The expressions of strangers: Our identity-independent representation of facial expression

Andrew L. Skinner

School of Experimental Psychology, University of Bristol, Bristol, UK

Christopher P. Benton

School of Experimental Psychology, University of Bristol, Bristol, UK

Evidence suggests that underlying the human system processing facial expressions are two types of representation of expression: one dependent on identity and the other independent of identity. We recently presented findings indicating that identity-dependent representations are encoded using a prototype-referenced scheme, in a manner notably similar to that proposed for facial identity. Could it be that identity-independent representations are encoded this way too? We investigated this by adapting participant to anti-expressions and asking them to categorize the expression aftereffect in a prototype probe that was either the same (congruent) or different (incongruent) identity to that of the adapter. To distinguish between encoding schemes, we measured how aftereffect magnitude changed in response to variations in the strength of adapters. The increase in aftereffect magnitude with adapter strength characteristic of prototype-referenced encoding was observed in both congruent and, crucially, incongruent conditions. We conclude that identity-independent representations of expression are indeed encoded using a prototype-referenced scheme. The striking similarity between the encoding of facial identity and both representations of expression raises the possibility that prototype-referenced encoding might be a common scheme for encoding the many types of information in faces needed to enable our complex social interactions.

Keywords: expression, identity-independent representation, adaptation

Introduction

Humans are remarkably proficient at recognizing facial expressions, not only those of individuals familiar to us but also those of people we have never met before. This capacity extends to recognizing the expressions of unfamiliar individuals of other cultures and is so striking that it is acknowledged as a fundamental cross-cultural human trait (Ekman, 1972). The existence of this trait implies an ability to encode facial expressions in an identity-independent manner; the current study examines the nature of this representation.

We recently investigated the representation of expressions using a new technique in which we adapted participants to anti-expressions (Skinner & Benton, 2010). Adapting to an anti-expression (e.g., anti-fear) produced an aftereffect in which perception was biased toward the matching expression (e.g., fear). We interpreted this as evidence that expressions are represented not as discrete entities but within a multidimensional framework, such has been proposed for representations of facial identity (Leopold, O'Toole, Vetter, & Blanz, 2001). We then investigated the type of encoding used within this framework by following an approach recently used to explore our encoding of gaze (Calder, Jenkins, Cassel, & Clifford, 2008) and measured

the effect of varying the strength of the adapter (where strength was the distance of the adapter from the probe) on the magnitude of the anti-expression aftereffect.

In a prototype-referenced scheme, pairs of opposing populations of neurons code the deviation of a face from a prototype face along a number of dimensions (Rhodes et al., 2005). Aftereffects in this kind of scheme are the result of differences in the level of activation within these pairs of neuron populations during adaptation: the bigger the difference in activation, the greater the aftereffect. Increasing the strength of an adapter by moving it further from a prototype probe will result in greater differences in activation within the pairs of populations during adaptation and, consequently, an increasing magnitude aftereffect (Robbins, McKone, & Edwards, 2007). In an exemplar encoding scheme, individual faces are coded by discrete, overlapping populations of neurons. Adaptation will only affect a probe if the populations coding the adapting face and the probe face are close enough together to overlap. Increasing the strength of the adapter by moving it further from the prototype probe increases the separation between the populations coding the adapter and probe faces, which clearly reduces the extent to which these populations will overlap, therefore leading to a decrease in the magnitude of the aftereffect (Robbins et al., 2007). Following this logic, the increase in anti-expression aftereffect magnitude

with increased adapter strength we observed was consistent with a prototype-referenced encoding scheme for expressions (Skinner & Benton, 2010).

One of our aims in this initial study of the encoding of expressions (Skinner & Benton, 2010) was to ensure that patterns of aftereffect magnitudes were not muddied by idiosyncratic variations in expressions across individuals. To help reduce this possibility, the adapter and probe stimuli we used were produced by averaging expressions from multiple identities. The averaging process is highly effective at reducing idiosyncrasies in faces and produces a set of prototypical expressions that are uniform in terms of their identity. As a result, it is possible that our previous study was tapping into a representation of expression that was identity specific and not a representation that would support the recognition of expressions in unfamiliar identities. Our ability to recognize the identity of individuals just from the idiosyncratic biological motion within facial expressions certainly seems to suggest that this kind of identity-dependent representation exists (Hill & Johnston, 2001; Knappmeyer, Thornton, & Bulthoff, 2001).

Indeed, evidence of the existence of representations of expression that are both dependent and independent of identity also comes from the adaptation literature. Fox and Barton (2007) measured expression aftereffects using adapters and probes that were both the same and different identities. Aftereffects were strongest when the adapting and probe expressions were the same identity; however, they were still present, albeit at an attenuated level, when the adapter and probe were different identities. Fox and Barton interpreted the difference in aftereffect magnitude as evidence that our representation of expression is not unitary but comprises two types of representational components—one that is identity independent, contributing to aftereffects whether the adapter and probe identities are either congruent or incongruent, and one that is identity dependent, providing an additional contribution to aftereffect magnitude only when adapting and probe identities are congruent. The same variation of expression aftereffect magnitude with adapter–probe identity congruency has been reported in a number of subsequent studies (Benton, 2009; Campbell & Burke, 2009).

There is, however, an alternative explanation for the reduction in aftereffect strength observed in these studies that warrants consideration. Traditional models of face perception (Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000) detail a component, prior to the point at which identity and expression bifurcate, that analyzes and codes the physical structure of faces. Bruce and Young's (1986) model describes a structural encoding function, and Haxby et al. (2000) suggest a neural subsystem for early perception of facial features, located in the inferior occipital gyri. Representations of faces at this structural level have been shown to be dissociable from representations of higher level constructs such as identity (Rotshtein, Henson, Treves, Driver, & Dolan, 2005). This means that it is possible that the patterns of aftereffects observed by

Benton (2009), Campbell and Burke (2009), and Fox and Barton (2007) could have been the result of adaptation in a number of stages in the face processing hierarchy, with some adaptation happening at the structural level, and feeding through to structurally independent stages elsewhere within the face perception system. In this case, the reduction in aftereffects observed with incongruent adapting and probe identities would not be attributable solely to differences in the high-level construct of identity but also to differences between the structural information in the adapting and probe faces. Just measuring the overall expression aftereffect gives a view of the combined effect of adaptation at these different stages and may not necessarily enable us to distinguish between them.

Fox and Barton (2007) explored the possibility that the reduction in aftereffect magnitude they observed in incongruent conditions was attributable not to differences in identity but to differences in lower level factors by including a "same person, different image" adaptation condition. They observed no difference in aftereffect magnitude between this and a "same person, same image" condition and concluded that the reduction in aftereffect magnitude seen across different identities was not the result of low-level image differences. However, as the authors acknowledged, this manipulation explored the contribution of adaptation to low-level image properties and did not address the possibility that adaptation at a cortical level capable of representing the shape and structure of faces may have been contributing to the magnitude of aftereffects.

Butler, Oruc, Fox, and Barton (2008) further explored which factors contribute to expression aftereffects by comparing aftereffects of intact faces to those produced by faces in which the individual features had been arranged in a random configuration. Aftereffects produced by faces with scrambled features were smaller than those produced by intact faces. The authors interpreted this as evidence that expression aftereffects were not the result of local adaptation to image elements such as those describing individual features. While this does indicate that adaptation to local image properties and image elements such as tilt and curvature are not major contributors to expression aftereffects, this still leaves the possibility that adaptation at a somewhat higher level, at which this local information is aggregated into a structural representation of a complete face, may play a significant role in any expression aftereffects observed.

We reasoned there are two likely ways that having different adapting and test identities might affect structural representations and reduce expression aftereffects. The first is simply that the relationships between facial features, such as the distance between the eyes and the distance between the nose and mouth, all clearly vary from person to person. These variations in the arrangement of features mean that the overall structure of each individual's face will be different. When adaptation to one identity biases the perception of a probe with a different identity, the differences in structure between the two faces could distort the

expression aftereffect in the probe, reducing the percept of a particular expression in the aftereffect. The second relates to the manner in which individuals animate facial features to produce expressions. Because the facial musculature controlling expressions depends on the underlying skull morphology, and this varies from individual to individual, there may be idiosyncratic variations in expression production linked to the underlying morphology of faces (Cohn, Schmidt, Gross, & Ekman, 2002; Pessa et al., 1998; Schmidt & Cohn, 2001). If the adapter and probe identities differ, the particular feature displacements that produced a perfectly natural looking expression in the adapter identity might be a poor fit to the probe face, distorting the aftereffect. This might well reduce the percept of a particular expression in the aftereffect.

In the present study, we aimed to address these two possibilities in ways that required the minimum of changes to the adapter and probe faces, preserving the naturalistic quality of our stimuli. First, we reduced the idiosyncratic variations in expression between identities. For this, we began by averaging all the exemplars for each expression to make prototypical expressions. We then averaged all the exemplars from all expressions to make an overall prototype face. The displacement of facial features between the overall prototype and a prototypical expression gave a standard, prototypical feature transformation for that expression that could be applied to any identity. This meant that the idiosyncratic, veridical expressions of an individual used here could be replaced by versions displaying standardized, prototypical expressions. If, after processing our adapter and probes in this manner, a reduction in aftereffect strength was still observed switching from congruent to incongruent identities, it would seem unlikely that idiosyncratic variations in the production of expression acting on a structural representation were the cause of the reduction.

To explore the possibility that the reduction in aftereffect was the result of differences in invariant aspects of facial features, we measured the differences in the shape and location of the key facial features between pairs of adapting and probe identities. If no correlation was observed between these feature difference scores and the magnitude of aftereffects, it would seem unlikely that a reduction in aftereffect magnitude in incongruent conditions could be attributable primarily to differences in invariant aspects of facial features acting on a structural level representational of faces.

The present study broadly followed the approach used in our previous investigation of the representation of expression (Skinner & Benton, 2010) but with two key differences. First, we replaced the prototypical expression stimuli with those from real identities. We use these to test for evidence of identity-dependent and identity-independent representations of expression by measuring the aftereffects produced by combinations of adapters and probes from either the same (congruent) identity or different (incongruent) identities. Second, we introduced the two

new procedures described previously that addressed the possibility that adaptation was happening at a structural level. If we are able to rule out an explanation for changes in aftereffect magnitude based on adaptation at a structural level, a reduction in aftereffect magnitude from congruent to incongruent conditions would support the existence of identity-dependent and identity-independent representations of expression. We could then investigate the encoding of putative identity-dependent and identity-independent representations of facial expression, by observing how the magnitude of identity-congruent and identity-incongruent expression aftereffects varies as the strength of adapter varies.

While the nature of identity-dependent representations has been the subject of previous investigation, that of identity-independent representations remains largely unexplored. Yet, it is these identity-independent representations that likely underlie our capacity to recognize the facial expressions of strangers with such astonishing proficiency. Understanding this ability, which is so key to successfully navigating our complex social world, rests on first understanding how we encode these representations. It is the nature of these representations that forms the focus of the current study.

Methods

Participants

Our initial plan was to run the experiment with two adapter strength conditions with 48 participants in each condition, and 48 participants (5 males, average age = 19.6 years) completed the experiment with adapters at 100% strength. However, on the basis of the results for the 100% strength condition (which we ran first), it was apparent that 24 participants per strength condition would be adequate, enabling us to test a further two strength conditions. Twenty-four participants (all females, average = 20.1 years) then completed the experiment with adapters at 50% strength and 24 (2 males, average = 19.4 years) with adapters at 25% strength. All participants were undergraduates at the University of Bristol.

Stimuli

All images were drawn from the Karolinska Directed Emotional Faces (KDEF) database (Lundqvist, Flykt, & Öhman, 1998). Two male identities (labeled M1 and M2 here) and 2 female identities (labeled F1 and F2) were selected for use as the test (adapter and probe) identities, on the basis that the expressions they displayed were free from acute idiosyncrasies. To produce the prototypical expression transformations that were used to reduce
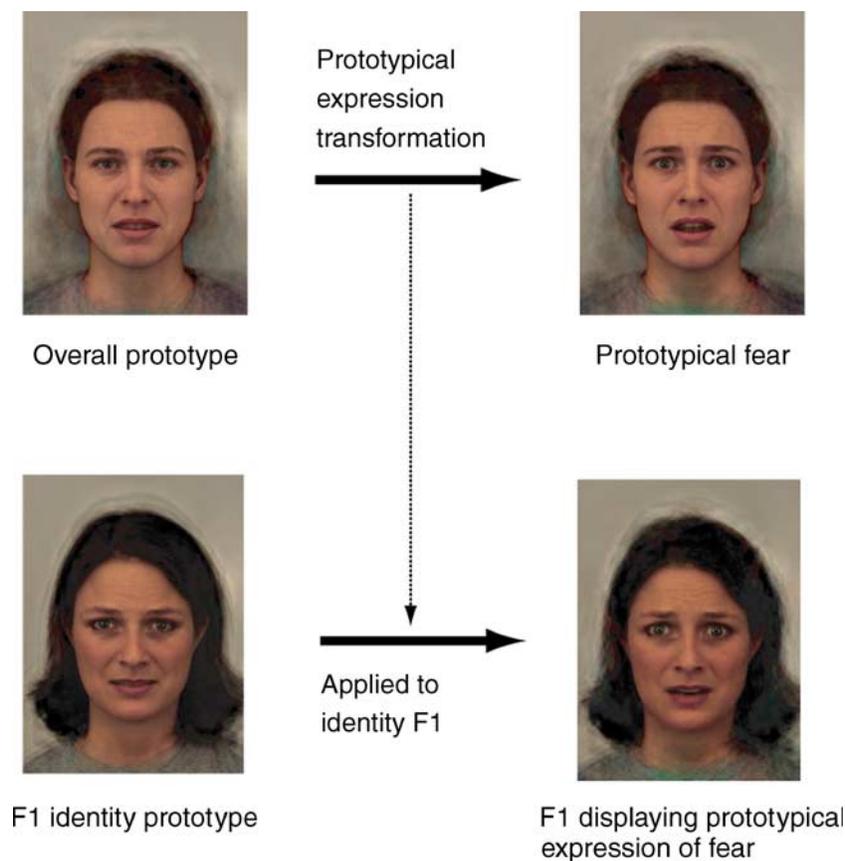
Figure 1. Construction of prototypical expressions.

idiosyncratic variations in expression production, 25 male and 25 female identities were selected from the KDEF database, each displaying expressions of the six basic emotions (happy, sad, fear, anger, surprise, and disgust) together with a neutral expression. For each expression, the 50 exemplars were averaged using the PsychoMorph application (Tiddeman, Burt, & Perrett, 2001) to produce a prototypical version of that expression.

To make an overall prototype, one approach would have been to average only neutral expressions. However, the use of neutral expressions in this context could reasonably be questioned, given evidence that neutral expressions contain affective content (Kesler-West et al., 2001; Lee, Kang, Park, Kim, & An, 2008) and can form expressions in their own right (Shah & Lewis, 2003). More importantly, at least for present purposes, the overall prototype produced from neutral expressions has certain feature characteristics (e.g., a closed mouth) that introduce distortions during the morphing procedure used to construct anti-expressions. These distortions are visible in the anti-expressions at strengths far lower than those required here. We used an alternative approach to make the overall prototype, in which we averaged across all expressions. This produced an overall prototype without a strong readily identifiable affective content (see Figure S1 in the Supplementary materials), with the benefit that it enabled

us to construct anti-expression stimuli at the higher strengths required for the current study. Note that we have used this approach previously (Skinner & Benton, 2010) and a similar approach has also been used by Cook, Matei, and Johnston (2011).

The difference between the overall prototype and a prototypical expression was essentially a transformation that described how the facial features were deformed to produce that expression. Using PsychoMorph, these transformations could be applied to any identity to produce an image of that identity displaying a prototypical version of that expression (see Figure 1). To make images of the four test identities display prototypical expressions, we first averaged all of the expressions for each identity to make four identity prototypes. We then applied each of the six prototypical expression transformations to each identity prototype using PsychoMorph. The veridical and prototypical expressions are shown together in Figure 2.

To produce anti-expressions, the shape and texture from a prototypical expression for an identity were morphed (Tiddeman et al., 2001) along a trajectory connecting that face with the relevant identity prototype, through to the other side and away from the identity prototype. The strength of the anti-expressions was modified by changing how far the morph continued along the trajectory away from the identity prototype. Morphing the face away from

Figure 2. Veridical and prototypical expressions for the four test identities. Left block shows veridical expressions, with columns for, from left, actors F1, F2, M1, and M2. In each column, expressions are, from top, happy, sad, fear, anger, surprise, and disgust. Right block shows the same but with prototypical expressions.

the identity prototype a distance equivalent to that between the prototypical expression and the identity prototype produced an anti-expression with a designated strength of 100%. Morphing the face away from the identity prototype to only half the distance between the prototypical expression and the identity prototype produced an anti-expression at 50% strength (see Figure 3).

The identity prototype, prototypical expressions, and anti-expressions for each test identity are shown in Figure 4. All stimuli were converted to grayscale, and the edges of each face were blurred to display mean luminance.

To ensure the prototypical expressions produced in this manner effectively conveyed the expected signals of emotion, we used a categorization task to measure the affect of each of the prototypical expressions for each identity.

The results of this task (included in the Supplementary materials) show that participants categorized the prototypical expressions with the expected emotions at levels consistent with natural expressions (Calvo & Lundqvist, 2008). In the same task, participants also categorized the prototypical anti-expressions. The results for these are also included in the Supplementary materials but are less indicative of the naturalistic quality of these faces, as there is no expectation that natural looking anti-expressions will convey predictable, clear patterns of affect. Inspection of the prototypical anti-expressions in Figure 4 reveals that some have a rather unnatural appearance—take the anti-expressions of happy, for example, with their narrow, down-turned mouth shapes. The less than natural appearance of certain prototypical anti-expressions does not, however,
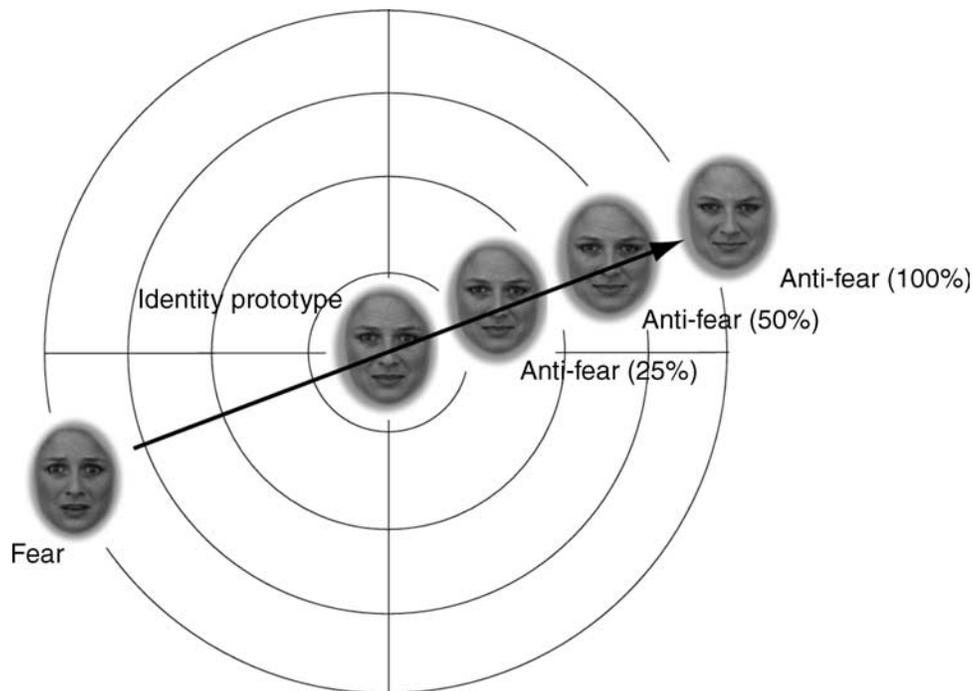
Figure 3. Construction of anti-expressions at different strengths.

necessarily mean that using these stimuli in adaptation tasks limits the applicability of results to actual faces. Indeed, many of the adapting faces used in classic face adaptation studies have been far from natural in appearance; for example, consider those used in studies of the face distortion aftereffect (e.g., Webster & MacLin, 1999). Furthermore, the prototypical anti-expressions shown in Figure 4 are the greatest strength anti-expressions used in this study (100%). As the strength of the anti-expressions is reduced to 50% and then 25%, their appearance becomes increasingly more natural (see Figure 5 for examples), yet they are still effective at producing robust aftereffects (see Results section below).

Design

Each of the four test identities was used as an adapter—with itself and each of the other three other identities as probes. The adapting images were the 6 anti-expressions for each identity, and the probe was always the identity prototype for the relevant identity. To keep testing sessions to within an hour, each experimental session comprised testing one identity adapter (with each of its six anti-expressions), with an identity-congruent probe (i.e., the same identity as the adapter) and one identity-incongruent probe. In half of the sessions, the incongruent probe was the same sex as the adapter; in the other half, it was the opposite sex. To test all combinations of adapters and probes, 12 different session types were required. Each participant completed one session,

and each session type was completed by four participants with adapters at 100% strength and by two participants with adapters at 50% and 25% strength. The inclusion of the 25% adapter strength condition is an important improvement on our previous study (Skinner & Benton, 2010), which used only 50% and 100% conditions. While the reduction in aftereffect magnitude we observed when adapter strength decreased from 100% to 50% was indicative of prototype-referenced encoding, it was possible that the reduction at 50% was the result of a dip in aftereffect strength around that adapter strength and that aftereffect strength increased as adapter strength decreased further. This would have been suggestive of exemplar rather than prototype-referenced encoding. The 25% strength condition gives us increased insight into what happens to aftereffect strength as adapter strength is decreased toward zero.

Within each session, trials were grouped into blocks of 6, with each anti-expression adapter used once in each block. The order in which the anti-expressions were used within a block was randomized for each block. Participants completed 3 blocks with identity-congruent probes and 3 with identity-incongruent probes, making a total of 36 trials. Each session lasted approximately 1 h, with roughly 45 min of testing.

Procedure

The experiment took place in a darkened room where the monitor was the only strong source of illumination.
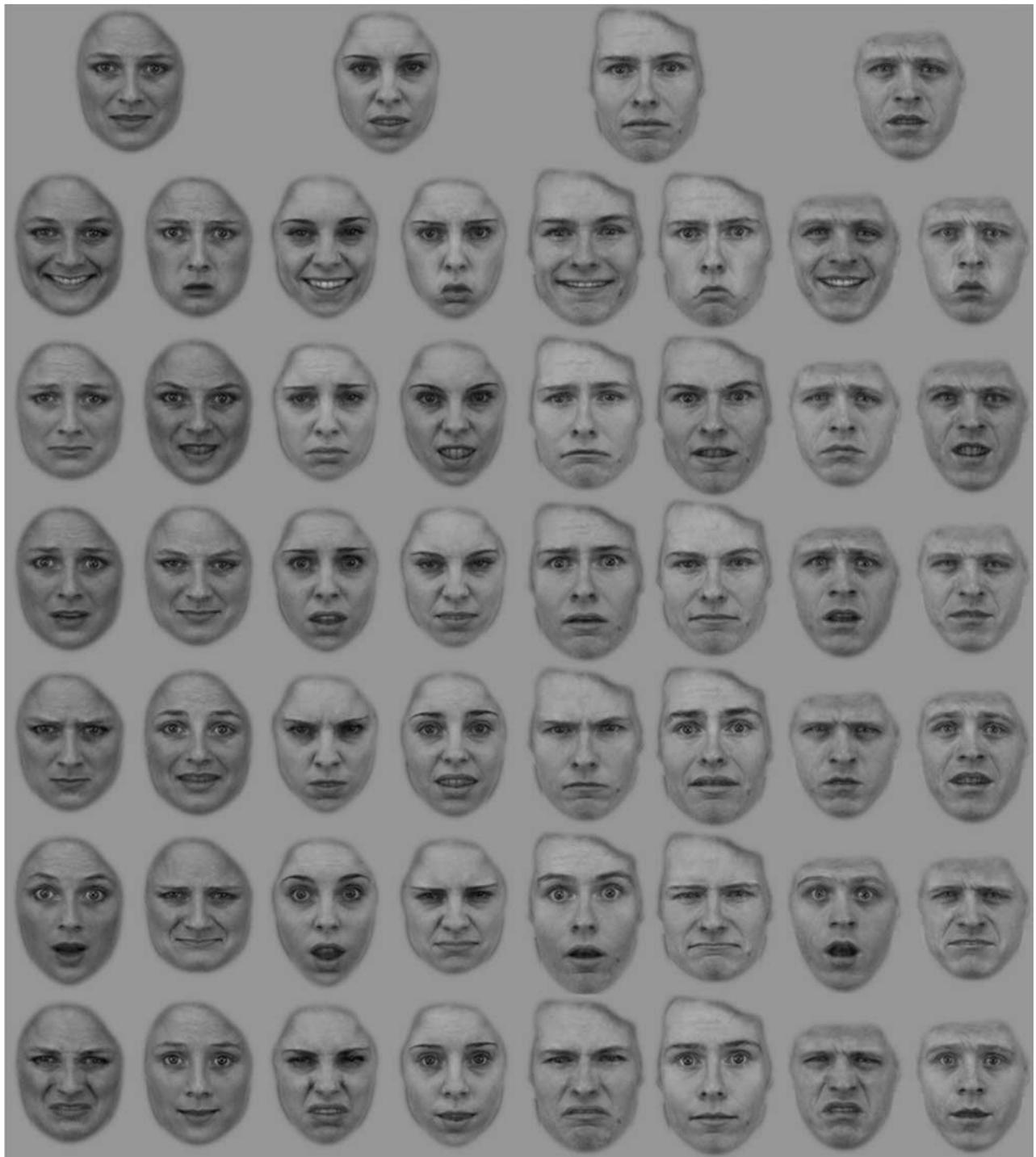
Figure 4. Prototypical identities, expressions, and anti-expressions for each test identity. Top row shows prototypical identities, with, from left, identities F1, F2, M1, and M2. Underneath each identity, the column to the left column shows their prototypical expressions, with, from the top, happy, sad, fear, anger, surprise, and disgust. The column to the right shows the corresponding prototypical anti-expressions.

Stimuli were presented on a Lacie Electron Blue IV 22″ monitor. The spatial resolution was set to 1024 × 768 pixels, and temporal resolution was set to 75 Hz. The monitor was linearized and had a mean luminance of 61.3 cd/m². Matlab was used to control the experiments, and the PsychToolBox extensions (Pelli, 1997) were used to display the stimuli. Participants viewed stimuli at a distance of 1 m. At this distance, face stimuli subtended visual angles of 6.8 degrees vertically and 5.1 degrees horizontally.

Figure 5. Prototypical anti-expressions at different strengths. Top row shows identity F1 with anti-expression of surprise, at strengths 25% (left), 50% (center), and 100% (right). Bottom row shows identity M1 with anti-expression of happy at same strengths.

In each trial, participants were presented with an adapting image of an anti-expression of a particular identity for 1 min. In order to prevent retinotopic adaptation, such as one finds in the tilt aftereffect (Dickinson, Almeida, Bell, & Badcock, 2010), participants were instructed to maintain fixation on a point in the center of the screen while the adapting image rotated around this point once every 5 s in a circular trajectory of 1-degree diameter (Benton, Jennings, & Chatting, 2006; Fang & He, 2005). Trajectory start position and direction were determined randomly for each stimulus. This was followed by a blank ISI of 500 ms, after which the identity prototype, either for the same identity as the adapter or one of the other 3 test identities, was shown for 500 ms. Participants were then presented with the names of the six emotions in text boxes along the bottom of the screen and used a handheld keypad to select the emotion that best described the expression of the average face after adaptation.

Feature difference scores

To generate a score reflecting the difference in structural information in the features of two faces, we utilized a particular aspect of the PsychoMorph application. Central to the use of this application is the association of a template with every face to be processed and every new face that is produced by the application. This template identifies the location of 219 points that delineate the key facial features and outline of the face (Brennan, 1985; Rowland & Perrett, 1995; Tiddeman et al., 2001). Each of the points in a template always marks the same part of a facial feature, so, for example, point 88 in every template always marks

the location of the left corner of the mouth. By taking the templates for two faces and scaling and aligning these using the centers of the pupils as registration points, it was possible, for each point in the template, to calculate the difference between the locations of that point in the two faces as a Euclidean distance. This distance was calculated for the 150 points in the templates marking the contours of the key facial features of identity prototypes (see Figure 6). The square root of the sum of squares of these distances gave a single score for the differences in shape and location of all facial features between pairs of adapter–probe faces, reflecting the differences in the morphological structure of these features.

Results

To compute aftereffect magnitudes, we calculated the percentage of responses, in identity-congruent and identity-incongruent conditions, in which each participant categorized the aftereffect with an emotion (e.g., fear) that matched the anti-expression adapter (i.e., anti-fear). We then calculated the mean of these scores across all participants for each adapter strength to give a score for aftereffect magnitude congruent and incongruent at each of the three adapter strengths (Skinner & Benton, 2010): These are shown in Figure 6. For completeness, the distributions of responses for individual anti-expressions are shown in Figures S3, S4, and S5 in the Supplementary
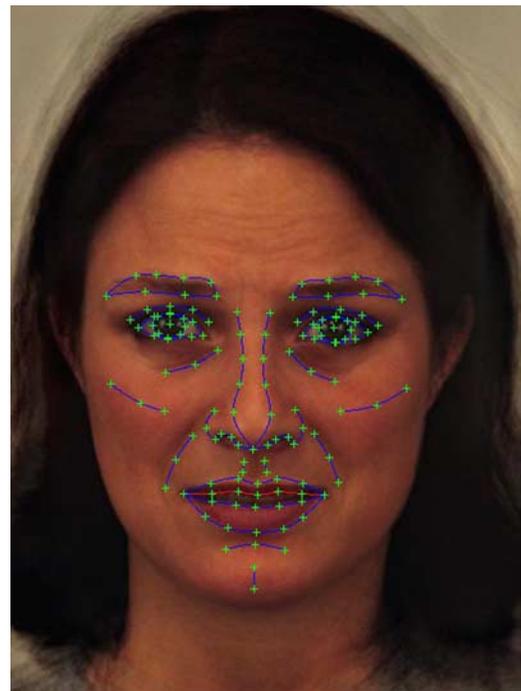

Figure 6. Location of feature delineation points in PsychoMorph templates used to compute the feature difference scores.

materials. We would counsel the reader to be careful in drawing conclusions from any patterns that they might perceive within these data. Without a clear idea of how expressions might differ from one another as they vary over adaptation strength and congruity, any comparisons are necessarily post hoc. The number of comparisons that one intrinsically makes while looking at such a large spread of data is necessarily large. To avoid apophenia, we have avoided making comparisons between expressions and have instead concentrated on looking across expressions with a clear eye to the predictions that are being made.

A mixed design ANOVA (with Greenhouse–Geisser correction) with congruency as within-subjects factor and adapter strength as between-subjects factor revealed a significant effect of congruency ($F(1,93) = 85.3$, $p < 0.001$) and adapter strength ($F(2,93) = 18.8$, $p < 0.001$), with no interaction between these ($F(2,93) = 2.43$, $p = 0.093$). Post-hoc analysis using Games–Howell (as adapter strength group sizes were different) indicated that in both congruent and incongruent conditions aftereffect magnitude was greater when adapter strength was 100% and 50% than

| Identity | Identity | | |
	M2	F1	F2
M1	10.7	12.4	11.6
M2	–	11.3	11.7
F1	–	–	9.85

Table 1. Feature difference scores between pairs of identities.

when it was 25% ($p < 0.001$) but that there was no difference in aftereffect strength between adapter strengths 100% and 50% ($p = 0.89$).

As adapter strength decreases below 25%, aftereffect magnitude will presumably decrease further until at some point it reaches the level of chance performance (16.67%). Measurements below this point will essentially be subject to a floor effect and will no longer yield meaningful information about the relationship between adapter strength and aftereffect magnitude. However, at the lowest adapter strength used here, 25%, the lowest magnitude aftereffect observed (in the incongruent condition) was still significantly above the level of chance, $t(23) = 3.10$, $p = 0.005$, effect size $d = 0.633$.

To explore whether or not the relationship between aftereffect magnitudes in congruent and incongruent conditions varied across the three adapter strength conditions, we calculated quotients at each of the adapter strengths. For each of the three groups, we divided each participant's mean congruent aftereffect magnitude by their mean incongruent magnitude. The overall means of the quotients for the three adapter strengths are shown in Figure 7. There was no significant difference in quotient values between adapter strengths ($F(2,95) = 0.148$, $p = 0.863$), indicating that the ratio of congruent to incongruent aftereffect magnitudes remained constant across adapter strengths.

To investigate the relationship between feature differences and anti-expression aftereffect magnitude, aftereffect magnitude was calculated (as detailed previously) for all identity-incongruent combinations of adapters and probes. Feature difference scores, between all combinations of the four identity prototypes, were also calculated (see Table 1).

The aftereffect magnitude and feature difference score for each identity-incongruent adapter–probe combination are shown in the scatter plot in Figure 8. A Spearman's rank test for correlation showed no relationship between structural difference scores and aftereffect magnitude for the 12 incongruent combinations ($r(10) = -0.1$, $p = 0.76$).
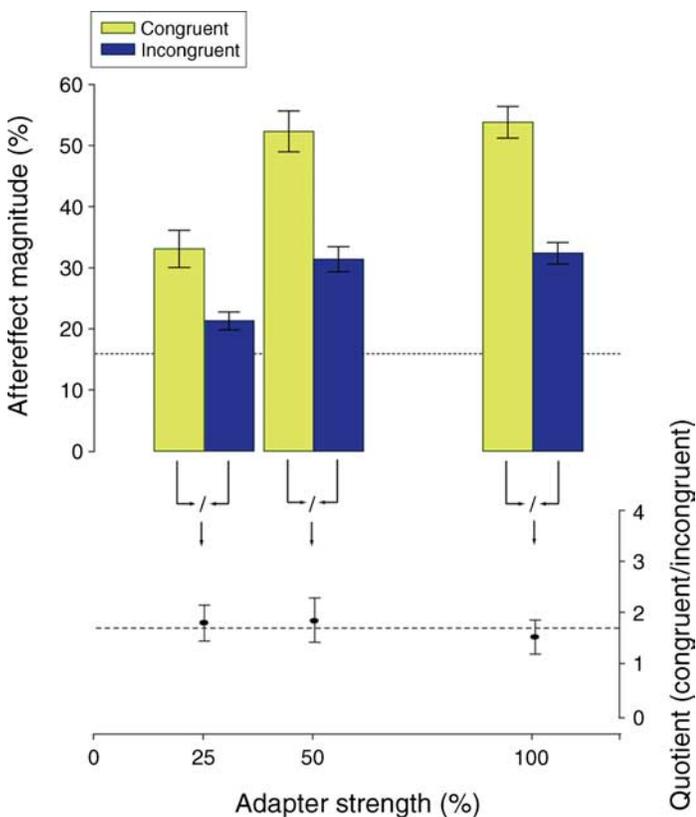

Figure 7. Anti-expression aftereffect magnitude in identity-congruent and identity-incongruent conditions as a function of anti-expression adapter strength. Dotted line shows level of chance performance. Below these are quotients resulting from division of congruent by incongruent aftereffect magnitudes, with dashed line showing the mean of quotients. All error bars show 95% confidence intervals.

Discussion

The primary purpose of the present study was to explore the encoding of identity-independent representations of expressions. We investigated this by measuring how varying the strength of anti-expression adapters affected
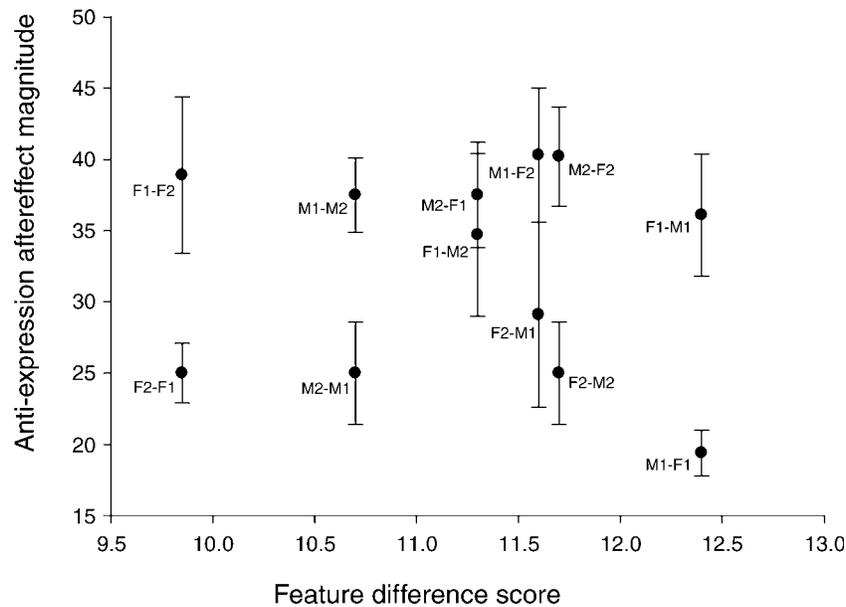
Figure 8. Variation of anti-expression aftereffect magnitude with feature difference score for all adapter–probe identity pairings. Points are labeled with adapter–probe identities. Error bars show 95% confidence intervals.

the magnitude of anti-expression aftereffects. As detailed previously, the two likely encoding schemes, prototype-referenced and exemplar, make different predictions about the effect of increasing adapter strength. Prototype-referenced encoding predicts that increasing adapter strength will increase aftereffect magnitude, while exemplar encoding predicts that increasing adapter strength will decrease aftereffect magnitude (Robbins et al., 2007). In the identity-incongruent conditions, aftereffect strength increased as adapter strength was increased from 25% to 50%, indicating the use of prototype-referenced encoding in the identity-independent representation of expressions.

In their study of neural models of the encoding of faces, Robbins et al. (2007) explored the encoding of a facial attribute that varied along a single dimension. Encoding facial expressions will, of course, require a representational framework with multiple dimensions, and we should consider if the logic from Robbins et al. still applies in this case. In an exemplar-based scheme, in which there is a direct, monotonic relationship between the similarity of faces and the proximity of the discrete neural populations encoding those faces, increasing the strength of the adapter and making it increasingly dissimilar to the probe face inevitably has the effect of increasing the separation of the populations of neurons coding the adapter and the prototype. This decreases the overlap between these populations and, consequently, results in a decreased magnitude aftereffect. No matter how many dimensions a representational framework has, moving the adapter further from the probe will always have the effect of increasing the separation of the neuron populations coding the adapter and probe, so it seems reasonable to assume that this principle remains true irrespective of the number of dimensions in a representation space.

In a prototype-referenced scheme, the adapter will be encoded as differences from the prototype along each of the dimensions in the space. The difference encoded along each dimensions may vary (the adapter face may have eyes that are open wider than those of the prototype but a mouth that is more closed), but however the adapter differs from the prototype in any dimension, this difference will increase in magnitude as the strength of the adapter increases and the adapter moves further from the prototype. This will result in increasing disparity in activation in the pools of neurons coding those dimensions and, assuming our system for detecting expressions aggregates information across these dimensions, increased magnitude aftereffects. Again, there seems no good reason to doubt that the predictions of Robbins et al. (2007) would apply in a multidimensional representational framework. The increase in aftereffect magnitude with adapter strength that we observed is most likely to indicate of the use prototype-referenced encoding in identity-independent representations of expressions.

As adapter strength was increased to 100%, there was no further increase in aftereffect magnitude. The reason aftereffect magnitude leveled off in this way is not entirely clear. However, given the finite range of activation of the populations of cells encoding deviations from a prototype in a prototype-referenced scheme, it does seem reasonable to assume that aftereffect magnitude could not increase indefinitely. Following this line of reasoning, Susilo, McKone, and Edwards (2010) recently described how the neural model of opponent encoding of faces proposed by Rhodes et al. (2005) predicts exactly this response. Despite this lack of a further increase in aftereffect magnitude as adapter strength increases from 50% to 100%, at no point was aftereffect magnitude observed to decrease with an

increase in adapter strength, in the way that would be expected in an exemplar encoding scheme. We would suggest the monotonic increase in aftereffect magnitude with adapter strength we observed sits best with the use of prototype-referenced encoding in the identity-independent representations of expression.

The response to adapter strength in the identity-congruent conditions followed almost exactly the same pattern as that in the identity-incongruent conditions. Indeed, the quotients we obtained by dividing the congruent by incongruent aftereffect magnitudes at the various levels of aftereffect strength showed that the congruent aftereffect magnitudes were approximately twice those of the incongruent aftereffect magnitudes at all levels of adapter strength. As the identity-congruent responses will have contributions from both identity-dependent and identity-independent representations of expression, the similarity in pattern suggests that the response of the identity-dependent representation is highly similar to that of the identity-independent representation, again displaying the monotonic increase in aftereffect magnitude with adapter strength characteristic of prototype-referenced encoding.

The reduction in aftereffect magnitude from identity-congruent to identity-incongruent conditions we observed in the current study followed the pattern of identity-contingent expression aftereffects observed in a number of previous studies (Benton, 2009; Campbell & Burke, 2009; Fox & Barton, 2007). Fox and Barton (2007) attributed this pattern of aftereffects to the existence of identity-dependent and identity-independent representations of expression. We addressed the possibility that the pattern of aftereffects might reflect not adaptation in high-level representations of expression and identity but adaptation in an earlier stage of processing, in the representation that codes the structure of faces. If adaptation were happening at this structural level, there appeared two likely ways that having different adapter and probe identities might result in a decrease in expression aftereffect.

The first relates to the invariant aspects of the facial features, such as the length of the nose and the distance between the eyes. The shapes and relationships between these structures vary from individual to individual, and these variations mean that it is possible that adapting to one identity biases a structural representation so that the shape and arrangement of the features in the aftereffect in the probe face of another identity look unnatural, reducing the strength of expression perceived in the aftereffect. To explore the possibility that structural differences between adapter and probe might be driving aftereffect magnitude, we checked for a correlation between aftereffect magnitudes and the scores describing structural differences in facial features between pairs of identities. No such correlation was observed. In addition, we observed a clear asymmetry in aftereffect magnitude in adapter–probe pairs (as shown by the lack of overlap in error bars; Cumming & Finch, 2005). The aftereffect magnitude with identity A as adapter and identity B as probe was generally different to the

magnitude with identity B as the adapter and A as the probe. If aftereffect magnitude were primarily driven by structural differences between adapter and probe, there seems no reason to expect this kind of asymmetry, as the structural differences between identities A and B are the same when A is the adapter and B is the probe as they are when B is the adapter and A is the probe.

The second possibility relates to idiosyncratic variations in the way the adapter and probe identities produce expressions, linked to individual differences in underlying face morphology and musculature. It is possible that these idiosyncratic differences in expressions might also reduce the strength of expression aftereffect. We used prototypical expression transformations to reduce the idiosyncratic variation in expressions in our test identities. Despite taking this step, the decrease in anti-expression aftereffect in incongruent conditions remained, so it seems unlikely that this decrease was caused by idiosyncratic variations of expressions acting on a structural representation. These findings speak against an explanation based on adaptation happening at the level of a structural representation. It seems reasonable to conclude, therefore, that the adaptation observed here reflects the operation of separate identity-dependent and identity-independent higher level facial representations.

Taken together, our results indicate that identity-dependent and, crucially here, identity-independent representations of expression are encoded using prototype-referenced schemes within multidimensional frameworks. These findings describe, for the first time, how we encode identity-independent representations of facial expression. One of the most striking aspects of these findings is the similarity between the way the results here suggest representations of expressions are encoded and the way we encode facial identity. Leopold et al. (2001) explored the representation of facial identity by comparing anti-identity aftereffects along various trajectories in identity space and, from the differences in aftereffects observed along different trajectories, concluded that facial identity is encoded using a prototype-referenced scheme. More recently, Loffler, Yourganov, Wilkison, and Wilson (2005) used an fMRI adaptation paradigm to show that a single neural population responded to changes in faces along an identity trajectory with its origin at the prototype face but not to equivalent distance changes in directions not corresponding to an identity trajectory originating at the prototype. This again supports the notion that identity is encoded as trajectories in a multidimensional space, with reference to a prototype at the origin of that space. The qualitative similarity between the findings of Leopold et al. and Loffler et al. and those from the current study, in particular those relating to the representation of expression that is independent of identity, raises a number of possibilities for the relationship between representations of identity and expression.

The similarity may reflect the use of a common approach to encoding across separate systems representing

different aspects of faces; this would likely offer some economy and might mitigate compatibility issues when integrating different types of information about faces in the numerous ways our complex social interactions demand. Equally, it may be a further indication that the representation of expression and identity are less separate than traditional models of face perception proposed (Bruce & Young, 1986; Haxby et al., 2000) and that more recent models suggesting much greater overlap of expression and identity in representational systems common to both (Calder, Burton, Miller, Young, & Akamatsu, 2001; Dailey, Cottrell, Padgett, & Adolphs, 2002) offer a better account of how we encode different types of information about faces.

Broadening our view from expression and identity to face processing in general, what are the wider implications of our findings for the encoding of other information in faces? Does the existence of an identity-dependent representation of expression speak to the existence of representations combining other social information from faces—age and sex for example? The results from recent contingent adaptation studies (Schweinberger et al., 2010) certainly appear to support the existence of some form of overlap between the representations of these kinds of information too. Whether this reflects the existence of multiple prototypes located as nodes throughout a single representational system (thus enabling the coding of combinations of different types of information) or the interactions between separate representation systems for those different types of information remains unclear. In both cases, however, the notion of a common encoding strategy would seem a beneficial, if not essential, requirement for the integration of different kinds of information.

In summary, we used anti-expression aftereffects to explore the way we encode representations of facial expression, in particular representations that are independent of identity. The observation of anti-expression aftereffects in both identity-congruent and identity-incongruent adaptation conditions suggests that identity-dependent and identity-independent aspects of expression are both encoded within multidimensional representational frameworks. The monotonic increase in anti-expression aftereffect magnitude in both congruent and incongruent conditions indicates that prototype-referenced schemes are used to encode identity-dependent representations and, crucially, the identity-independent representations that underlie our ability to recognize the expressions worn by unfamiliar faces.

Acknowledgments

Commercial relationships: none.
Corresponding author: Andrew L. Skinner.
Email: andy.skinner@bristol.ac.uk.
Address: School of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK.

References

Benton, C. P. (2009). Effect of photographic negation on face expression aftereffects. *Perception, 38,* 1267–1274.

Benton, C. P., Jennings, S. J., & Chatting, D. J. (2006). Viewpoint dependence in adaptation to facial identity. *Vision Research, 46,* 3313–3325.

Brennan, S. E. (1985). Caricature generator, the dynamic exaggeration of faces by computer + illustrated works. *Leonardo, 18,* 170–178.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327.

Butler, A., Oruc, I., Fox, C. J., & Barton, J. J. S. (2008). Factors contributing to the adaptation aftereffects of facial expression. *Brain Research, 1191,* 116–126.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research, 41,* 1179–1208.

Calder, A. J., Jenkins, R., Cassel, A., & Clifford, C. W. G. (2008). Visual representation of eye gaze is coded by a nonopponent multichannel system. *Journal of Experimental Psychology: General, 137,* 244–261.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods, 40,* 109–115.

Campbell, J., & Burke, D. (2009). Evidence that identity-dependent and identity-independent neural populations are recruited in the perception of five basic emotional facial expressions. *Vision Research, 49,* 1532–1540.

Cohn, J. F., Schmidt, K., Gross, R., & Ekman, P. (2002). Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. In *Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces* (pp. 491–496). Los Alamitos, CA: IEEE Computer Soc.

Cook, R., Matei, M., & Johnston, A. (2011). Exploring expression space: Adaptation to orthogonal and anti-expressions. *Journal of Vision, 11*(4):2, 1–9, http://www.journalofvision.org/content/11/4/2, doi:10.1167/11.4.2. [PubMed] [Article]

Cumming, G., & Finch, S. (2005). Inference by eye—Confidence intervals and how to read pictures of data. *American Psychologist, 60,* 170–180.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience, 14,* 1158–1173.

Dickinson, J. E., Almeida, R. A., Bell, J., & Badcock, D. R. (2010). Global shape aftereffects have a local substrate: A tilt aftereffect field. *Journal of Vision, 10*(13):5, 1–12, http://www.journalofvision.org/content/10/13/5, doi:10.1167/10.13.5. [PubMed] [Article]

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation, 1971* (pp. 207–283). Lincoln, NE: University of Nebraska Press.

Fang, F., & He, S. (2005). Viewer-centered object representation in the human visual system revealed by viewpoint aftereffects. *Neuron, 45,* 793–800.

Fox, C. J., & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research, 1127,* 80–89.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4,* 223–233.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology, 11,* 880–885.

Kesler-West, M. L., Andersen, A. H., Smith, C. D., Avison, M. J., Davis, C. E., & Kryscio, R. J. (2001). Neural substrates of facial emotion processing using fMRI. *Brain Research. Cognitive Brain Research, 11,* 213–226.

Knappmeyer, B., Thornton, I. M., & Bulthoff, H. H. (2001). Characteristic motion of human face and human form. *Perception, 30,* 33.

Lee, E., Kang, J. I., Park, I. H., Kim, J. J., & An, S. K. (2008). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Research, 157,* 77–85.

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

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

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces—KDEF*. CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, ISBN 91-630-7164-9.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Pessa, J. E., Zadoo, V. P., Adrian, E. K., Yuan, C. H., Aydelotte, J., & Garza, J. R. (1998). Variability of the midfacial muscles: Analysis of 50 hemifacial cadaver dissections. *Plastic and Reconstructive Surgery, 102,* 1888–1893.

Rhodes, G., Robbins, R., Jacquet, E., McKone, E., Jeffery, L., & Clifford, C. W. G. (2005). Adaptation and face perception: How aftereffects implicate norm-based coding of faces. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world* (pp. 213–240). Oxford, UK: Oxford University Press.

Robbins, R., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: Adapter position effects and neural models. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 570–592.

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8,* 107–113.

Rowland, D. A., & Perrett, D. I. (1995). Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications, 15,* 70–76.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. In *Yearbook of physical anthropology* (vol. 44, pp. 3–24). New York: Wiley-Liss.

Schweinberger, S. R., Zaske, R., Walther, C., Golle, J., Kovacs, G., & Wiese, H. (2010). Young without plastic surgery: Perceptual adaptation to the age of female and male faces. *Vision Research, 50,* 2570–2576.

Shah, R., & Lewis, M. B. (2003). Locating the neutral expression in the facial-emotion space. *Visual Cognition, 10,* 549–566.

Skinner, A. L., & Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial Expressions. *Psychological Science, 21,* 1248–1253.

Susilo, T., McKone, E., & Edwards, M. (2010). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research, 50,* 300–314.

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

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