Vision Research 127 (2016) 1–10

Contents lists available at ScienceDirect

Vision Research

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

Contributions of feature shapes and surface cues to the recognition of facial expressions

Mladen Sormaz, Andrew W. Young, Timothy J. Andrews *

Department of Psychology, University of York, York YO10 5DD, UK

ARTICLE INFO

Article history:
Received 11 December 2015
Received in revised form 7 June 2016
Accepted 7 July 2016

Keywords:
Face
Expression
Shape
Surface
Texture
Image

ABSTRACT

Theoretical accounts of face processing often emphasise feature shapes as the primary visual cue to the recognition of facial expressions. However, changes in facial expression also affect the surface properties of the face. In this study, we investigated whether this surface information can also be used in the recognition of facial expression. First, participants identified facial expressions (fear, anger, disgust, sadness, happiness) from images that were manipulated such that they varied mainly in shape or mainly in surface properties. We found that the categorization of facial expression is possible in either type of image, but that different expressions are relatively dependent on surface or shape properties. Next, we investigated the relative contributions of shape and surface information to the categorization of facial expressions. This employed a complementary method that involved combining the surface properties of one expression with the shape properties from a different expression. Our results showed that the categorization of facial expressions in these hybrid images was equally dependent on the surface and shape properties of the image. Together, these findings provide a direct demonstration that both feature shape and surface information make significant contributions to the recognition of facial expressions.

1. Introduction

The human face has a complex musculature that allows it to create a remarkable variety of facial expressions (Du, Tao, & Martinez, 2014). Although there are individual differences between people in the precise anatomical arrangement of the facial muscles, those muscles involved in producing facial expressions of what are considered to be basic emotions (which include happiness, sadness fear, anger, and disgust) are highly consistent across individuals (Waller, Cray, & Burrows, 2008). These muscles allow a person to move critical expressive features such as the eyebrows, eyes, nose and mouth in ways that can change their shapes (e.g. raising or lowering the corners of the lips, widening or narrowing the eyes), their positions (raising or lowering the eyebrows), or often both (wrinkling the nose, or lowering the jaw to open the mouth).

Despite this well-known anatomical background, the nature of the visual information underlying recognition of facial expressions is poorly understood. While an obvious place to begin looking for critical visual cues might seem to be in the patterns of movement themselves, these are difficult to define and the good recognition of photographs of normal intensity basic emotions shows that the apex of a set of muscle contractions often creates an easily recognisable expressive configuration of the facial features. Moreover, notational systems such as the Facial Action Coding System (FACS: Ekman & Friesen, 1978) depend on the fact that the underlying pattern of muscle contractions that create an expression is evident even in a static image. Many studies therefore begin by exploiting the recognisability of well-validated photographs of facial expressions such as the Ekman and Friesen (1976) series, as we do here.

There are many ways of thinking about the visual information conveyed by a photograph of a face, but one that has proved very useful is in terms of its shape and surface properties. Any facial image consists of a set of edges created by abrupt changes in reflectance due to the shapes and positions of facial features and a broader pattern of reflectance based on the surface properties of the face – also known as texture or albedo (Bruce & Young, 1998, 2012). Shape properties can be operationally defined by the spatial locations of fiducial points that correspond to facial features; note that in this sense 'shape' properties will include both the feature shapes and their positions. In contrast, surface properties result from the pattern of reflectance of light due to the combination of ambient illumination, the face's pigmentation, and shape from shading cues.

* Corresponding author.
E-mail address: timothy.andrews@york.ac.uk (T.J. Andrews).

http://dx.doi.org/10.1016/j.visres.2016.07.002
0042-6989/

The distinction of shape from surface properties is widely used in face perception research (Bruce & Young, 1998, 2012) and is implicit in standard approaches to computer image manipulation (Tiddeman, Burt, & Perrett, 2001). These image manipulation techniques allow quasi-independent changes to a face's shape or surface properties. Such changes cannot be fully independent, of course, because many of the shape and surface properties of images will necessarily covary. For example, the surface property of shading is clearly affected in part by the face's shape. None the less, such methods allow us to hold face shape fixed as closely as possible (by using the same fiducial positions for a set of images) or to hold the surface properties fixed as closely as possible (by using the same surface brightness patterns in a set of images). This then allows a direct test of the relative contributions of shape and surface information. Studies based on this approach have demonstrated independent contributions of shape and surface properties to the perception of a range of facial characteristics including gender, age, attractiveness and dominance (Burt & Perrett, 1995; Russell, 2003; Torrance, Wincenciak, Hahn, DeBruine, & Jones, 2014).

Thinking of facial images as broadly consisting of shape (feature positions) and surface (pigmentation, shading patterns) properties has also helped our understanding of facial identity recognition, where it is clear that both shape and surface cues can contribute (Russell, Sinha, Biederman, & Nederhouser, 2006; Troje & Bülthoff, 1996), but that the role of surface cues becomes more salient for familiar faces (Burton, Jenkins, Hancock, & White, 2005; Russell & Sinha, 2007).

In contrast to the established role of surface cues in the perception of facial identity, judgements of expression are often thought to be based primarily on the shapes and positions of critical expressive features such as the eyebrows, eyes, nose and mouth. This makes sense because these shape changes are a direct consequence of facial muscle movements. Evidence for the primary importance of shape cues in facial expression recognition comes from contrast reversal (as in a photo negative). In a contrast-reversed image the edges that define feature shape properties remain in the same positions, despite the huge change in overall surface properties. Although contrast negation is well-known to be very disruptive of facial identity recognition (Bruce & Young, 1998, 2012), it turns out that judgements based on facial expression are still possible in contrast-reversed images (Bruce & Young, 1998; Harris, Young, & Andrews, 2014a; Magnussen, Sunde, & Dyrnes, 1994; Pallett & Meng, 2013; White, 2001). Similarly image manipulations that completely remove surface information, such as line drawings of faces, also show relatively preserved expression perception (Etcoff & Magee, 1992; McKelvie, 1973). Using such evidence, most current accounts posit shape information to be the most important cue in the perception and recognition of expression (Bruce & Young, 2012; Calder, Young, Perrett, Etcoff, & Rowland, 1996).

Although previous studies have suggested that feature shape is the dominant cue for the perception and recognition of facial expressions, there are grounds for thinking that surface information might also play a role (Benton, 2009; Calder, Burton, Miller, Young, & Akamatsu, 2001). For example, Benton (2009) found a decrease in the emotional expression aftereffect to facial expressions when images were negated, suggesting that the perception of facial expression can be affected by changes in surface information. Using Principal Component Analysis (PCA), Calder et al. (2001) found that principal components (PCs) that convey variation in surface information could be used to categorize different facial expressions, albeit to a lesser extent than PCs that convey variation in shape. However, while these findings show a potential role for surface cues, they do not provide a direct test of whether surface properties are actually used for the recognition of facial expression. None the less, there are obvious ways in which surface properties might be useful to facial expression recognition. For example the feature shape change of opening the mouth will be accompanied by a bright region if the teeth are bared or a relatively dark region if the teeth are retracted; these different surface brightnesses are a direct reflection of muscle movements that clearly convey different expressions. Moreover, there are also indirect effects of underlying muscle movements such as the skin folding around the mouth and eyes resulting from smiling. These changes do not correspond to specific facial features, and are largely evident from their impact on surface shading patterns.

The aim of the current study was therefore to investigate the contribution of changes in the shapes of key expressive features (such as the eyebrows, eyes, nose and mouth) and changes in surface brightness patterns (such as those resulting from showing the teeth, or furrowing the brow) to the categorization of facial expression. In Experiment 1, we manipulated images to create facial expressions that varied primarily in shape or primarily in surface cues. This was achieved by reshaping images of different expressions to standardise the locations of the fiducial positions across the images, or by standardising the surface properties as far as possible by overlaying the same averaged surface onto the fiducials that characterise each expression. Because many of the shape and surface properties of images will necessarily covary, this method does not orthogonally manipulate shape and surface information, but it does allow us to hold the shape fixed as closely as possible (by using the same fiducial positions for all images) or to hold the surface properties fixed as closely as possible (by using the same surface brightness patterns in all images). This then allows a direct test of whether the information that remains free to vary across images can actually be used for the categorization of facial expression. In Experiment 2, we used contrast-reversed versions of the images used in Experiment 1 to further probe the role of shape and surface properties in the recognition of facial expressions. In Experiment 3, we then created hybrid images that combined the surface properties from one expression with the shape of a different expression. This approach offers a complementary method to that used in Experiment 1 and 2 for determining the relative contribution of surface and shape cues to the categorization of facial expressions.

2. Experiment 1

2.1. Method

2.1.1. Participants

Participants (n = 20, female = 10, mean age = 24.8 years, SD = 3.8) were drawn from an opportunity sample of students and staff at the University of York. Participants gave informed consent and were paid or given course credit for their participation. All data were collected in accordance with the ethical guidelines determined by the Psychology Department of the University of York and were in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.1.2. Stimuli

Fig. 1 and Supplementary Fig. 1 shows the stimuli for the three conditions used in Experiment 1 (original, shape varying, and surface varying; these are the 5×5 image matrices that form the leftmost columns in Fig. 1). Static images of expressions were presented as these are well-recognised as long as they represent the apex of the pattern of muscle movements involved in producing the expression (see Bruce & Young, 2012). Five models (females F5, F6, F8, males M1, M6) were selected from the FEEST set (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002) of Ekman and

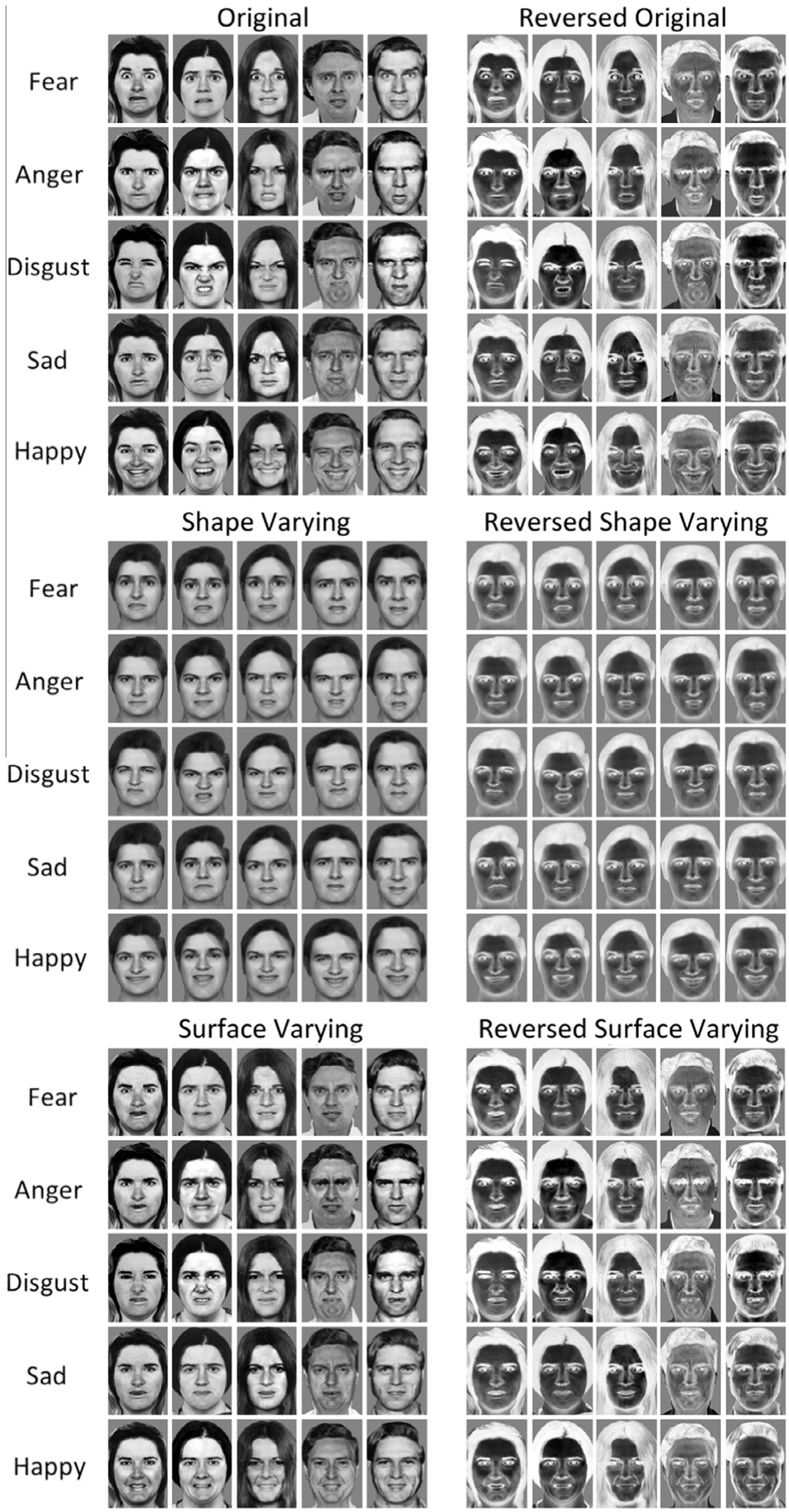

Fig. 1. Images used in Experiment 1 and Experiment 2. Original images are taken from the FEEST image set (Young et al., 2002). These show 5 models posing 5 expressions. The shape-varying images were created by superimposing the average surface of all images onto the original shapes. The surface-varying images were created by superimposing the original surfaces onto the average shape of all images. Experiment 1 used the normal contrast versions of each of the images shown on the left of the display. Experiment 2 used the normal contrast versions of the original images and the contrast-reversed versions of all images shown on the right of the display.

Friesen (1976) photographs on the basis of high recognisability of their facial expressions and the similarity of the action units (muscle groups) used to pose each of the expressions from Ekman's coding of the action units given in the FEEST test manual (Young et al., 2002). For each model, images of expressions of fear, anger, disgust, sadness and happiness were used. These unmodified images from the Ekman and Friesen (1976) series formed the 'original' image condition (see Fig. 1). Although also present in FEEST, expressions of surprise were omitted because the status of surprise as a basic emotion has been questioned (Oatley & Johnson-Laird, 1987); one can be pleasantly or unpleasantly surprised (Du et al., 2014).

The aim of Experiment 1 was to measure the perception of images that vary primarily in shape or primarily in surface properties. To create the 'shape-varying' images, Psychomorph software was used to manually delineate 179 facial fiducial points on each of the 25 original images (Tiddeman et al., 2001). All 25 original images were then reshaped to the average shape (as defined by the fiducial positions) of all 25 images and averaged together to arrive at the average surface brightness across all 25 images. This averaged surface brightness was then reshaped back to the original shape (as defined by the original fiducial positions) of each of the 25 images. In this way, 25 new images were created, each of which had lost any expression-specific surface brightness information and only contained the unique shape cues associated with each individual expression.

A 'surface-varying' set of images was then created, in which primarily the surface information rather than shape information provided the cue to the posed expression. To do this the surface of each of the 25 original images was reshaped into the average shape across all 25 images. This removes most of the underlying shape cues to expression (as all images now shared exactly the same set of fiducial points), but leaves the surface information relatively unchanged (each image retains its surface brightness pattern).

2.1.3. Procedure

Participants performed a 5-AFC (alternative forced choice) expression categorization task in which they indicated the perceived expression (Fear, Anger, Disgust, Sadness or Happiness) by a button press. Images were presented using PsychoPy (Peirce, 2008) at a viewing distance of 40 cm. Each trial began with a 500 ms fixation cross followed by a 1000 ms central presentation of one image ($16° \times 11°$). Faces from the normal contrast original, shape-varying and surface-varying sets were presented in a randomised order and participants viewed each of the faces twice during the experiment, so each participant completed 150 trials. We recorded both accuracy and reaction time data for all participants. The experiment began with 15 practice trials using one variant of each expression from each experimental condition in a randomised order. The actors used in the practice trials were different from the ones in the main experimental trials.

2.2. Results

Our principal analyses used overall accuracies and reaction times to determine whether it was possible to identify the facial expression in images that primarily varied in surface or shape. Fig. 2 shows that accuracy in the categorization task was above chance (20%) for the original (88.1 ± 2.2%), surface-varying (68.7 ± 3.5%) and shape-varying (70.2 ± 4.1%) images. An ANOVA showed a significant effect of condition ($F(1.42, 27) = 86.41$, $p < 0.001$, partial $n^2 = 0.82$). The main effect of condition was due to higher accuracy for the original images compared to both the surface-varying ($t(19) = 14.06$, $p < 0.001$) and shape-varying ($t(19) = 13.1$, $p < 0.001$) images. There was no significant difference between surface-varying and shape-varying conditions ($t(19) = 1.77$, $p = 0.1$).

We also found a significant effect of condition on reaction time ($F(2, 38) = 40.93$, $p < 0.001$, partial $n^2 = 0.68$). This main effect was due to faster reaction times to original images compared to either surface-varying ($t(19) = 7.3$, $p < 0.001$) or shape-varying ($t(19) = 7.23$, $p < 0.001$) images. There was no difference ($t(19) = 0.18$, $p = 0.86$) in reaction time between surface-varying (1557 ± 73 ms) or shape-varying (1553 ± 75 ms) images.

A subsidiary analysis was used to do determine whether different image cues are differentially important for different expressions, by including the five expressions as a factor in the accuracy analysis (Fig. 3). A 2 way ANOVA showed a significant interaction between image type and expression ($F(8,152) = 6.12$, $p < 0.001$, partial $n^2 = 0.45$). All post hoc paired t tests carried out to compare different conditions and all stated p values were Bonferroni-Holm adjusted and tested at a critical alpha level of 0.05. Accuracy for anger ($t(19) = 3.38$, $p = 0.014$ was higher for surface-varying images compared to shape-varying images. In contrast, accuracy for shape-varying images compared to surface-varying images was higher for sad ($t(19) = 5.01$, $p < 0.001$) and happy ($t(19) = 10.22$, $p < 0.001$) expressions. Accuracy for the original images was higher compared to surface-varying images for

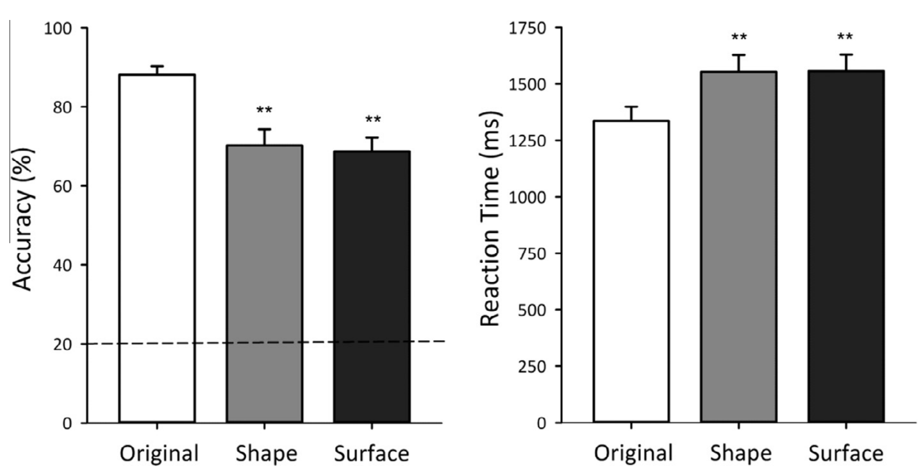

Fig. 2. Experiment 1: Accuracy and reaction times for the categorization task averaged across all expressions. Accuracy was above chance levels (horizontal line) but lower than for the original images for both the shape-varying (shape) and surface-varying (surface) conditions. There was no significant difference in the accuracy or reaction time between the shape-varying and surface-varying conditions. Error bars represent standard error of the mean. Significant differences compared to original images: $^{**}p < 0.001$.

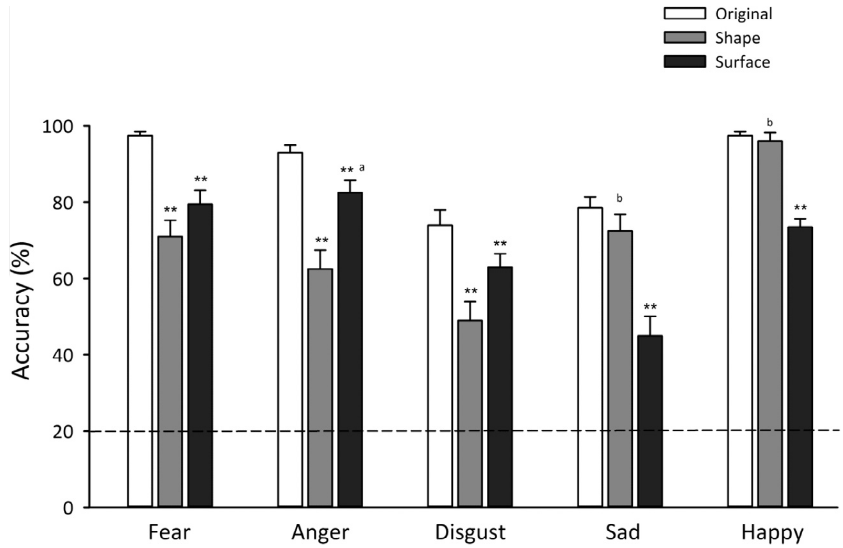

Fig. 3. Experiment 1: Accuracy for the categorization task for each expression. Error bars represent standard error of the mean. Accuracy for shape-varying (shape) and surface-varying (surface) conditions was above chance level (horizontal line) for each expression. However, the relative importance of shape and surface properties differed across expressions, with shape cues being relatively important for happy and sad expressions, and surface cues for anger and disgust. ** indicates significant differences compared to original: $p < 0.001$, [a] indicates surface significantly greater than shape, [b] indicates shape significantly greater than surface.

fear ($t(19) = 7.21$, $p < 0.001$), anger ($t(19) = 3.5$, $p = 0.014$), sadness ($t(19) = 6.98$, $p < 0.001$) and happiness ($t(19) = 12.42$, $p < 0.001$). Accuracy for the original images was higher compared to shape-varying images for fear ($t(19) = 6.58$, $p < 0.001$), anger ($t(19) = 6.97$, $p < 0.001$) and disgust ($t(19) = 5.08$, $p = 0.001$). Accuracy for original images was not significantly greater to shape-varying images for happy ($t(19) = 1$, $p = 0.79$) and sad ($t(19) = 1.16$, $p = 0.79$) expressions.

2.3. Discussion

Facial expression images that were manipulated to have fixed surface properties (the shape-varying images) or to have fixed shapes (the surface-varying images) were less well-recognised than original images from the Ekman and Friesen (1976) series, with longer reaction times and higher error rates. This shows that both shape and surface properties contribute to facial expression recognition.

Although there was no difference in the overall impact of holding fixed the shape or surface properties of the expressions, there were clear differences in how these contributed to recognising different emotions. For happiness and sadness, images with averaged surfaces (the shape-varying images) were recognised better than images with averaged shapes (the surface-varying images), showing the importance of shape cues for recognising these expressions. The opposite pattern held for anger and disgust, for which surface cues made a stronger contribution.

A potential limitation of our method for controlling shape is that only 179 fiducial positions were used to define the shape of each expressive feature, and human perceivers are known to be very sensitive to small differences in curvature that underlie the perception of feature shapes (Kosslyn, Hamilton, & Bernstein, 1995). We therefore ran an experiment using contrast-reversed images to confirm that the surface-varying images (i.e. those with the fixed fiducial positions) did not contain significant residual shape information. Contrast reversal has no effect on the positions of edges in the images, as the abrupt discontinuities in brightness values that create the perception of edges are still present, so it does not markedly affect what we here call shape information. Instead, contrast reversal has a substantial effect on surface properties because it inverts the relationships between all the relatively light and dark areas in the image. We therefore predicted that contrast reversal should result in a decrease in the recognition of expressions from the surface-varying images (as their fixed fiducial positions should largely have eliminated differences in the shape information that can survive contrast reversal), but should have no effect on shape-varying images (as these images do not contain any useful surface information that could be disrupted by contrast reversal).

3. Experiment 2

3.1. Method

3.1.1. Participants

Participants ($n = 20$, female = 10, mean age = 20.3 years, $SD = 1.8$) were recruited in the same way as for Experiment 1. Participants gave informed consent and data were again collected in accordance with the ethical guidelines determined by the Psychology Department of the University of York.

3.1.2. Stimuli

Stimuli are shown in Fig. 1. They comprised contrast-reversed versions of the images used in Experiment 1 (the 5×5 matrices of images shown in the rightmost columns of Fig. 1). Normal contrast versions of the original set of 25 images from the FEEST set (forming the 5×5 image matrix positioned in the upper left part of Fig. 1) were also included as a point of comparison.

3.1.3. Procedure

The procedure followed that established for Experiment 1, with the exception that four sets of stimuli were use in Experiment 2 (normal contrast original images, contrast-reversed original images, contrast-reversed shape varying images, and contrast-reversed surface varying images).

3.2. Results

Mean accuracies and correct reaction times for recognition of emotion in each condition are shown in Fig. 4.

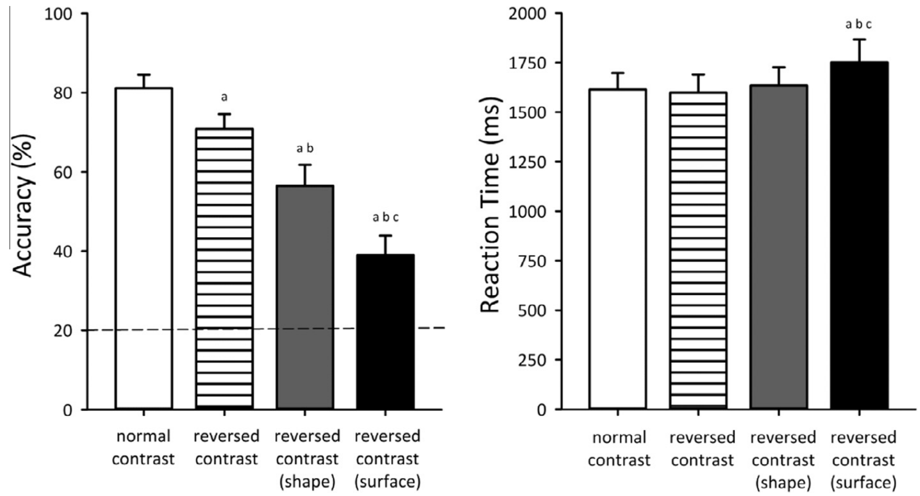

Fig. 4. Experiment 2: Accuracy and reaction times for the categorization task averaged across all expressions. Accuracy was above chance levels (horizontal line) for all conditions, but the biggest drop in performance was evident for the contrast-reversed surface-varying images. The right panel shows reaction times in each of the experimental conditions for correct trials. Error bars represent standard error of the mean. [a] indicates significantly different to normal contrast, [b] indicates significantly different to reversed contrast, [c] indicates significantly different to reversed contrast (shape).

Our principal analyses again compared overall accuracies and reaction times across conditions. Fig. 4 shows that accuracy in the categorization task was above chance (20%) for original images (81.1 ± 3.4%), reversed original (70.9 + 3.7%), reversed surface-varying (39 ± 4.9%) and reversed shape-varying (56.5 ± 5.3%) images. An ANOVA of the accuracy data with Greenhouse-Geisser corrected degrees of freedom (due to violation of sphericity) showed a significant effect of condition ($F (2.17, 41.3) = 174.95$, $p < 0.001$, partial $n^2 = 0.90$). Post hoc paired t tests p values were again Bonferroni-Holm adjusted. The main effect of condition was due to higher accuracy for the normal contrast images compared to the reversed contrast ($t (19) = 8.2$, $p < 0.001$), reversed contrast, surface-varying ($t (19) = 22.42$, $p < 0.001$), and reversed contrast, shape-varying ($t (19) = 10.07$, $p < 0.001$) conditions. There was also a significantly higher recognition accuracy for the reversed contrast images than for contrast-reversed, surface-varying ($t (19) = 19.96$, $p < 0.001$) and contrast-reversed, shape-varying conditions ($t (19) = 7.85$, $p < 0.001$). Recognition accuracy was significantly higher in the reversed shape-varying condition than in the reversed surface-varying condition ($t (19) = 7.38$, $p < 0.001$).

Fig. 4 shows that the categorization task reaction times were similar in the original image (1614 ± 83 ms), contrast-reversed original image (1599 ± 91 ms), and contrast-reversed shape-varying (1635 ± 92 ms) conditions, and slowest in the contrast-reversed surface-varying (1751 ± 116 ms) condition. This pattern was confirmed with a second ANOVA (also with Greenhouse-Geisser corrected degrees of freedom) which demonstrated a significant effect of condition on reaction time ($F (1.93, 36.6) = 6.03$, $p = 0.006$, partial $n^2 = 0.24$). Post hoc paired t tests p values were again Bonferroni-Holm adjusted. This main effect reflected significantly faster reaction times to original images compared to contrast-reversed surface-varying ($t (19) = 2.88$, $p = 0.04$) images. There was no significant difference in reaction time between original images and contrast-reversed original images ($t (19) = 0.22$, $p = 1$) or contrast-reversed shape-varying ($t (19) = 0.23$, $p = 1$) conditions. There were significantly faster reaction times in the contrast-reversed original condition compared to contrast-reversed surface-varying ($t (19) = 2.76$, $p = 0.04$) but not contrast-reversed shape-varying conditions ($t (19) = 0.38$, $p = 1$). The reaction times in the contrast-reversed shape-varying condition were significantly faster than in the contrast-reversed surface-varying condition ($t (19) = 3.54$, $p = 0.01$).

In Experiment 1, we found no difference in accuracy for shape-varying and surface-varying original-contrast images, whereas in Experiment 2 there was greater accuracy for shape-varying compared to surface-varying contrast-reversed images. To confirm that this difference in the pattern of results was statistically reliable, we took the accuracy data from the three conditions of Experiment 1 (original, shape-varying, and surface-varying images) and the three corresponding contrast-reversed conditions in Experiment 2 (contrast-reversed original, contrast-reversed shape-varying, and contrast-reversed surface-varying images) and submitted these to a 2×3 ANOVA with image format (normal or contrast-reversed) as a between-groups factor and experimental condition (normal, shape-varying, surface-varying) as a within-group factor. This ANOVA revealed a significant main effect of image format ($F (1, 38) = 75.73$, $p < 0.001$, partial $n^2 = 0.67$), a significant main effect of condition ($F (1.48, 56.06) = 202.54$, $p < 0.001$, partial $n^2 = 0.84$), and a significant interaction between image format and condition ($F (1.48, 56.06) = 10.12$, $p = 0.001$, partial $n^2 = 0.21$). The image format $\times$ condition interaction confirms that the pattern of effects differed between the normal-contrast image format used in Experiment 1 and the reversed-contrast images used in Experiment 2.

3.3. Discussion

Experiment 2 found that contrast reversal had a greater impact on recognising the surface-varying than the shape-varying images created for Experiment 1. Contrast reversal has little effect on the edge cues that define features shapes but completely changes the brightness values that define surface patterns. On this basis, if the images created for Experiment 1 did indeed minimise the roles of surface or shape cues as intended, we would expect to find a substantial decrement of contrast reversal on the recognition of surface-varying images and less impact for shape-varying images. That this was precisely the pattern observed provides support to the view that the image manipulation techniques used to create the stimuli for Experiment 1 had achieved the intended effects.

Further support for the importance of surface properties in the recognition of facial expression is shown by the effect of contrast reversal on the original and shape-varying images. We found that contrast reversal of the original images significantly lowered the recognition of facial expression. The only difference between the normal contrast compared and reversed contrast image sets is the surface properties. This implies that surface properties are important for the recognition of facial expression. We also found that contrast reversed original images were recognised significantly better than contrast reversed shape-varying images. Again, the only difference between these image sets is the surface properties (reversed contrast of original images compared to reversed contrast of averaged images). This implies that the surface properties continue to contribute to the recognition of facial expression even in the contrast reversed original images. Thus, contrast reversal disrupts rather than eliminates surface cues.

To further explore the contributions of shape and surface properties we introduced a complementary method for Experiment 3, in which we tested the categorization of all possible combinations of the average shape of one expression with the averaged surface of another expression. This method pits shape against surface properties, allowing a test of which dominates the expression seen in the hybrid image. For example, we can ask whether a hybrid of fear shape and happy surface will be seen as fear (shape dominance), as happiness (surface dominance) or as some other expression (because of the inconsistent cues).

4. Experiment 3

4.1. Method

4.1.1. Participants

The participants used in Experiment 3 were those used in Experiment 1. They gave informed consent and were tested within the ethical guidelines of the University of York Psychology Department.

4.1.2. Stimuli

The aim of Experiment 3 was to compare the relative contribution of shape and surface properties to the perception of facial expression with a technique that would be complementary to that used in Experiment 1. Hybrid images were created that combined the average surface properties from one expression with the average shape from another expression. To generate the average surface properties from each expression, all 25 original images were reshaped to the average across all 25 images. The five images of each expression were then averaged to create an average surface for each of the five expressions. To generate the shape properties for each expression, we averaged the position of the fiducial points across the five images from each expression. This gave rise to one shape image for each expression. Finally, we combined the averaged surface for each expression with the averaged shape for each expression to create a matrix of 25 images (Fig. 5) in which images on the top left to bottom right diagonal have the average surface and shape properties from the same expression. All other images in Fig. 5 have the average surface properties from one expression and the average shape properties from a different expression, allowing us to estimate whether these hybrid expressions are recognised primarily from of their shape or their surface properties.

As noted already, covariation between shape and surface properties means that completely independent change in one property without also changing the other property cannot be achieved. None the less, our intention was to achieve relatively greater changes in image shape or surface. To verify that image shapes were manipulated as intended, we applied an edge detector (Sobel filter) to the images in Fig. 5. The output of the edge detector is shown in Supplementary Fig. 2. This shows that images with the same shape (columns) appear to be similar, whereas images with different shapes (rows) appear more different. To quantify this effect, we measured the difference in gray value of these filtered images across columns or across rows. Consistent with the perceptual impression given by the images, Supplementary Fig. 3 shows that the absolute difference in gray value was greater per pixel ($t(49) = 8.81$, $p < 0.001$) for images that vary in shape (rows: mean = 5.17, SEM = ± 0.03) compared to images that have the same shape (columns: mean = 5.57, SEM = ± 0.04).

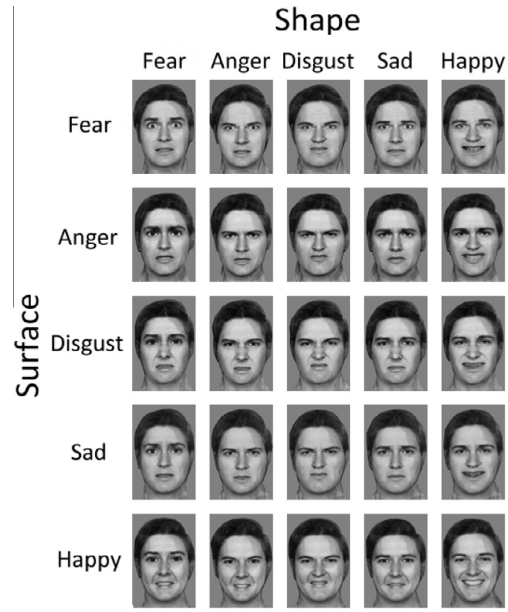

Fig. 5. Hybrid images used in Experiment 3. Images were created by combining shape and surface properties from averaged facial expressions. Images on the top left to bottom right diagonal have the averaged shape and surface properties of the same facial expression. All other images have average shape and surface properties from different expressions. For example, the bottom left image combines an averaged fear shape with an averaged happy surface.

4.1.3. Procedure

The same 5AFC categorization task and presentation parameters were used as for Experiment 1, except that in Experiment 3 each of the 25 faces was presented five times, leading to 125 trials.

4.2. Results

In Experiment 3, we directly probed the relative contributions of shape and surface properties to the categorization of facial expression using hybrid images with the average surface properties from one expression and the average shape properties from another expression. Because these images were created from the averaged shapes and surfaces of each expression across the 5 models from the FEEST set, we began by checking that they were none the less seen as the intended emotion when the shape and surface of the same expression were combined (the images falling along the diagonal from top left to bottom right in Fig. 5). Participants categorized these images in which the surface and shape cues conveyed the same expression with an accuracy of 92%, showing at least as good recognition as the original Ekman and Friesen (1976) images from Experiment 1.

Next, we determined the response for hybrid images in which the surface and shape cues conveyed different expressions. Fig. 6 shows that participants reported the expression that corresponded to the surface properties of the image on 41.1% of trials and the

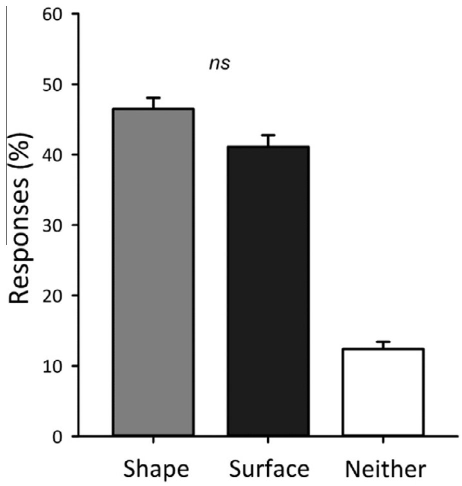

Fig. 6. Experiment 3: Responses indicating whether the categorized expression corresponded to the Shape or Surface properties of the image, or when the response did not correspond to the shape or surface information in the image (Neither). Responses based on Shape and Surface were significantly higher than Neither shape nor surface, but there was no significant difference (*ns*) between the responses based on Shape and Surface themselves.

expression that corresponded to the shape of the images on 46.5% of trials. Only a small proportion of responses were based on neither shape nor surface. An ANOVA showed a significant effect of condition ($F(1.53, 29.08) = 95.52$, $p < 0.001$, partial $n^2 = 0.83$). Correct responses based on shape and surface properties were significantly higher compared to responses that did not correspond to the expression from either the surface or shape properties (both $p < 0.001$). There was no overall difference in the proportion of trials in which participants chose surface compared to shape ($t(19) = 1.73$, $p = 0.1$).

To investigate the subsidiary issue of whether different image cues might be differentially important for different expressions, the data were separated by the expression identified in participants' responses, as shown in Fig. 7. An ANOVA revealed a significant interaction between condition and expression ($F(4.41, 83.86) = 63.09$, $p < 0.001$, partial $n^2 = 0.77$). Post hoc paired t tests p values were again Bonferroni-Holm adjusted. In this analysis, the surface properties were more dominant than shape for disgust ($t(19) = 3.11$, $p = 0.006$) and sadness ($t(19) = 4.95$, $p < 0.001$). In contrast, shape cues were more dominant for happiness ($t(19) = 16.28$, $p < 0.001$).

4.3. Discussion

Like Experiment 1, the overall pattern of findings from Experiment 3 showed that both shape and surface properties are important to the recognition of facial expressions. As for Experiment 1, the analysis by expressions also showed that the relative importance of shape and surface cues varied across expressions. However, these detailed patterns differed somewhat across experiments, with the consistent findings being that perception of happiness is largely determined by feature shapes and disgust by surface properties. We can speculate that this reflects the salience of the distinctive mouth shape in happy expressions and the way that shading patterns enhance the nose wrinkling that characterises many expressions of disgust (Rozin, Lowery, & Ebert, 1994). The most important point, though, is again that both shape and surface properties contribute.

5. General discussion

The aim of the present study was to investigate the roles of feature shapes and surface properties of the face in the recognition of facial expressions. To achieve this, we used complementary converging methods. In Experiment 1, we created images that held either shape or surface properties as constant as possible, to investigate the usefulness of each source of information in relative isolation. In Experiment 2, we validated the general method for varying shape and surface properties by testing recognition of contrast-reversed versions of the images used in Experiment 1. In Experiment 3, we created hybrid images that combined the averaged shapes and surfaces of different expressions, to investigate which type of information would dominate the perceived expression.

Our results show clearly that both shape and surface properties are useful for the recognition of facial expression. Indeed, despite the widely-shared opinion that the shapes of expressive features created by muscle movements are particularly important to

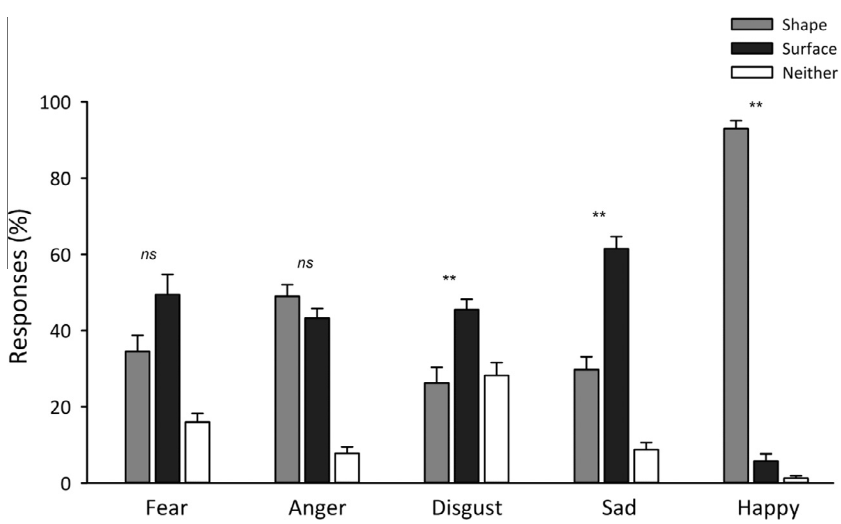

Fig. 7. Experiment 3: Responses for each expression indicating whether the expression was categorized based on the Shape or Surface properties of the image, or when the response did not correspond to the shape or surface information in the image (Neither). Surface responses were significantly higher than Shape responses for disgust and sadness. Shape responses were significantly higher than Surface responses for happiness. ** indicates significant differences between Shape and Surface conditions, [ns] indicates no significant differences between Shape and Surface conditions.

expression, we instead found that both shape and surface information contributed more or less equally overall, albeit with some differences between different expressions. In Experiment 1, we found that identification accuracy for images that varied primarily in surface properties was well above chance (around 70% correct in 5AFC) and not significantly different from images that varied primarily in shape. In Experiment 2 we validated the properties of the images created for Experiment 1 by demonstrating a substantial decrement of contrast reversal on the recognition of expression for the surface-varying images and less impact for shape-varying images. This pattern is as expected because contrast reversal has little effect on the edge cues that define features shapes but completely changes the brightness values that define surface patterns. In Experiment 3, we directly compared the relative contributions of shape and surface cues to the categorization of facial expression. We found that participants were equally likely to use the surface or shape information to categorize facial expressions when viewing hybrid images that contain the surface properties from one facial expression and the shape properties from another facial expression.

So, while the present findings provide further support for the long held assertion that feature shape cues are important for the perception and recognition of expression, they also show that the importance of surface information in the representation of facial expressions has been underestimated. The novel finding from this study is that images that mainly contain variation in their surface properties can convey facial expression. Indeed, removing either shape or surface information impairs perception of expressions approximately equally. In both Experiments 1 and 3 categorization of surface-only and shape-only images was significantly lower than for images containing appropriate shape and surface information. Taken together these findings show that the perceptual mechanisms that underpin the recognition of facial expression are tightly linked to both shape and surface information.

Impairment of shape or surface cues did not have an equivalent effect for each expression. This suggests that the informative cues in the face may vary for each expression, for example a smile signifying happiness produces a consistent shape change of an upturned mouth across all actors, making shape a salient and prominent cue. Conversely, a facial expression signifying disgust although also highly identifiable may produce less consistent or more subtle shape cues, therefore leaving the viewer to rely also on surface cues not present for other expressions. This reliance on both types of cue may reflect the natural covariance between shape and surface cues within many expressions. For example, fear expressions involve opening the mouth and widening the eyes (shape cues) and this creates salient contrast changes in the eye and mouth regions (surface cues). Moreover, some of the critical surface cues involve skin folding and other perturbations that are actually the result of movements of the facial features, allowing them to act as proxies for these even when the correlated feature shape changes themselves have been eliminated.

At first sight, these findings contrast with previous studies showing that the disruption of surface cues does not substantially impair perception of expressions (Harris et al., 2014a; Pallett & Meng, 2013; White, 2001). However, more careful, examination of such studies shows clear pointers to the conclusion we have reached. For example, White (2001) showed in 2 of his 4 experiments a small but non-significant increase in error rate for contrast negated expressions when compared to normal images of facial expression. Similarly, although Pallett and Meng (2013) found preserved recognition accuracy for contrast-negated expressions they noted a reduced adaptation aftereffect for contrast-negated expressions. Indeed, as already noted Benton (2009) also found a reduced expression adaptation aftereffect to negated expressions, suggesting a role for surface information. However, our data go beyond these previous studies by showing that when the only differences between images are in the surface information, it can still be used to correctly identify some facial expressions. In other words, surface information alone can be sufficient.

Of course the photographs and computer-manipulated stimuli we used were two-dimensional static images, whereas faces we see in everyday life are three-dimensional and nearly always moving. However, the use of 2D static images brings advantages in terms of experimental control, and we do not think it has serious limitations. The role of movement in assisting facial expression recognition is most clearly evident with much more subtle and variable expressions than the type we have used here (e.g. Ambadar, Schooler, & Conn, 2005; Krumhuber, Kappas, & Manstead, 2013). The good recognition of photographs of normal intensity basic emotions shows, for these emotions at least, the apex of the set of muscle contractions forms a recognisable configuration and studies have not found much in the way of differences between neural responses to moving and static expressions of basic emotions (Harris, Young, & Andrews, 2014b; Johnston, Mayes, Hughes, & Young, 2013). So despite its importance, we need to be careful not to overstate the role of movement in facial expression recognition.

Much the same point applies to three-dimensionality. Although the face is a complex 3D structure, decades of research has established that humans are good at recognising age, sex, familiar identity and expression from 2D photographs (Bruce & Young, 2012). What may be important here, though, is that some of this excellent performance with photographs may reflect the presence of 3D cues in the form of shape from shading, but the standard image-manipulation methods we use here will assign shading information to surface properties and not to the feature shapes per se. This does not seem to us a serious limitation as there are no grounds at present for thinking that (say) a smile will be less obvious because a person has a protruding nose. Where it does have an influence, though, is that we have been careful to point out that the skin folds that result from raising the corners of the mouth, wrinkling the nose, or screwing up the eyes will also be treated as strongly covarying surface properties and not as changes in the feature shapes themselves.

In conclusion, we show that both shape and surface information can be used to identify facial expressions. We also show that the relative importance of shape and surface varies across different expressions, presumably reflecting the extent to which either cue can be diagnostic of a particular facial expression. In some ways, finding that both shape and surface properties play important roles fits what we noted in the Introduction concerning the perception and recognition of age, sex and identity. But because prominent theories (Haxby, Hoffman, & Gobbini, 2000; see also Calder & Young, 2005) draw a strong distinction between changeable (expression) and relatively invariant facial characteristics (age, sex, identity), it was important to investigate what happens in the case of facial expression. The key new finding is thus that both shape and surface information play important roles in the recognition of facial expressions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2016.07.002.

References

Ambadar, Z., Schooler, J. W., & Conn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science, 16*(5), 403–410.

Benton, C. P. (2009). Effect of photographic negation on face expression aftereffects. *Perception, 38*(9), 1267–1274.
Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception* : . Oxford University Press.
Bruce, V., & Young, A. (2012). *Face perception* : . Psychology Press.
Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings Biological Sciences/The Royal Society, 259*(1355), 137–143.
Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51*, 256–284.
Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research, 41*(9), 1179–1208.
Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641–651.
Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Catgorical perceptions of morphed facial expressions. *Visual Cognition, 3*(2), 81–117.
Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National academy of Sciences of the United States of America, 111*(15), E1454–E1462.
Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.
Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement* : . Consulting Psychologists Press.
Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*(3), 227–240.
Harris, R. J., Young, A. W., & Andrews, T. J. (2014a). Brain regions involved in processing facial identity and expression are differentially selective for surface and edge information. *NeuroImage, 97*, 217–223.
Harris, R. J., Young, A. W., & Andrews, T. J. (2014b). Dynamic stimuli demonstrate a categorical representation of facial expression in the amygdala. *Neuropsychologia, 56*(1), 47–52.
Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233.
Johnston, P., Mayes, A., Hughes, M., & Young, A. W. (2013). Brain networks subserving the evaluation of static and dynamic facial expressions. *Cortex, 49*(9), 2462–2472.
Kosslyn, S., Hamilton, S., & Bernstein, J. (1995). The perception of curvature can be selectively disrupted in prosopagnosia. *Brain and Cognition, 27*(1), 36–58.
Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review, 5*(1), 41–46.
Magnussen, S., Sunde, B., & Dyrnes, S. (1994). Patterns of perceptual asymmetry in processing facial expression. *Cortex, 30*(2), 215–229.
McKelvie, S. J. (1973). The meaningfulness and meaning of schematic faces. *Perception & Psychophysics, 14*(2), 343–348.
Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion, 1*(1), 29–50.
Pallett, P. M., & Meng, M. (2013). Contrast negation differentiates visual pathways underlying dynamic and invariant facial processing. *Journal of Vision, 13*(14), 1–18.
Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10.
Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology, 66*(5), 870–881.
Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception, 32*(9), 1093–1107.
Russell, R., & Sinha, P. (2007). Real-world face recognition: The importance of surface reflectance properties. *Perception, 36*(9), 1368–1374.
Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception, 35*(6), 749–759.
Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications, 21*(5), 42–50.
Torrance, J. S., Wincenciak, J., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2014). The relative contributions of facial shape and surface information to perceptions of attractiveness and dominance. *PLoS ONE, 9*(10), e104415.
Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*(12), 1761–1771.
Waller, B. M., Cray, J. J., & Burrows, A. M. (2008). Selection for universal facial emotion. *Emotion, 8*(3), 435–439.
White, M. (2001). Effect of photographic negation on matching the expressions and identities of faces. *Perception, 30*(8), 969–981.
Young, A. W., Perrett, D. I., Calder, A. J., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: Stimuli and tests (FEEST)*. Bury St Edmunds, England: Thames Valley Test Company (January).