Vision Research 49 (2009) 1877–1885

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The composite illusion requires composite face stimuli to be biologically plausible

Jessica Taubert *, David Alais

School of Psychology, The University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 22 January 2009
Received in revised form 28 April 2009

Keywords:
Face perception
Holistic processing
Composite effect

ABSTRACT

Composite stimuli are whole faces comprised of two halves taken from different individuals. When asked to decide if two identical top halves are the 'same', subjects are more accurate (or faster to respond) in misaligned trials, than in aligned trials. This performance advantage for misaligned trials is referred to as the composite face effect (CFE). The proposed explanation is that aligned features are automatically fused together and form a global identity that interferes with the recognition of smaller components (the composite face illusion, CFI). However, when composite faces are misaligned, it appears to be much easier to ignore the identity of the whole face and process individual features. Here we are interested in why misalignment impairs holistic face perception. In Experiment 1 we tested the difference between horizontal and vertical misalignment and found that holistic interference persists when the vertical distance between features is increased. Is this because vertical misalignment leaves features in the correct vertical arrangement, or because vertically stretched faces are biologically plausible? Experiment 2 tested the difference between these two accounts by measuring the CFE when the two halves of a composite face were separated in stereo-depth and demonstrates that vertical symmetry alone is not sufficient for holistic processing. However, when the faces were slanted through stereo-depth (to an equivalent extent), subjects continued to be inaccurate. Overall, these experiments provide strong evidence that the composite illusion depends on biological plausibility in that the faces must be globally coherent.

1. Introduction

The appearance of a discrete facial feature is changed by the presence of other facial features. This interdependence has been attributed to holistic processing and neatly accounts for both the observation that a whole face can be more easily recognized than a feature (the part whole effect: Davidoff & Donnelly, 1990; Donnelly & Davidoff, 1999; Tanaka & Farah, 1993; also see Leder & Carbon, 2005) and the observation that a whole face will interfere with the recognition of a feature (the composite face illusion). In the original investigation of the composite face illusion (CFI), subjects found it more difficult to name the target half of a familiar face when it was paired with a distractor half taken from a different face (Young, Hellawell, & Hay, 1987). Since then, a number of studies have found that two identical top halves of a face (from the horizontal midline up) are more likely to be judged as 'different', when they are presented with different bottom halves (de Heering, Rossion, Turati, & Simion, 2008; Goffaux & Rossion, 2006; Hole, 1994; Hole, George, & Dunsmore, 1999; Le Grand, Mondloch, Maurer, & Brent, 2004; McKone, 2008; Robbins & McKone, 2003, 2007; Rossion & Boremanse, 2008). When the top half of a composite face is shifted away from the bottom half, along the common horizontal axis, the interference caused by the CFI is reduced. Increased accuracy (or decreased reaction time) in these misaligned trials, compared to their aligned counterparts, is a reliable face perception phenomenon that is commonly known as the composite face effect (CFE).

It is generally presumed that misalignment reduces holistic interference because it fractures what all faces have in common, their first-order configuration (see Maurer, Le Grand, & Mondloch, 2002; McKone, Kanwisher, & Duchaine, 2007). All faces have the same features, arranged in a common spatial layout. For example, the eyes are always above the nose. In this context, the term 'first-order configuration' merely refers to the presence of the correct facial features, in their correct positions, relative to each other. Thus, when the features of a composite face are presented in the aligned format, they form a pattern that satisfies the first-order configuration of a whole face and triggers holistic face processing. Features presented in the misaligned format, however, are not processed holistically because they break the first-order configuration of a face and as a result feature processing improves (de Heering et al., 2008; Goffaux & Rossion, 2006; Hole, 1994; Hole et al., 1999; Le Grand et al., 2004; McKone, 2008; Robbins & McKone, 2003, 2007; Rossion & Boremanse, 2008).

Here we reason that if the first-order configuration of a face induces holistic interference, then the CFE would be dependent on misalignment being horizontal. That is, if the upper and lower face

---

* Corresponding author.
E-mail address: jtaubert@psych.usyd.edu.au (J. Taubert).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.04.025

halves were displaced vertically, the first-order configuration would remain essentially unchanged (i.e., there would still be two eyes above a nose and a mouth). Furthermore, using stereopsis to misalign the features of a composite face in depth would leave the first-order configuration intact and, thus, not result in a CFE. Alternatively, if misalignment *per se* were ultimately responsible for overcoming holistic interference, then any misalignment of features (whether two- or three-dimensional) would result in the efficient processing of local features. We test these predictions in two experiments.

2. Experiment 1

A recent investigation of spatial relationships between facial features concluded that vertical and horizontal relationships between features were dissociable (Goffaux & Rossion, 2007). Sensitivity to vertical relationships decreased when face stimuli were turned upside down, whereas, sensitivity to horizontal relationships and feature details remained relatively high. Given that picture-plane inversion is widely known to disrupt face recognition (Diamond & Carey, 1986; Moscovitch, Winocur, & Behrmann, 1997; Rossion & Gauthier, 2002; Yin, 1969), Goffaux and Rossion (2007) provide evidence of a strong link between that the vertical relationships that exist between features and the discrimination of face stimuli (also see Goffaux, 2008). Furthermore, Dakin and Watt (2009), who were recently interested in evaluating the contribution of orientation information to the structure of faces, concluded that the vertical alignment of features forms part of a reliable "bar code" that could be used to distinguish faces from other nonface stimuli.

Accepting that the vertical organization of facial features makes a dissociable contribution to face perception holds potentially important implications for the CFE because horizontal misalignment disrupts the vertical symmetry of features. Experiment 1 was designed to test the difference between horizontal and vertical misalignment. Drawing on a large body of previous research (see McKone et al., 2007) the "horizontal misaligned" condition is expected to result in a performance advantage for matching a half-face when compared to the "horizontal aligned" condition, due to the reduction of holistic interference (see examples in Fig. 1a). There are, however, divergent predictions for the "vertical misaligned" condition. If the minimum requirement for holistic interference is the first-order configuration of a face, then the CFI should persist in the vertical misaligned condition because the top half is merely moved up from the bottom half and vertical organization of facial features is preserved (Fig. 1a). Thus, there should be no difference in accuracy or reaction time between the aligned and vertically misaligned trials. Alternatively, if misalignment of any kind suffices to reduce holistic interference, then better performance would be predicted in both the horizontally and vertically misaligned conditions.

2.1. Methods

2.1.1. Subjects

Thirty three undergraduate students from Macquarie University participated in this experiment for course-credit (seven males). All

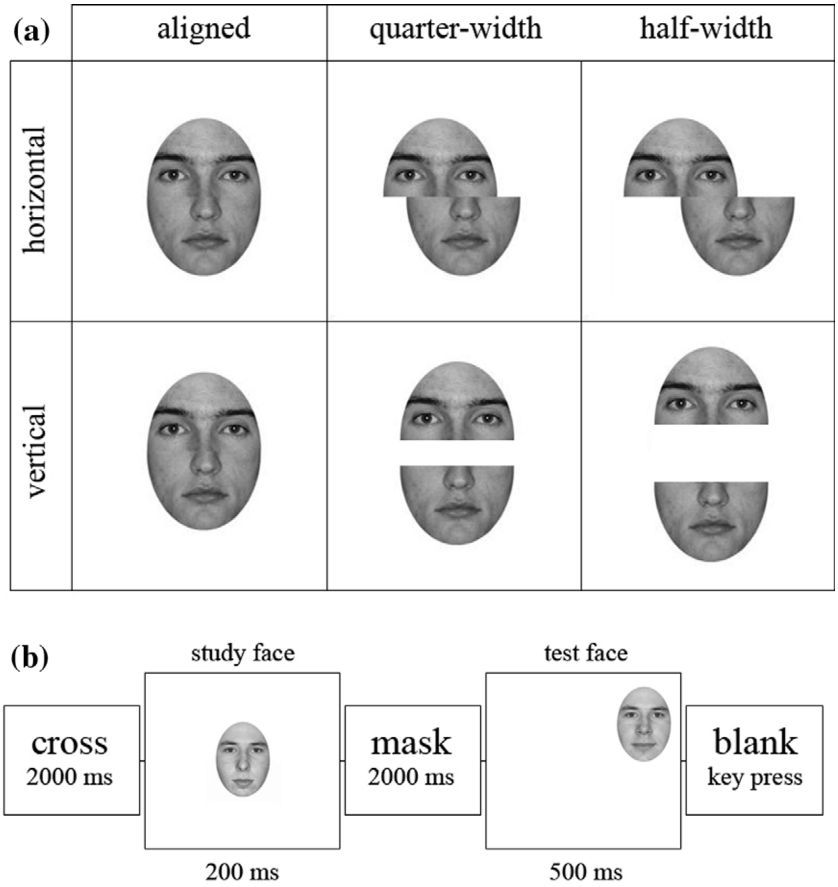

Fig. 1. Experiment 1: (a) examples of the experimental stimuli used, presented as a function of misalignment axis and misalignment distance and (b) an illustrative example of a 'same' trial.

subjects were aged between 18 and 25 years. They all had normal or corrected-to-normal vision.

2.1.2. Stimuli

For Experiment 1, the stimulus set was comprised of 10 digital photographs taken of 10 different human faces (256 greyscale, 8-bit). The images were matched for lighting direction and image quality. Each photograph depicted a full-face view of a clean-shaven Caucasian male with a neutral facial expression (no glasses, facial hair, or distinctive blemishes). Differences in brightness and contrast were reduced as much as possible. The experimental stimuli were created in the following ways using Adobe Photoshop CS2 software (www.adobe.com).

2.1.2.1. Aligned faces. The 10 faces were cropped using a rectangle tool, placed on a white canvas, and resized so that they were all precisely 170 pixels wide. The top half of each face, from the horizontal mid-line up, was then cropped using the rectangle tool, with a hard edge, and placed on a new canvas. The bottom halves were reassigned to different top halves. Thus, 10 novel exemplars were created from rejoining the parts of the original 10 faces. These will be from now on referred to as the 'study' stimuli (see Fig. 1). To create the 'test' stimuli, an additional 10 composite faces were then formed via a repeated process of reassignment; the upper halves were, again, isolated and paired with another lower half. The 10 composite faces that comprised the 'test' stimuli were then changed in one final but important way.

2.1.2.2. Quarter-width faces. The 10 study and 10 test stimuli that were described above, were transformed into the quarter-width faces for both the horizontal and vertical condition. All quarter-width faces had their upper and lower halves separated by the width of a quarter face (42 pixels). For the horizontal misaligned condition, all 20 aligned composite faces were copied and then their two halves were separated along the common horizontal axis so that the upper half was shifted to the left of the lower half. The lateral offset of the two halves represents traditional misalignment. For vertical condition, however, the misalignment was vertical; the entire upper half was shifted up by 42 pixels (see Fig. 1b).

2.1.2.3. Half-width faces. The half-width faces, in both the horizontal and vertical condition, were created using the same process as the quarter-width faces. The only difference was that the distance was increased to 85 pixels (see Fig. 1b).

2.1.3. Design

The experimental design had a 2 × 3 repeated-measures structure, with every subject completing all six unique conditions in a single experimental block. The first factor, referred to as axis, had two levels (horizontal v vertical). The second factor described the distance between the two halves (aligned v quarter-width v half-width). Three levels of width were incorporated in the design because it was not clear from the available literature what the optimal width would be.

The experiment was comprised of 600 trials in total, of which 300 were 'same' trials. There were 50 'same' trials that were generated, per condition, when the 10 study stimuli were repeated five times. For these 'same' trials, the associated test stimulus was a composite face with the same upper half as the study face, but a different lower half. There were an equal number of 'same' and 'different' trials. For the 'different' trials the test stimuli were reassigned so that the task-relevant information, in the upper half changed between the study and test phase. Therefore, across the entire experiment, the 10 study stimuli were seen a total of 60 times (30 'same'; 30 'different'), and the same was true of the test stimuli.

2.1.4. Procedure

The experiment was programmed using Superlab 4 (OS X) software. The visual stimuli were presented on a Sony Tritron (RBG) CRT monitor being driven by a Macintosh G4 powerPC processor. Subjects were positioned in front of the monitor with a viewing distance of approximately 45 cm. Before the experiment began, subjects were instructed to complete a sequential same/different task. Accordingly, the subjects were asked to judge whether the upper half of any test face was the 'same' as, or 'different' from, the preceding study face. They were to indicate their decision using two keys on a USB keyboard (v for 'same'; m for 'different'). Subjects were asked to make their decisions as quickly and accurately as possible. There were no practice trials before the experiment began and the subjects were given no feedback during the experiment. Every experimental trial began with the presentation of a central fixation cross for 2000 ms. When the fixation cross disappeared it was immediately followed by the study phase, where the subject saw a composite face for 200 ms (no response was required from the subject at this time). The study face was always positioned in the centre of the monitor. The onset of the test phase was subsequent to the presentation of the visual mask (made up of randomly jumbled facial features taken from the images of faces not used in the experiment), which was visible for 1000 ms. The test stimulus would appear for 500 ms, in one of 10 possible screen locations. The position was randomized across the experimental trials. The test phase could only be terminated by a key press, so following the presentation of the test face, subjects would sit in front of a blank screen until they responded. The computer recorded the time between the onset and termination of the test phase, in addition to subject accuracy. Order of trial presentation was fully randomised across subjects.

2.2. Results and discussion

2.2.1. Accuracy

2.2.1.1. Same trials. A summary of the data collected during Experiment 1 is provided in Table 1. A two-way ANOVA model was used to analyse the axis (2) × distance (3) design, where both factors were manipulated within subjects. Collapsing the three levels of distance, identical top halves in the horizontal condition were matched more accurately than identical top halves in the vertical condition ($F(1,32) = 46.13$, $MSE = 0.70$, $p < 0.01$). The manipulation of distance (aligned v quarter-width v half-width) also had a significant impact on accuracy ($F(3,64) = 87.59$, $MSE = 0.90$, $p < 0.01$). Critically, the overall analysis yielded a significant interaction between axis and distance ($F(2,64) = 10.12$, $MSE = 0.70$, $p < 0.01$).

Subsequent to the overall analysis, a series of simple contrasts (two tailed) were run, designed to isolate any significant CFEs. The reported p-values have been appropriately adjusted for planned comparisons using the Bonferroni decision rule. In the horizontal condition, it was found that shifting the top half of a face made the matching task easier, regardless of how far the face was shifted (aligned face v quarter-width, $t(32) = 8.70$, $p < 0.01$; aligned face v half-width, $t(32) = 18.87$, $p < 0.01$). Consistent with the prediction that there would be a CE found in the vertical shift condition as it would approximately preserve the first-order configuration, there was no evidence of improved performance between the aligned and quarter-width vertical-shift trials ($t(32) = 2.02$, $p > 0.05$). However, subjects had less difficulty matching top halves during the half-width vertical-shift condition compared to the aligned condition ($t(32) = 10.42$, $p < 0.01$). There was no evidence of a difference between the two aligned conditions (horizontal aligned v vertical aligned, $t(32) = 1.93$, $p > 0.05$), nor between the two half-width conditions (horizontal half-width v vertical half-width, $t(32) = 2.15$, $p > 0.05$). Subjects were, however,

Table 1
A summary of the results of Experiment 1 ($N = 33$): Mean accuracy (in percent correct) and mean correct response times (in ms) are given for all experimental conditions. Significant differences between conditions are indicated.

	Distance	Axis		
		Horizontal (H)	Vertical (V)	H – V
Mean accuracy in ‘same’ trials (±*SEM*)	Aligned (A)	58.61 (2.28)	54.12 (1.86)	4.49 (2.32)
	Quarter-width (Q)	74.36 (1.47)	58.79 (1.71)	15.58 (1.96)[**]
	Q – A	15.76 (1.81)[**]	4.67 (2.31)	
	Half-width (H)	80.30 (1.60)	76.30 (1.12)	4.00 (1.86)
	H – A	21.70 (2.83)[**]	22.18 (2.13)[**]	
Mean correct reaction time in ‘same’ trials (±*SEM*)	Aligned (A)	976.18 (28.61)	1007.39 (24.19)	–31.21 (28.14)
	Quarter-width (Q)	878.07 (21.25)	985.46 (20.49)	–107.40 (23.18)[**]
	Q – A	–98.12 (27.50)[**]	–21.93 (25.43)	
	Half-width (H)	795.80 (22.76)	797.84 (24.09)	–2.03 (22.91)
	H – A	–180.38 (24.66)	–209.56 (30.18)	

[*] $p < 0.05$.

[**] $p < 0.01$.

more accurate in the quarter-width horizontal-shift condition than in the quarter-width vertical-shift condition ($t(32) = 7.96$, $p < 0.01$).

2.2.1.2. Different trials. Although there was no experimental prediction for the ‘different’ trials, the means were inspected because systematic variation across the ‘different’ trials might suggest a signal-detection analysis was appropriate. A 2 × 3 overall ANOVA test was again used to analyse the data because the structure of the ‘different’ trials was identical to the structure of the ‘same’ trials. None of the main effects approached significance (axis, $F(1, 32) = 0.10$, $MSE = 0.07$, $p = 0.75$; distance, $F(2, 64) = 0.91$, $MSE = 0.09$, $p = 0.41$) nor was there any evidence of an interaction between axis and distance ($F(2, 64) = 0.62$, $MSE = 0.07$, $p = 0.54$).

2.2.2. Reaction time

Correct reaction times collected during the ‘same’ trials were also analysed using an independent 2 × 3 repeated-measures ANOVA model (see Table 1). Averaging across distance, subjects were faster to respond to the horizontal trials than the vertical trials ($F(1, 32) = 10.73$, $MSE = 10136.75$, $p < 0.01$). The overall analysis also revealed that distance had a significant impact on reaction time, averaging across the two levels of axis ($F(2, 64) = 51.92$, $MSE = 12673.93$, $p < 0.01$). The key interaction between axis and distance was significant ($F(2, 64) = 4.77$, $MSE = 10226.49$, $p < 0.05$). Importantly, the differences in the reaction-time data were inconsistent with a speed/accuracy trade off.

Planned pair-wise comparisons revealed that, in the horizontal condition, misalignment generally allowed subjects to respond correctly faster than when the composite faces were aligned (aligned face v quarter-width, $t(32) = 3.57$, $p < 0.01$; aligned face v half-width, $t(32) = 7.31$, $p < 0.01$). When the composite faces belonged to the vertical condition, there was no evidence of an advantage for quarter-width trials over aligned trials ($t(32) = 0.86$, $p > 0.1$). Subjects were faster to accurately match vertical half-width composites compared to the vertical aligned composites ($t(32) = 6.94$, $p < 0.01$). Consistent with the accuracy data, the reaction-time data also revealed that there was no difference between the two aligned conditions ($t(32) = 1.11$, $p > 0.1$). The correct responses were recorded just as quickly in the two half-width conditions ($t(32) = 0.09$, $p > 0.5$). When the two halves of the composite faces were separated by a quarter-width, subjects responded significantly faster in the horizontal condition than the vertical condition ($t(32) = 4.63$, $p < 0.01$).

3. Experiment 2

Holistic processing of composite faces is impaired when two facial halves are presented in the misaligned format (see Maurer et al., 2002; McKone et al., 2007). Experiment 1 extended this accepted finding by investigating the potential difference between horizontal and vertical misalignment. The axis manipulation influenced the ease with which subjects were able to respond to the feature-matching task. In the horizontal condition, the two halves were manipulated in the standard way; misalignment was achieved by shifting the top half of a composite face along the common horizontal axis. This manipulation allowed subjects to match more easily the task-relevant information. In contrast, there was evidence of continued holistic interference when the top half was presented directly above the bottom half, separated by a vertical distance (the width of a quarter-face). We note that performance dramatically improved in the half-width vertical condition when compared to the aligned control implying that a vertical separation also breaks the global configuration of a face when the distance between faces is unrealistic compared to anthropomorphic norms. What these results show, therefore, is that holistic processing is more tolerant of vertical shifts than horizontal shifts, although ultimately a large vertical shift will also prevent holistic processing.

Why, then, we do not see evidence of a CFE in the quarter-width vertical condition, whereas the same distance allowed for better performance in the comparable horizontal condition? This observation refutes the assertion that the CFE is purely a consequence of misalignment. The behavioural responses of subjects to Experiment 1 suggest that the features of the vertically misaligned faces fuse together, into a single unit of analysis – provided the separation is not too great. If holistic processing occurs as part of an early, automatic response to the detection of a face-like pattern (two eyes above a mouth), then this might explain why symmetry around the vertical axis is sufficient for a holistic representation to be built (Dakin & Herbert, 1998; Rock, 1983; Scognamillo, Rhodes, Morrone, & Burr, 2003; Wagemans, 1997; also see Rhodes, Peters, Lee, Morrone, & Burr, 2005; Tsao & Livingstone, 2008). If this were true and holistic representations are built in response to the first-order configuration of a face, then using stereo-depth to misalign the features of a composite face but should preserve the CFI because the symmetry around the vertical axis would be intact.

A subtle variation of this account for the outcome of Experiment 1 is that the difference between horizontal and vertical misalignment reflects a difference in biological plausibility. Vertically misaligned faces might be considered organic because it is possible to encounter long (i.e., tall) faces. Not only is there is natural variation in the length of faces, but there are also expressions and movements during speech that temporarily stretch facial features in the vertical dimension. Thus, the mechanism responsible for building holistic representations of composite faces might be more tolerant of vertical misalignment because it easily maps onto common changes in expression or viewpoint. This explanation would also account for the observation in Experimental 1 that when a vertical distance that exceeded biological norms separated the two halves of a composite face, the illusion of a whole face was broken. We hypothesize that if holistic processing depends on biological plausibility, and features that belong to a single face are usually in the same depth plane, then, dividing a composite face across perceived depth should break holistic interference and make it easier to match identical features than when the same face is presented synoptically.

One foreseeable caveat in this line of reasoning is that individual features that appear at different depths in the visual environment will be processed by different areas of retinotopic cortex, perhaps preventing them from being integrated at a later stage. Therefore, a control condition is required to rule out this low-level interpretation of the results of Experiment 2. What is needed is evidence that holistic interference could still occur when the features of a single face are separated by stereo-depth but maintain biologically plausible global coherence. For this, we used stereoscopic slant (i.e., a gradual change of depth, rather than an abrupt one). When a face is tilted forward, the features in the top half of the face are closer than the features in the bottom half but the continuity and coherence of the face is maintained and can therefore be considered as biologically plausible. If we assume that holistic processing acts on biologically plausible faces, the features that belong to an aligned composite face that is slanted forward (or backward) in stereo-depth should be strongly integrated, despite differences between the stereo-depth of the discrete facial features.

3.1. Methods

3.1.1. Subjects

There were 18 volunteers tested during Experiment 2 (six male). 17 subjects were naïve to the experimental objectives. To ensure independence, none of the subjects from Experiment 1 were recalled to participate in Experiment 2. The mean age of a subject was 29.5 years and all had normal, or correct-to-normal vision.

3.1.2. Visual stimuli

3.1.2.1. Faces. The digital photographs of five Caucasian males were taken under similar lighting conditions. These digitized images were then greyscaled (256; 8-bit) and matched for brightness and contrast as much as possible. The size of the faces were resized such that the same facial features fit into a standard oval region (170 pixels wide). To create the aligned composite faces, the top half of each of the five exemplars was lifted and paired with each of the 4 other bottom halves taken from the remaining four faces. Thus, there were 20 aligned composite faces in total. Unlike in Experiment 1, the position of the composite faces was constant during the experiment because of the limitations imposed by the stereoscope. To compensate for the constant position of the stimuli, a second set of aligned composite faces was created for the test phase. Identical copies of the original 20 composite faces were made and then changed so that the faces that appeared in the study and test phase would differ in their level of brightness by a random amount ('random' differences were determined by www.random.com).

These 40 composite faces were then imported into Matlab (version 7.4.0) to soften the outer edge of the oval cut stimuli. A horizontal black line (six pixels wide) was superimposed across the front of the stimuli to conceal the join. The composite faces were then each placed in the centre of a much bigger visual display (1024 × 786 pixels in size) and copied to create identical stereo pairs. The faces that belonged to an identical pair were then positioned so that they appeared at the same height and were distributed evenly along the horizontal midline of the display (135 pixels apart). A white square reference frame was placed around the faces to assist with the fusion of the stereo pairs.

3.1.2.2. Synoptic (aligned) faces. For the synoptic condition, the visual displays were not manipulated and, as such, subjects were presented with the same image in both eyes (with no disparity). Hence there was no stereoscopic information at all in the synoptic condition and the images appeared to lie on a single fronto-parallel plane.

3.1.2.3. Stereoscopic faces: split-depth. In the uncrossed and crossed conditions, the upper and lower face halves appeared to lie on different depth planes. This was achieved taking the visual displays from the synoptic condition and laterally shifting the upper half of each eye's face either closer together (crossed disparity) or further apart (uncrossed disparity) (see Fig. 2). In the crossed condition, the upper half appeared to lie on a nearer plane, and in the uncrossed condition the upper half appeared to lie on a farther plane, with the bottom half at zero disparity. Two levels of horizontal disparity (and therefore depth separation) were used, six and 12 pixels. Note that the faces themselves were not presented with stereoscopic three-dimensionality, but rather as two-dimensional flat images with the upper and lower halves presented at different depths. Note also that the small horizontal offset visible in the monocular images is not perceived in the binocularly viewed image: the stereo image is perceived as perfectly aligned but with the upper half at a different depth plane from the lower half. This stereoscopic condition will be referred to as the 'split-depth condition'.

3.1.2.4. Stereoscopic faces: stereo-slant. The second stereoscopic condition will be referred to as the 'stereo-slant condition' because the illusion of depth was created via an affine transformation using Matlab software that sheared the tops of the stereo pairs either towards each other (crossed) or away from each other (uncrossed). The degree of the shear was carefully calibrated to create conditions where the disparity of the eyes (the most salient features in the top half of the face) were disparate by amounts of either six or 12 pixels (either crossed or uncrossed) to match the disparities used in the split-depth conditions. When viewed as a fused stereo image through the stereoscope, these pairs of tilted images appear as a single perfectly upright, two-dimensional face but tilted in depth (i.e., stereo slant), appearing to lean either towards the viewer (if the upper halves are closer) or away from the viewer. The affine transformation was chosen because it preserves the internal geometry of the images and therefore maintains the relationships between the facial features. Although the result is a slightly elongated image, this is does not pose a problem because it has been shown that geometric distortions which stretch a face vertically do not interfere with face recognition (Hole, George, Eaves, & Rasek, 2002).

3.1.3. Design

A factorial repeated measures design was implemented. For both kinds of stereoscopic information (split-depth v stereo-slant) there were five levels of disparity (uncrossed 12 v uncrossed 6 v synoptic v crossed 6 v crossed 12). Both the uncrossed (u) and crossed (c) conditions were necessary to ensure that the results could be contributed to the break in stereo-depth and not, more simply, the salience of features that are perceived as being closer. Two levels of disparity were introduced into the design, namely six and 12 pixels, because of the uncertainty surrounding when subjects would experience diplopia. The shifted and tilted trials were blocked separately. The order that these blocks were completed was counterbalanced across subjects.

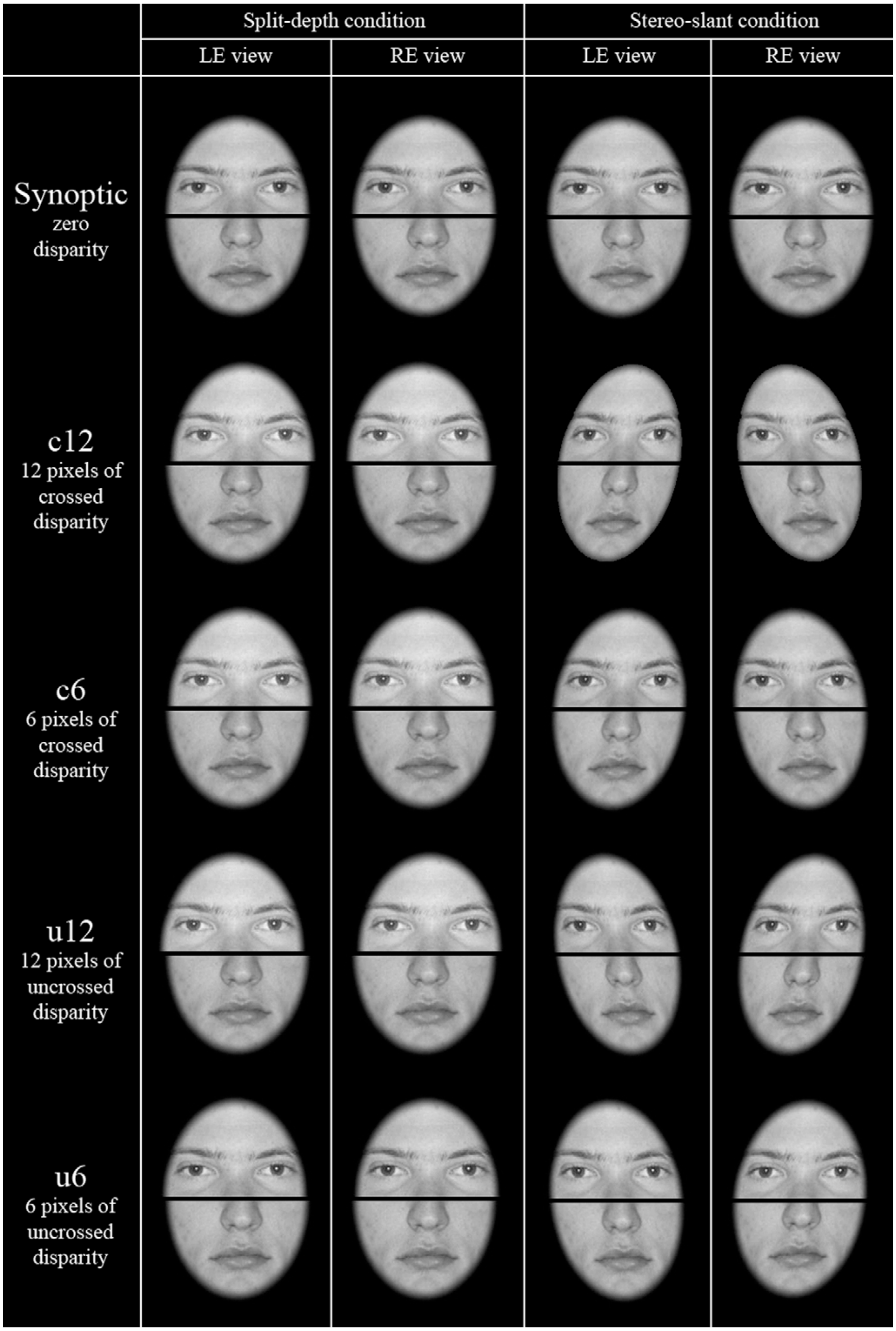

Fig. 2. Experiment 2: an example of a stereo pair from each of the 10 unique conditions.

Within each block, subjects were required to complete five conditions that were separated by a short break (c12, c6, synoptic, u6, u12). The order that these five conditions were completed was pseudo randomised to ensure that every subject completed the 10 total conditions in a different order. Each condition was comprised of 40 discrete trials. For 20 of these trials the correct response was ‘same’.

3.1.4. Procedure

For subjects, the procedure during Experiment 2 was the same as the sequential same/different procedure described for Experiment 1, with the exception that during the Experiment 2 the visual displays were presented on a 20” Mitsubishi diamond CRT monitor and viewed through a Stereo Aids ScreenScope stereoscope (www.stereoaids.com.au). The stereoscope provided a nose rest that helped maintain the viewing distance at 45 cm. Instead of a fixation cross, which is difficult to fuse, each trial in Experiment 2 began with the presentation of a green oval region in the centre of the fusion frame, which was replaced after 2000 ms with a study face, which was only visible for 200 ms. The study phase was followed by a grey oval region for 2000 ms, which was, in turn, replaced by the test phase. In the test phase of any given trial, the associated test face appeared briefly, for 200 ms, before a red oval shape signaled the opportunity to respond.

3.2. Results and discussion

3.2.1. Accuracy

3.2.1.1. Same trials. Mean accuracy (% correct) across subjects was analysed using a 2×5 repeated-measures ANOVA procedure (see Fig. 3 for a summary of the results). Averaging across the five levels of disparity, there is evidence to suggest that subjects were more accurate during the split-depth condition ($M = 72.5$, $SEM = 2.3$) than the stereo-slant condition ($M = 55.7$, $SEM = 3.1$; $F(1, 17) = 35.04$, $MSE = 0.04$, $p < 0.001$). The significant main effect of disparity ($F(4, 68) = 7.23$, $MSE = 0.01$, $p < 0.001$) appears to have been driven by the difference between the aligned condition ($M = 56.8$, $SEM = 3.5$) and the 4 disparate conditions (c12, $M = 66.4$, $SEM = 2.1$; c6, $M = 65.6$, $SEM = 1.9$; u6, $M = 64.4$, $SEM = 2.7$; u12, $M = 67.3$, $SEM = 3.1$). The interaction between stereo information and disparity was also significant ($F(4, 68) = 11.53$, $MSE = 0.01$, $p < 0.001$) indicating that the pattern of differences across the levels of disparity depended on the stereo manipulation.

Eight planned comparisons were run on the accuracy data to determine the outcome of the experimental hypotheses. These comparisons were calculated within the interaction family (using the error term associated with the interaction family), and were thus adjusted appropriately using the Bonferroni rule. One hypothesis predicted that, for the split-depth and slanted stereo conditions, alike, there would be no difference between the crossed and uncrossed conditions. However, for it to be appropriate to test the difference between crossed and uncrossed, there should be no difference across the levels of disparity. It was found that whether the faces belonged to a crossed or uncrossed condition, there was no difference between 6 and 12 pixels of disparity (split-depth c12 v c6, $t(17) = 0.68$, $p > 0.5$; split-depth u12 v u6, $t(17) = 2.78$, $p > 0.1$; stereo-slant c12 v c6, $t(17) = 0.68$, $p > 0.5$; stereo-slant u12 v u6, $t(17) = 0.91$, $p > 0.5$). Therefore, the means for the two levels of disparity were averaged together and compared. When matching the split-depth composites, there was no evidence of difference between the crossed and uncrossed condition ($t(17) = 1.12$, $p > 0.5$). There was no comparable difference in the stereo-slant data ($t(17) = 1.02$, $p > 0.5$). Finally, the most important contrasts for the experimental hypotheses were the comparisons between the synoptic and stereoscopic conditions. These comparisons revealed that in the split-depth condition, subjects found it easier to match the task relevant features when they were presented in a different depth plane from the bottom half, compared to in the synoptic condition when all features were presented in a single plane of depth ($t(17) = 5.51$, $p < 0.01$). This was the predicted result if the correct first-order configuration of a face was necessary, but not sufficient for the CFE. The same was not true when the stereo pairs were slanted as to create the perception of a composite face leaning towards (or away). In the stereo-slant condition, subjects were as accurate in stereo conditions as they were in the synoptic condition ($t(17) = 0.90$, $p > 0.5$). This result helps to rule out a low-

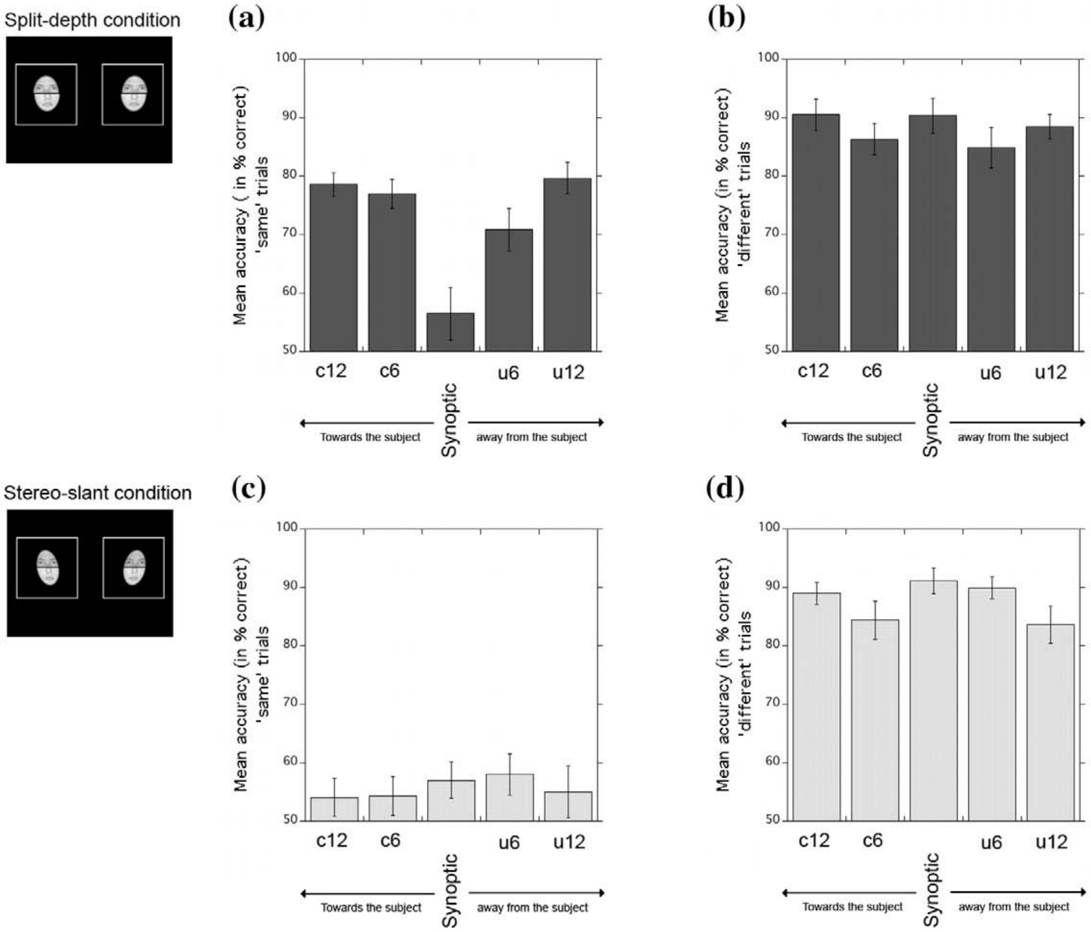

Fig. 3. The results for Experiment 2: (a) the mean accuracy for all split-depth conditions, 'same' trials only; (b) the mean accuracy for all split-depth conditions, 'different' trials only; (c) the mean accuracy for all stereo-slant conditions, 'same' trials only; and (d) The mean accuracy for all stereo-slant conditions, 'different' trials only.

Table 2
A summary of the correct reaction times for all experimental conditions ($N = 18$).

	Disparity	Stereo condition	
		Split-depth	Stereo-slant
Mean correct reaction time (±*SEM*) 'same' trials only	c12	796.3 (80.1)	939.9 (97.9)
	c6	891.4 (85.0)	982.3 (98.1)
	Synoptic	868.6 (64.1)	894.6 (85.3)
	u6	811.5 (74.0)	913.1 (83.5)
	u12	855.4 (77.6)	947.2 (91.1)

level explanation for the outcome of the split-depth condition, because although the features were presented at different depths, they were still globally integrated into a single unit of analysis, thereby causing holistic interference.

3.2.1.2. Different trials. An overall analysis was also run on mean accuracy in the 'different' trials. As expected, performance was generally very good when the task-relevant information differed between the study and test phase (c12, $M = 89.7$, $SEM = 1.5$; c6, $M = 85.4$, $SEM = 2.5$; synoptic, $M = 90.8$, $SEM = 2.2$; u6, $M = 87.4$, $SEM = 2.3$; u12, $M = 86.0$, $SEM = 2.1$). The main effects of stereoscopic information ($F(1, 17) = 0.06$, $MSE = 0.02$, $p > 0.5$) and disparity ($F(4, 68) = 2.38$, $MSE = 0.01$, $p > 0.05$) were not significant. Importantly, there was also no evidence of an interaction between the two factors ($F(4, 68) = 1.50$, $MSE = 0.01$, $p > 0.1$).

3.2.2. Reaction time

When a correct response was made to a 'same' trial, the response time was recorded. The reaction-time data was analysed using the appropriate repeated-measures ANOVA model for a 2×5 factorial design. The results are summarised in Table 2. The overall ANOVA suggests that there was no systematic variation in reaction time across the experimental conditions. The main effects of stereoscopic information and disparity, together with the interaction effect, were not found to be significant (stereoscopic information, $F(1, 17) = 1.82$, $MSE = 203588.90$, $p > 0.1$; disparity, $F(4, 68) = 1.60$, $MSE = 20411.30$, $p > 0.1$; interaction, $F(4, 68) = 0.78$, $MSE = 20623.78$, $p > 0.5$). The mean reaction times that were recorded across the 10 unique conditions were highly homogeneous. The grand mean was 890.04 ms ($SEM = 71.13$) with all cell means falling within two standard errors (see Table 2). Upon closer inspection, there was a trend for an accurate response in the stereo-slant condition ($M = 935.4$, $SEM = 86.2$) to take longer than in the split-depth condition ($M = 844.7$, $SEM = 70.3$). This observation is of some importance because the trend is inconsistent with a speed–accuracy tradeoff; the advantage for the split-depth condition, over stereo-slant conditions, in the accuracy data was not due to longer reaction times.

4. General discussion

The CFE is an accepted marker of holistic interference and shows that local details are harder to identify when presented in a global (composite) context. The aim of Experiment 1 was to determine whether vertical or horizontal displacements of the half-faces would break the influence of global configuration. The results showed that holistic processing was reduced (i.e., identification of local features was facilitated) when the half-faces were horizontally displaced by as little as a 1/4 face-width. In contrast, vertically displacing the half-faces by the same amount was not sufficient to break the influence of the global context. This provides strong evidence for the previous assertion that the vertical and horizontal relations between face features do not make equally important contributions to upright face perception (Goffaux, 2008; Goffaux & Rossion, 2007). Our result implies that symmetry around the vertical axis is a highly discriminating feature for the visual system in detecting faces, as even small horizontal displacements reduce holistic interference attributed to the CFI. The reason for this might be that holistic face processing requires what is common to all faces, a symmetrical pattern of facial features (see Dakin & Watt, 2009; Rhodes et al., 2005; Tsao & Livingstone, 2008). Small vertical displacements, on the other hand, will distort a face but still maintain its vertical organization and its ability to produce holistic interference. Large vertical displacements (1/2 face width) will break holistic interference as Experiment 1 shows, presumably because the feature spacing becomes implausibly large given anthropometric norms.

The CFE was further explored in Experiment 2 where we used binocular disparity to test the effect of displacement of the half-faces in the depth dimension. This manipulation is of interest because depth displacements preserve both horizontal and vertical relations between face features, as the half-faces remain unchanged two-dimensionally in the picture plane and differ only in depth. The results clearly demonstrated that simply having two eyes above a mouth (i.e., first-order configuration) is not sufficient for holistic processing. In the split-depth stereo condition, despite the vertical and horizontal organization of the composite faces remaining intact, there was a significant improvement in performance when the task relevant half-face was presented in a different depth plane from the irrelevant half-face, compared to the synoptic condition where both half-faces were presented in the same depth plane. This advantage was independent of the amount and the sign of disparity, as both six and 12 pixel disparities, in the crossed or uncrossed direction, were found to break the influence of holistic processing and significantly reduce the CFI. This indicates that holistic face processing is sensitive to ecologically invalid distances between face features. More fundamentally, the results of this experiment suggest that holistic processing is bound by a set of rules that extends beyond simple (two-dimensional) first-order configuration.

Experiment 2 also included a stereo-slant condition in which the faces were tilted forward or backward in depth. Although the degree of disparity increases continuously from the face's horizontal mid-point, this condition was calibrated so that the eyes were disparate by either six or 12 pixels (to match the split-depth condition). The main difference with the split-depth condition is that in the stereo-slant condition the disparity varies continuously so that the face is globally coherent and should therefore remain a plausible face. It appears that the holistic processes underlying face perception are sensitive to this as holistic interference was evident in both the six and 12 pixel disparity stereo-slant conditions, for both forward and backward slant. This finding is significant in two ways. First, it rules out the low-level explanation of the split-depth stereo condition that holistic interference failed to occur simply because the face features were presented at different depths. Second, it is consistent with the assumption that holistic processing depends on the plausibility of the stimulus: two half-faces presented at different depths clearly cannot belong to the same face, whereas a continuous depth change could plausibly arise from a tilted face or a change in viewpoint.

In related research supporting the second conclusion, McKone (2008) found evidence that the CFE is still observed when changes in viewpoint around the vertical (or yaw) axis are included. Face recognition, on the other hand, is known to be susceptible to changes in viewpoint. Observers can very quickly identify the front view of faces, but they tend to take longer or be less accurate when identifying the profile view (Hill & Bruce, 1996; Hill, Schyns, & Akamatsu, 1997; McKone, 2008). The origin of this viewpoint effect is possibly due to the relative rarity of profile views, but to the best of our knowledge there is no available evidence that front views are seen more frequently than profile views. In any case, this dissociation between the CFE and face identification for changes in viewpoint holds two important implications. The first is that holistic processing is not the only factor contributing to accurate face identification (see also Moscovitch et al., 1997), and the second is that the minimum requirement for holistic interference during a composite task is not simply *two* eyes above, a nose and a mouth.

In other research relating to the plausibility of faces, it has been found that the magnitude of the CFE is increased when the faces have been low-pass filtered to remove high spatial frequencies (Goffaux & Rossion, 2006). The fact that holistic representations of face stimuli can be built in the absence of high spatial frequencies is consistent with the common experience of seeing a face approach from a distance. The amplitude of high spatial frequencies is attenuated with distance and fine spatial features become difficult to see. However, despite being blurry (low-frequency dominated) they can still be considered plausible faces because low spatial frequencies are sufficient to establish the first-order configuration of eyes, nose and mouth. A good example of an implausible face would be an upside-down face, as these are very seldom encountered. Consistent with our proposed role for plausibility in holistic face processing, the CFE is dissolved by inversion in the picture plane and performance in aligned trials improves (Hole, 1994; Robbins & McKone, 2003). The observations that the CFE is independent of both viewpoint and spatial frequency content, but dependent on canonical orientation, converges with our proposal that holistic interference requires a biologically plausible face stimulus.

In summary, we propose that biological plausibility is a good predictor of when holistic interference should be expected, however it also has a broader implication. In 2006, Goffaux & Rossion drew on neural and behavioural evidence to support their claim that holistic processing plays a role in the early stages of face recognition (possibly at the level of face detection). The advantage of early holistic processes would be that it would quickly bind facial features into a coherent global stimulus that could then be further processed by a specialized face recognition mechanism. The proposal that holistic processing gates face recognition squares with the sensitivity of the CFE to unnatural displacements of features (see Experiments 1 and 2) because accurate face detection would require extracting information that distinguishes faces from other objects (Dakin & Watt, 2009; Tsao & Livingstone, 2008) and the visual environment (Dakin & Herbert, 1998; Rhodes et al., 2005; Rock, 1983; Scognamillo, Rhodes, Morrone, & Burr, 2003; Wagemans, 1997).

References

Dakin, S. C., & Herbert, A. M. (1998). The spatial region of integration for visual symmetry detection. *Processings of the Royal Society B, 265*, 659–664.

Dakin, S. C., & Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision, 9*(4), 1–10. http://journalofvision.org/9/4/2/.

Davidoff, J., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes. *Acta Psychologica, 73*(3), 225–243.

de Heering, A., Rossion, B., Turati, C., & Simion, F. (2008). Holistic face processing can be independent of gaze behaviour: Evidence from the face composite effect. *Journal of Neuropsychology, 2*, 183–195.

Diamond, R., & Carey, S. (1986). Why faces are and not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107–117.

Donnelly, N., & Davidoff, J. (1999). The mental representations of faces and houses: Issues concerning parts and wholes. *Visual Cognition, 6*(3–4), 319–343.

Goffaux, V. (2008). The horizontal and vertical relations in upright faces are transmitted by different spatial frequency ranges. *Acta Psychologica, 128*, 119–126.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial" – Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1023–1039.

Goffaux, V., & Rossion, B. (2007). Face Inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 995–1002.

Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 986–1004.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201–222.

Hole, G. J. (1994). Configural factors in the perception of unfamiliar faces. *Perception, 23*(1), 65–74.

Hole, G. J., George, P. A., & Dunsmore, V. (1999). Evidence for holistic processing of faces viewed as photographic negatives. *Perception, 28*, 341–359.

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31*, 1221–1240.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science, 15*(11), 762–768.

Leder, H., & Carbon, C.-C. (2005). When context hinders! Learn-test compatibility in face recognition. *Quarterly Journal of Experimental Psychology, 58A*, 235–250.

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. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 310–327.

McKone, E., Kanwisher, N., & Duchaine, B. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11*, 8–15.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Psychology, 9*(5), 555–604.

Rhodes, G., Peters, M., Lee, K., Morrone, C. M., & Burr, D. (2005). Higher-level mechanisms detect facial symmetry. *Proceedings of the Royal Society B, 272*, 1379–1384.

Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition, 88*, 79–107.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103*, 34–79.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT.

Rossion, R., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Vision*, *8*(4): 1–13, http://journalofvision.org/8/4/3/.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioural and Cognitive Neuroscience Reviews, 1*, 63–75.

Scognamillo, R., Rhodes, G., Morrone, C., & Burr, D. (2003). A feature based model of symmetry detection. *Proceedings of the Royal Society B, 270*, 1727–1733.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology A, 46*, 225–245.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience, 31*, 411–437.

Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences, 1*(9), 346–352.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141–145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747–759.