is different from the one Dakin and Omigie (2009, Fig. 2) used to explain their results. Their model included only *one* pool of neurons, with tuning similar in general shape to one of our two pools, but with the critical difference that response to the average face (and anti-identity faces) was zero. We prefer our version because (a) a one-pool model cannot, in itself, explain the existence of face/anti-face adaptation aftereffects, and (b) neurophysiology shows neural response to the average face is not zero but intermediate between responses to positive and negative attribute values (Freiwald et al., 2009; Leopold et al., 2006).



**Fig. 10.** A “separated-S” tuning model. Each pool has an S-shaped response, but rather than the pools crossing at the steepest part of the curve (cf. standard S-shaped model in Fig. 2), the S-shapes have been shifted apart so that the pools cross at a shallower region. This model predicts poor discrimination around average faces, better discrimination around faces intermediate in distance from the average, and poorer discrimination again around more extreme faces.

2001), and proposed that this idea might apply to faces (Rhodes et al., 2005; Webster, Werner, & Field, 2005). The prediction was that adaptation should *improve* discrimination around the adapted value and/or the new norm position (e.g., adaptation to male should improve discrimination of male faces, and adaptation to anti-Bill should improve discrimination of faces shifted away from the average in the anti-Bill direction). The findings of both studies rejected this prediction.

While neither study was able to provide a complete explanation of why the prediction was rejected, their results are quite straightforward to explain using the particular response model that we argue underlies their *baseline* results (i.e. the condition without adaptation, namely pre-adaptation in Ng et al., 2008, and where subjects adapted to the average face in Rhodes et al., 2007). We previously noted that the baseline discrimination data in Experiment 1 of Rhodes et al. (2007) showed equal sensitivity to all interocular distance changes regardless of the distance of the starting point face from the average, which indicates *linear* tuning. Linear tuning predicts that adaptation will *not* alter discriminability, because adaptation will not alter the property that the system is equally sensitive to changes at all points along the continuum. This is exactly what they found: no change of any sort (improvement or worsening) following adaptation. For gender, the findings of Ng et al. (2008) were similar. Their baseline discrimination results supported a linear model, and their post adaptation results showed no change in discriminability. Finally, the baseline discrimination results in Rhodes et al.'s (2007) Experiments 2–4, using an identity-strength manipulation, supported the separated-S pattern: discrimination was worse around the average than further from the average. This type of tuning curve predicts that adaptation will affect discriminability. However, rather than *improving* discriminability around the new norm, it predicts *worsening*. This is exactly what Rhodes et al.'s Experiment 2 found. The poor discriminability that was originally around the average face shifted, after adaptation, to the identity-strength value that corresponded to the new norm.

These analyses of response curve shapes allow a coherent explanation of what were an apparently confusing set of results. They also allow an additional prediction: it should be possible for adaptation to improve discriminability as Rhodes et al. (2007) and Ng et al. (2008) originally predicted. This will occur only where an S-shaped model applies, where there is a *peak* of discriminability around the average in the unadapted condition. From the baseline discrimination results we reviewed earlier, this predicts that adaptation should improve discrimination around a new post adaptation average in the circumstances tested by Wilson et al. (2002) and Tanaka and Corneille (2007).

#### 6.3.4. Single unit recording in monkeys

Single unit recording in monkeys directly determines the shape of tuning curves. In face-selective cells of the temporal “middle face patch”, Freiwald et al. (2009) measured response functions to cartoon faces varied from one end of an extreme (e.g., mouth down) to the other (mouth up). The majority of cells showed maximum response at one (or other) *end* of the extreme (i.e., monotonic response functions consistent with two-pool opponent coding). Of these cells, the majority (67%) showed linear tuning (see Fig. 4D of Freiwald et al.) Other cells showed various nonlinear patterns. Using realistic faces, Leopold et al. (2006; Supplementary Figs. 4 and 5) presented relevant raw tuning curves (but no formal analysis), plotting the response of a neuron as the identity strength of a face was increased towards one identity (zero to positive numbers) or increased towards another identity (zero to negative numbers; note this identity was not the anti-face of the positive-coded identity). Plots indicate some cases of linearity, but also other cases of nonlinearity of various forms. Thus, single unit recording results support our own finding that linear response curves can occur, and indeed imply that linearity may be common, but also support our conclusions from other human studies that various nonlinear patterns can occur.

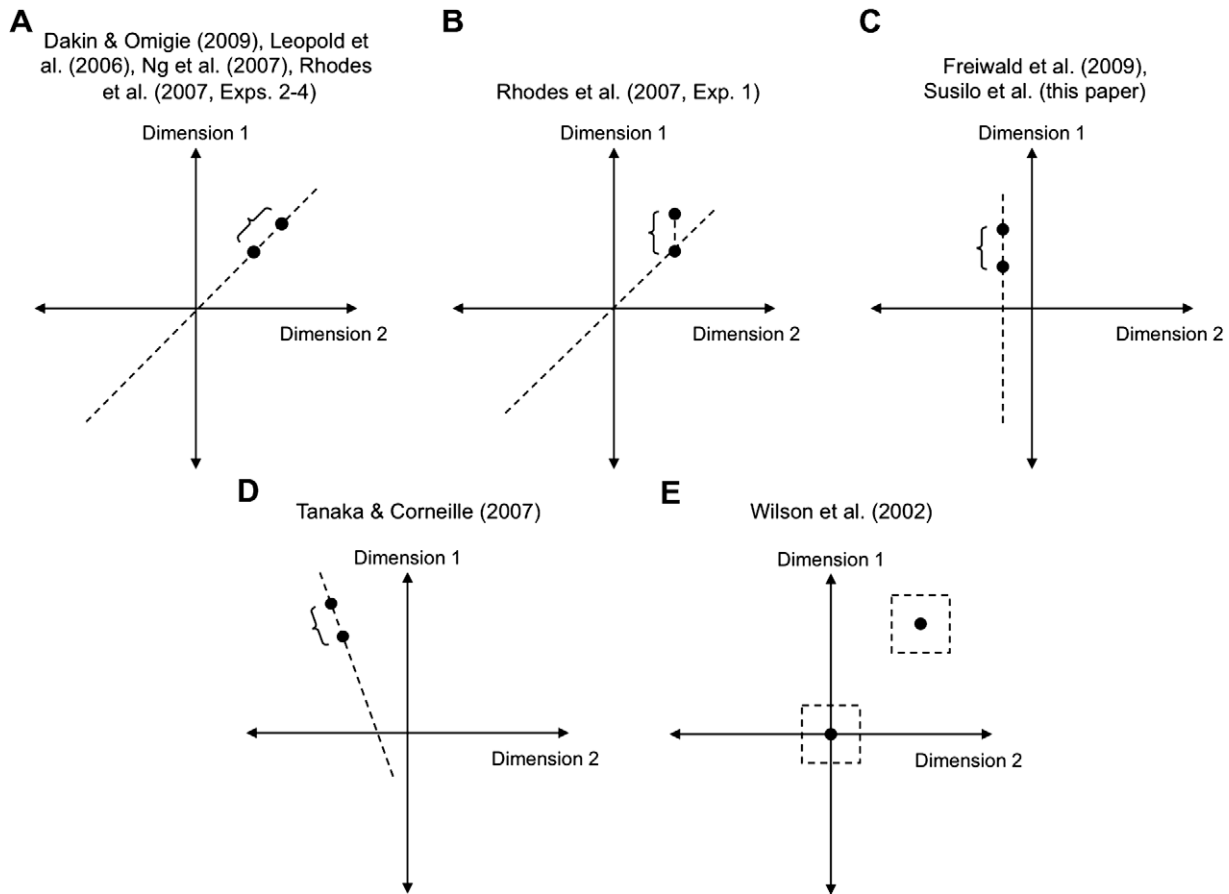
#### 6.3.5. What explains the heterogeneity?

Across the various types of studies there is clear agreement that the tuning functions of neural responses in face space show considerable heterogeneity. In particular, it appears that face space uses linear tuning in some situations, S-shaped tuning in other situations, and separated-S tuning in still other situations. What explains this heterogeneity?

At present, it is difficult to see any simple factor that can explain the full pattern of results. *Task differences* have been raised as possibly relevant by Rhodes et al. (2007) and Dakin and Omigie (2009). However, task cannot explain the full patterns across all data now available. For example, in humans, threshold discrimination has produced both S-shaped pattern (Wilson et al., 2002) and separated-S pattern (Dakin & Omigie, 2009), while above-threshold similarity ratings has also produced both S-shaped pattern (Tanaka & Corneille, 2007) and separated-S pattern (Rhodes et al., 2007). And, in Leopold et al.'s (2006) study, the monkeys' task was held constant, yet individual neurons produced all three patterns.

Second, an idea of “*simple*” versus “*complex*” manipulation types does not work either, defining a “*simple*” change as an alteration in local spacing and a “*complex*” change as a manipulation that affects information across the entire facial region simultaneously. For example, in humans, linear pattern has been produced by “*simple*” manipulations of eye/mouth height (present study) and interocular distance (Rhodes et al., 2007), but linear pattern has also been produced by the “*complex*” manipulation of gender (Ng et al., 2008). In monkeys, the study always used a “*complex*” manipulation of identity strength, yet individual neurons produced all three patterns (Leopold et al., 2006).

Third is what we label *location and direction in face space* (Fig. 11). Here, we expand on an idea of Dakin and Omigie (2009), who noted that stimulus changes in a discrimination task could be *on-* or *off-axis*. In *on-axis* procedures (Fig. 11A), the axis along which the *to-be-discriminated change* is made is the same axis along which distance of the *starting point* from the average is varied; this occurred, for example, in Dakin and Omigie's study (discrimination was of “amount of Bill” with starting point varied in identity strength along the “Bill” axis) and in Ng et al. (2008; discrimination of gender with starting point varied in gender). In *off-axis* procedures (Fig. 11B), the *to-be-discriminated changes* and the starting point are varied in different directions in face space; this occurred, for example, in Experiment 1 of Rhodes et al. (2007; discrimination was of interocular distance but starting



**Fig. 11.** Illustration of properties of different discrimination experiments using a simplified, two-dimensional face space: (A) on-axis discrimination with axis passing through centre of face space; (B) off-axis discrimination with axis passing through centre of face space; (C) on-axis discrimination with axis passing through zero value for the tested attribute but not centre of face space; (D) on-axis discrimination on trajectory between a random typical and distinctive face; and (E) off-axis change around the average face in the centre of face space and a random away-from-average face.

point was varied in identity strength). There are also other variations across studies. In the examples given so far (Fig. 11A and B), the starting point trajectory passed through the centre of face space (i.e., through the average face). In other studies, this was not the case. In our own experiments (Fig. 11C, also Freiwald et al. (2009)) the manipulations were on-axis, and the axis passed through the centre point of the attribute being varied (e.g., zero eye height) but not through the centre of face space (i.e., our zero eye height stimulus was an individual who was not average in other facial attributes). Fig. 11D illustrates the approach of Tanaka and Corneille (2007), where the manipulation was on-axis and the axis (running between a random typical face and a random distinctive face) was very unlikely to have passed through the centre of face space, and in fact might not have passed through the zero value for any facial attribute. Finally, Fig. 11E illustrates the procedure of Wilson et al. (2002), where the changes were off-axis in the sense that changes were made in orthogonal directions from a given starting face, but there was no underlying axis of starting face values.

These observations can perhaps be used to understand the variation of findings across studies. Dakin and Omigie (2009) showed that, at least where the axis passes through the centre of face space, an on- versus off-axis manipulation can matter. They demonstrated that discrimination of identity-strength-of-Face1 with starting point varied in identity-strength-of-Face1 produced separated-S pattern, while discrimination of identity-strength-of-Face 1 with starting point varied in identity-strength-of-Face 2 produced S-shaped pattern. They proposed this idea could explain

the conflict between their results and those of Wilson et al. (2002). Unfortunately, however, a simple on- versus off-axis idea is insufficient to explain the full pattern of results across studies. In humans, on-axis changes have produced linear pattern (present study; Ng et al. (2008)), S-shaped pattern (Tanaka & Corneille, 2007) and separated-S pattern (Dakin & Omigie, 2009; Rhodes et al., 2007). Off-axis changes have produced linear pattern (Rhodes et al., 2007, Experiment 1) and S-shaped pattern (Wilson et al., 2002). In Leopold et al. (2006), all changes were on-axis yet all three patterns of results were produced. Thus overall, we have to conclude that the origin of the different results across studies remains a mystery. Given that Dakin and Omigie (2009) have shown that on versus off axis *can* matter, it is possible that future investigation of the more complex combination of properties shown in Fig. 11A–E may cast some light on the issue.

#### 6.4. Linear versus nonlinear response curves in face space: Summary

Putting our results together with those of the previous literature leads to the conclusion that the opponent response curves underlying face space trajectories use a variety of tuning shapes – including linear, S-shaped, and separated-S-shaped. We also conclude that the particular circumstances that produce different types of tuning are currently poorly understood, and that extensive further research might be needed to resolve this issue.

We emphasise the present study has made several contributions towards this eventual end. First, our study is the first that explicitly argues for a linear model with supporting data. The



findings of both Rhodes et al. (2007, Experiment 1) and Ng et al. (2008) support a linear model, yet neither of these studies presented such a model and so were not able to provide a complete theoretical explanation for their results. Other papers have illustrated a linear model (e.g., Tsao & Freiwald, 2006), but presented no empirical evidence nor stated any explicit justification for choosing linear curves in their illustration of two-pool opponent coding. Second, our study is the first to make explicit the relationship between a variety of response models and the predictions for the baseline discrimination task. It is also the first to draw together the results of the various discrimination studies in the light of these models. Third, we have introduced the adaptor-position manipulation as a new method for testing response curve shape, and have shown that this produces results that converge with those of the discrimination task. When we found that linear tuning applies using discrimination (i.e., for eye height and mouth height), we found independent confirmation of the same model from the adaptor-position technique. Fourth, by placing baseline discrimination results in light of theoretical consideration of response curve models, we have explained why previous studies have failed to find discrimination improvement following adaptation (Ng et al., 2008; Rhodes et al., 2007). We have also shown that the idea of better discrimination following adaptation only holds for one particular type of response curve model (S-shaped), and that the results of previous studies – either no change in discrimination after adaptation, or a *worsening* of discrimination around the new norm – are in fact exactly as predicted by the models – linear and separated-S respectively – which we derived from their baseline discrimination data.

Finally, we remind readers that we measured psychophysical responses rather than neural responses. However, consistent with findings in low- and mid-level vision, results suggest our psychophysical data provided good information about neural response tuning curves. Our two different methods (adaptor-position and discrimination) both agreed with the predictions of a linear neural tuning model, despite measuring *perceptual* responses. In addition, our results were consistent with single-cell tuning curves in monkeys (Freiwald et al., 2009), where 67% of cells showed linear responses to manipulations conceptually similar to those in the present study (e.g., eyes shifted apart).

#### 6.5. Linear face space coding can continue far beyond the normal range

A final major finding of our study is that linear coding in face space can continue far beyond the normal range. Facial anthropometry data for young adult Caucasian males (same age, race, and sex as face stimuli used in this study) show the distance between the middle of the eyes and the hairline has a mean of 67.1 mm and a standard deviation of 7.5 mm, while the distance between the middle of the lips and the chin has a mean of 50.7 mm and a standard deviation of 4 mm (Farkas, Hreczko, & Katic, 1994). In our stimuli, these numbers correspond to approximately a  $\pm 15$  pixel deviation for eye height, and a  $\pm 8$  pixel deviation for mouth height, which is far below the maximum we tested ( $\pm 50$  pixels for eyes, and  $\pm 25$  pixels for mouth). Consistent with this, in a normality rating experiment using our present test faces, observers rated eye heights up to  $\pm 10$  pixels as completely normal,  $\pm 20$  pixels as noticeably less normal, and  $\pm 50$  as highly bizarre (McKone et al., 2005).

Our finding of face space coding far beyond the normal range is anticipated by the established fact that extreme face distortions that make the face appear distinctly weird produce aftereffects on the average face (Rhodes et al., 2003; Robbins et al., 2007; Watson & Clifford, 2003; Webster & MacLin, 1999). However, our finding that coding for changes can be as sensitive around extreme bizarre faces as around faces in the normal range is novel. By test-

ing positions far outside the natural range, and showing that aftereffect magnitude continues to increase linearly for extreme adaptor positions (Experiments 1 and 2), and that discrimination accuracy remains constant (Experiments 3 and 4), the present study has shown that the linearity of response curves for eye height and mouth height extends even to unnatural faces.

This finding is of strong theoretical importance. In terms of neural models, it argues that – for the face space trajectories of eye height and mouth height – there is no flattening out at all of response curves, and coding instead remains as sensitive to changes around extreme values as around values within the normal range.<sup>9</sup> This in turn speaks to the theoretical principle of optimal coding, the idea that neural responses in the face system should be allocated to match the statistical distribution of faces in the surrounding environment (Rhodes et al., 2005; Webster et al., 2005). In low-level vision, this idea has been fruitful in understanding coding mechanisms of contrast (Laughlin, 1990) and colour (MacLeod & von der Twert, 2003). In both cases, response characteristics of the low-level neurons are matched to the distribution of contrasts and colours in natural scenes. Our present results for faces, however, is inconsistent with this idea. Our linear coding implies that – at least for the face attributes of eye height and mouth height – neural responses are allocated equally for all possible values in a given face dimension, regardless of the frequency with which those values occur.

Why would the neural resources of face space be allocated to detailed coding of values far outside those experienced in everyday life? We see at least two possibilities.

One is that, although the range of attribute values in our stimuli is beyond the normal range *relative to the other internal features* in front view faces, it is not beyond the normal range if face space were coding identity-related information: (a) with respect to external head outline, and (b) not only for front view faces but also for faces with other up-down head rotations around the horizontal axis (“pitch”). Robbins et al. (2007) proposed that two-pool opponent coding might allow for pitch rotations, noting that feature positions which are extremely high or low relative to other features in the frontal view change their apparent position when the head is rotated upward or downward (e.g., eyes can appear to almost touch the hairline if the face is viewed sufficiently far from below). Therefore, what initially appears to be face space coding beyond the normal range in our stimuli could actually be coding *within* the normal range allowing for pitch rotations. This idea predicts that pitch tuning of at least some face neurons must be broad, a result observed in monkeys: for example, Perrett et al. (1985) reported one cell which responded maximally to faces rotated 45°-up reduced its spike rate only by 1/3 of the peak value for a front view face. In humans, there are no studies of neural overlap for faces with different pitches. However, on the presumably related topic of “yaw” tuning (left–right rotations of the head), broad tuning has been demonstrated, with shape distortions (expansion/contraction) producing significant transfer of adaptation across views separated by 45° (Jeffery, Rhodes, & Busey, 2006) and even 60° (Jeffery, Rhodes, & Busey, 2007).

Alternatively, the natural range for front view faces could be larger than it might initially appear, if it covers not only faces of adult humans, but also those of children and babies. Height of the eyes

<sup>9</sup> This raises the question of whether the response curves carry on increasing infinitely. In unpublished work in our laboratory (McKone & Edwards, in preparation), we have found that shifting the eyes past the hairline immediately drops the aftereffect magnitude to zero. This suggests that breaking the first order configuration of the face simply switches off responses of the neurons that code eye height (an idea that might also account for the very small aftereffects resulting from asymmetric changes to eye height in Robbins et al., 2007). Note that the dropoff cannot reflect any form of turnaround of response curves, because turnaround falsely predicts the existence of face *metamers* (i.e., two face stimuli with physically different attribute values yet perceived as the same) (Robbins et al. 2007, Fig. 12).

and mouth in humans, relative to the external outline of the face, changes noticeably across development, from the large-forehead-tiny-chin arrangement of the newborn to the smaller-forehead-larger-chin arrangement of the adult. Potentially, tuning of eye height pools in our adult observers might also cover the faces of children and babies. One likely argument against this view is the traditional presumption that the coding range of face space are constructed from ongoing lifetime experience with faces, and are therefore not fixed but instead tuned to maximise discrimination amongst the recent diet of faces to which a person has been exposed (Rhodes et al., 2005; Webster et al., 2005). However, the evidence for this presumption is less than compelling. Recent developmental studies have shown that extended lifetime experience is not required to produce face space: face space is present, and shows at least qualitatively adult-like tuning, at the youngest ages tested (e.g., 4–5 year olds using distinctiveness effects, McKone & Boyer, 2006; 8 year olds using adaptation aftereffects, Nishimura, Maurer, Jeffery, Pellicano, & Rhodes, 2008; also see Crookes and McKone (2009), for a discussion of whether face space shows quantitatively adult tuning in children). Second, the phenomenon traditionally taken as evidence for ongoing tuning of face space dimensions – the other-race effect, in which discrimination of other-race faces is poorer than discrimination of own-race faces – has recently been given alternative interpretations that make no reference to differences in face space coding sensitivity (such as social outgroup categorisation, e.g., Bernstein, Young, & Hugenberg, 2007; and own-group effects on holistic processing, e.g., Susilo, Crookes, McKone, & Turner, 2009). Third, the idea that face space is constantly updated to improve sensitivity to the particular range of face types to which we have been exposed predicts that adaptation to a particular face type should improve discrimination. As we saw previously, however, published studies to date have found either that adaptation worsens discrimination around the norm value or leaves it unaffected (Ng et al., 2008; Rhodes et al., 2007). Overall, there seems to be no intrinsic reason to reject the hypothesis that face space may include dimensions that are very broadly tuned because they are designed to cover facial structure of adult humans, children, and babies.

## Acknowledgments

This research was supported by Australian Research Council Grants DP0450636 and DP0984558 to EM and ME. TS is grateful for scholarship support from ANU Centre for Visual Sciences and overseas student fee waiver from ANU Department of Psychology.

## References

- 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* (pp. 363–375). Cambridge: Cambridge University Press.
- Barton, J. J. S. (2003). Attending to faces: Change detection, familiarization, and inversion effects. *Perception*, 32, 15–28.
- Barton, J. J. S., Keenan, J. P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. *British Journal of Psychology*, 92, 527–549.
- Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, 18(8), 706–712.
- Calder, A. J., Jenkins, R., Cassel, A., & Clifford, C. W. (2008). Visual representation of eye gaze is coded by a nonopponent multichannel system. *Journal of Experimental Psychology: General*, 137(2), 244–261.
- Catz, O., Kampf, M., Nachson, I., & Babkoff, H. (2009). From theory to implementation: Building a multidimensional space for face recognition. *Acta Psychologica*, 131(2), 143–152.
- Cohen, J., MacWhinney, B., Flatt, M., & Provost, P. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments and Computers*, 25, 257–271.
- Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, 17, 140–147.
- Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face-encoding, or face-space. *Cognition*, 111(2), 219–247.
- Dakin, S., & Omigie, D. (2009). Psychophysical evidence for a non-linear representation of facial identity. *Vision Research*, 49(18), 2285–2296.
- Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, 24(4), 419–430.
- Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic patients. *Neuropsychologia*, 44, 576–585.
- Fang, F., & He, S. (2005). Viewer-centered object representation in the human visual system revealed by viewpoint aftereffects. *Neuron*, 45, 793–800.
- Farkas, L. G., Hreczko, T. A., & Katic, M. J. (1994). Craniofacial norms in North American Caucasians from birth (one year) to young adulthood. In L. G. Farkas (Ed.), *Anthropometry of the head and face* (2nd ed., pp. 241–335). New York: Raven Press.
- Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, 12, 1187–1196.
- Frisby, J. (1979). *Seeing: Illusion, brain and mind*. Oxford: Oxford University Press.
- Jeffery, L., Rhodes, G., & Busey, T. (2006). View-specific coding of face shape. *Psychological Science*, 17(6), 501–505.
- Jeffery, L., Rhodes, G., & Busey, T. (2007). Broadly tuned, view-specific coding of face shape: Opposing figural aftereffects can be induced in different views. *Vision Research*, 47, 3070–3077.
- Jenkins, R., Beaver, J. D., & Calder, A. J. (2006). I thought you were looking at me: Direction-specific aftereffects in face perception. *Psychological Science*, 17(6), 506–513.
- Laughlin, S. B. (1990). Coding efficiency and visual processing. In C. Blakemore (Ed.), *Vision: Coding and efficiency*. Cambridge: Cambridge University Press.
- Lawson, R. P., Clifford, C. W. G., & Calder, A. J. (2009). About turn: The visual representation of human body orientation revealed by adaptation. *Psychological Science*, 20(3), 363–371.
- Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature*, 410, 890.
- Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology*, 53A(2), 513–536.
- Leopold, D. A., Bondar, I. V., & 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 aftereffects. *Nature Neuroscience*, 4(1), 89–94.
- Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 5(3), 212–228.
- Luria, S. M., & Strauss, M. S. (1978). Comparison of eye movements over faces in photographic positives and negatives. *Perception*, 7(3), 349–358.
- MacLeod, D. I. A., & von der Twer, T. (2003). The pleistochrome: Optimal opponent codes for natural colours. In D. Heyer & R. Mausfeld (Eds.), *Colour perception: From light to object*. London: Oxford University Press.
- Maddess, T., McCourt, M. E., Blakeslee, B., & Cunningham, R. B. (1988). Factors governing the adaptation of cells in area-17 of the cat visual cortex. *Biological Cybernetics*, 59, 229–236.
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255–260.
- McKone, E., Aitkin, A., & Edwards, M. (2005). Categorical and coordinate relations in faces, or Fechner's law and face space instead? *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1181–1198.
- McKone, E., & Boyer, B. L. (2006). Sensitivity of 4-year-olds to featural and second-order relational changes in face distinctiveness. *Journal of Experimental Child Psychology*, 94, 134–162.
- McKone, E., & Edwards, M. (in preparation). The importance of external face contour in coding internal feature position.
- McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Towards a new theory of holistic processing. *Psychonomic Bulletin and Review*, 16(5), 778–797.
- Movshon, J. A., & Lennie, P. (1979). Pattern-selective adaptation in visual cortical neurons. *Nature*, 278, 850–852.
- Muller, K.-M., Wilke, M., & Leopold, D. A. (2009). Visual adaptation to convexity in macaque area V4. *Neuroscience*, 161(2), 655–662.
- Ng, M., Boynton, G. M., & Fine, I. (2008). Face adaptation does not improve performance on search or discrimination tasks. *Journal of Vision*, 8(1), 1–20.
- Nishimura, M., Maurer, D., Jeffery, L., Pellicano, E., & Rhodes, G. (2008). Fitting the child's mind to the world: Adaptive norm-based coding of facial identity in 8-year-olds. *Developmental Science*, 11(4), 620–627.
- Orban de Xivry, J.-J., Ramon, M., Lefèvre, P., & Rossion, B. (2008). Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. *Journal of Neuropsychology*, 2, 245–268.
- Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze

- direction. *Proceedings of the Royal Society of London, Series B*, 223(1232), 293–317.
- Regan, D., & Hamstra, S. J. (1992). Shape discrimination and the judgement of perfect symmetry: Dissociation of shape from size. *Vision Research*, 32(10), 1845–1864.
- 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., Maloney, L. T., Turner, J., & Ewing, L. (2007). Adaptive face coding and discrimination around the average face. *Vision Research*, 47, 974–989.
- 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(6), 558–566.
- Rhodes, G., Jeffery, L., Watson, T. L., 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 after-effects in high-level vision* (pp. 213–240). Oxford: Oxford University Press.
- Robbins, R., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: Adaptor position effects and neural models. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3), 570–592.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1193–1216.
- Sirovich, L., & Kirby, M. (1987). A low dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3), 519–524.
- Susilo, T., Crookes, K., McKone, E., & Turner, H. (2009). The composite task reveals stronger holistic processing in children than adults for child faces. *PLoS ONE*, 4(7), e6460.
- Suzuki, S. (2005). High-level pattern coding revealed by brief shape aftereffects. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and after-effects in high-level vision* (pp. 135–172). Oxford: Oxford University Press.
- Tanaka, J. W., & Corneille, O. (2007). Atypicality bias in face and object perception: Further tests of an attractor model. *Perception & Psychophysics*, 69, 619–627.
- Tsao, D. Y., & Freiwald, W. A. (2006). What's so special about the average face? *Trends in Cognitive Sciences*, 10(9), 391–393.
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71–84.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology A*, 43(2), 161–204.
- Valentine, T., & Bruce, V. (1986). The effect of distinctiveness in recognising and classifying faces. *Perception*, 15, 525–535.
- 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.
- Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception*, 6(3), 313–326.
- Watson, T. L., & Clifford, C. W. G. (2003). Pulling faces: An investigation of the face-distortion aftereffect. *Perception*, 32, 1109–1116.
- Watson, T. L., & Clifford, C. W. G. (2006). Orientation dependence of the orientation-contingent face aftereffect. *Vision Research*, 46(20), 3422–3429.
- Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin and Review*, 6(4), 647–653.
- Webster, M. A., Werner, J. S., & Field, D. J. (2005). Adaptation and the phenomenology of perception. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and after-effects in high-level vision* (pp. 241–277). Oxford: Oxford University Press.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: 1. Fitting, sampling, and goodness-of-fit. *Perception & Psychophysics*, 63(8), 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: 2. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63(8), 1314–1329.
- Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research*, 42, 2909–2923.
- Yamashita, J. A., Hardy, J. L., DeValois, K. K., & Webster, M. A. (2005). Stimulus selectivity of figural aftereffects for faces. *Journal of Experimental Psychology: Human Perception and Performance*, 31(3), 420–437.
- Zhao, L., & Chubb, C. (2001). The size-tuning of the face-distortion after-effect. *Vision Research*, 41, 2979–2994.