Facial expression megamix: Tests of dimensional and category accounts of emotion recognition

Andrew W. Young[a],*, Duncan Rowland[b], Andrew J. Calder[a], Nancy L. Etcoff[c], Anil Seth[a], David I. Perrett[b]

[a]*MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK*
[b]*School of Psychology, University of St. Andrews, St. Andrews, Fife KY16 9JU, UK*
[c]*Department of Psychiatry, Harvard Medical School, Massachusetts General Hospital, Room 9150, Building 149, 13th Street, Charlestown, MA 02129, USA*

Abstract

We report four experiments investigating the perception of photographic quality continua of interpolated ('morphed') facial expressions derived from prototypes of the 6 emotions in the Ekman and Friesen (1976) series (happiness, surprise, fear, sadness, disgust and anger). In Experiment 1, morphed images made from all possible pairwise combinations of expressions were presented in random order; subjects identified these as belonging to distinct expression categories corresponding to the prototypes at each end of the relevant continuum. This result was replicated in Experiment 2, which also included morphs made from a prototype with a neutral expression, and allowed 'neutral' as a response category. These findings are inconsistent with the view that facial expressions are recognised by locating them along two underlying dimensions, since such a view predicts that at least some transitions between categories should involve neutral regions or identification as a different emotion. Instead, they suggest that facial expressions of basic emotions are recognised by their fit to discrete categories. Experiment 3 used continua involving 6 emotions to demonstrate best discrimination of pairs of stimuli falling across category boundaries; this provides further evidence of categorical perception of facial expressions of emotion. However, in both Experiment 1 and Experiment 2, reaction time data showed that increasing distance from the prototype had a definite cost on ability to identify emotion in the resulting morphed face. Moreover, Experiment 4 showed that subjects had some insight into which emotions were blended to create specific morphed images. Hence, categorical perception effects were found even though subjects were sensitive to physical properties of these morphed facial expressions. We suggest that rapid classification of prototypes and better across boundary discriminability reflect the underlying organisation of human categorisation abilities.

* Corresponding author.

1. Introduction

People are very skilled at understanding each other's facial expressions. We know that babies are very interested in faces (Johnson et al., 1991), and that they show precocious ability to respond to different facial expressions (Field et al., 1982). We also know that, for tests using a fixed range of alternative choices, certain configurations of facial features resulting from specific patterns of facial muscle movements are recognised throughout the world as corresponding to particular basic emotions (Ekman, 1992, 1994). Moreover, selective changes in ability to recognise emotion from the face have been reported after brain injury; sometimes, patients may remain able to recognise other social cues such as identity from the face, even though they have problems in recognising facial emotion (Calder et al., 1996b; Etcoff, 1984; Sprengelmeyer et al., 1996; Young et al., 1993). The physiological literature also suggests differences in the neural coding of facial identity and expression in other primates (Desimone, 1991; Hasselmo et al., 1989), and PET studies of humans have shown differences between brain regions involved in the analysis of identity and expression (Sergent et al., 1994), and demonstrated the possible existence of emotion-specific responses to facial expressions (Morris et al., 1996).

These facts are consistent with the long evolutionary history of facial expressions of emotion (Darwin, 1872; Ekman, 1973), but we know little about the perceptual basis of how emotions are recognised. One of the fundamental issues that is still disputed concerns whether facial expressions are perceived as varying continuously along certain underlying dimensions, or as belonging to qualitatively discrete categories (Ekman, 1982; Ekman et al., 1972). This issue has been difficult to resolve because many data can be accommodated within either view, and hybrid models are sometimes proposed. For example, Woodworth and Schlosberg (1954) identified happiness (in which they included love and mirth), surprise, fear (including suffering), anger (and determination), disgust and contempt as distinct, recognisable categories of emotion, but then suggested on the basis of their confusabilities that they can be considered to be located around the circumference of a circle (running happiness–surprise–fear–anger–disgust–contempt–happiness) with two orthogonal diagonals corresponding to the dimensions pleasant–unpleasant (running from the region of happiness to the region of anger) and attention–rejection (running from the surprise and fear boundary to the boundary between disgust and contempt). This suggestion is shown in Fig. 1a. A modern variant of this idea is the Russell (1980) circumplex model, in which more extreme degrees of an emotion fall around the edge of a two-dimensional emotion space encoding orthogonal bipolar dimensions of pleasure and arousal, with milder emotions falling more toward the centre.

The intention of such models is to create a two-dimensional solution to the problem of perceptually classifying facial expressions. Woodworth and Schlosberg (1954) drew an analogy to the colour circle, in which hue is arranged around the circumference and saturation along red–green and blue–yellow opponent axes.

In the present study, we took advantage of image-manipulation techniques to

(a)

LOVE, MIRTH, HAPPINESS

CONTEMPT

PLEASANT

SURPRISE

REJECTION   ATTENTION

DISGUST

FEAR,
SUFFERING

UNPLEASANT

ANGER,
DETERMINATION

(b)

HAPPINESS

ANGER   SURPRISE
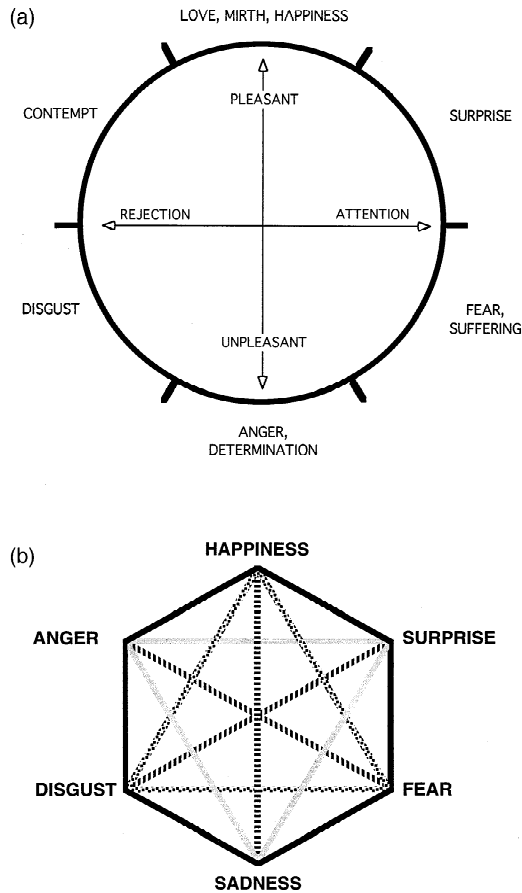
DISGUST   FEAR

SADNESS

Fig. 1. Schematic representations of relation between different facial expressions. (a) As proposed by Woodworth and Schlosberg (1954). (b) Continua used in Experiment 1.

examine contrasting predictions which follow from accounts in which facial expressions of emotion are perceived and recognised as discrete categories or by locating them in terms of a small number of underlying dimensions.

A consequence of dimensional accounts is that linear (or near-linear) physical transitions from one emotional expression to another will be accompanied by characteristic changes in identification. For example, in Woodworth and Schlosberg's (1954) schema (Fig. 1a), a transition from a happy face to an angry face cannot be effected without passing through a region close to the origin of the 'emotion space', where the expression should become neutral or indeterminate. This must happen in the Woodworth and Schlosberg (1954) account because happiness–anger corresponds to one of the hypothesised dimensions of the perceptual emotion space (pleasant–unpleasant), but the same will be true of any transition between expressions lying at opposite points in the emotion space.

Similarly, Fig. 1a shows that a transition from a happy to a frightened expression will not cross the origin of the two-dimensional emotion space, and thus need not involve any region where the expression becomes indeterminate, but it will enter a region where the expression should be seen as one of moderate surprise. In contrast, a transition from happiness to surprise does not involve entering the region of any other emotion, and may, therefore, be relatively abrupt, passing directly from one emotion to the other.

Dimensional accounts, then, can be used to predict the consequences for identification of physically transforming one facial expression to another. A method for making such transformations is available from image-manipulation procedures, for which algorithms have been devised to allow manipulation of photographic-quality images of faces (Benson and Perrett, 1991; Burt and Perrett, 1995). This is achieved by specifying the locations of a large number of facial feature points on prototype images. To create interpolated ('morphed') images depicting the continuum between two faces (say, pictures of a person with happy and sad expressions) in photographic quality, the positions of the features in one photograph are moved toward their positions in the other photograph, as if the image lies on a rubber sheet. The technique is described in detail later, and our figures show that highly regular changes can be achieved.

It is possible, then, to change one facial expression to another by computer morphing, and from the Woodworth and Schlosberg (1954) schema we can predict the consequences for identification if this dimensional account is correct. These predictions contrast with those of an account of the perceptual identification of facial expressions as discrete categories. From the discrete category standpoint, Fig. 1a is not an accurate geometric representation, since the perceptual space will be multi-dimensional. A transition from any one category to any other category may, therefore, never involve passing through any region which will not be assigned to one end-point category or the other. From this perspective, changes in identification will always be relatively abrupt, with no need for any region in any continuum between two prototype expressions to correspond to a third emotion.

Of course, predictions derived from Woodworth and Schlosberg's (1954) schema can only work fully to the extent that it is an accurate geometric representation, based on the correct underlying dimensions. This is important because we noted that other two-dimensional accounts, such as Russell's (1980) circumplex model, do use different dimensions and hence postulate slightly different positioning of some emotions. There can thus be differences in the precise predictions made by different variants of two-dimensional accounts. However, what is important is that all two-dimensional accounts of emotion classification must share the property that they will predict that transitions between facial expressions of emotion will be variable in their effects, according to how these particular expressions align in the emotion space, and that at least some transitions between two expressions will involve regions of indeterminacy or a third emotion. Regardless of the exact predictions as to when these types of transition will arise, the dimensional accounts thus contrast with category-based models, for which this property is not a necessary feature; in a category-based

account, the transition between any two emotions can always be perceptually abrupt.

In previous studies of identification of morphed images of facial expressions by normal perceivers (Calder et al., 1996a; Etcoff and Magee, 1992), only a limited number of continua have been tested at any given time. In a typical paradigm, subjects may be given a face varying along the continuum from expression '*x*' to expression '*y*', and asked whether its expression looks more like *x* or *y*. Sharp changes in identification along such continua have been found, but this Procrustean method may artificially exaggerate the abruptness of any transition, and hide the kinds of indeterminacy or outright shift to a different category predicted by dimensional accounts, since only the categories *x* or *y* are permitted as choices.

In Experiments 1 and 2, then, we tested the contrasting predictions of dimensional and category-based accounts by morphing between *all* possible pairs of facial expressions of basic emotions from the Ekman and Friesen (1976) series, and examining recognition of stimuli from all continua in a common set of trials. Our results were inconsistent with two-dimensional accounts, and favoured a category-based model.

A second type of prediction can be derived from the view that facial expressions of emotion are recognised as belonging to distinct perceptual categories. This prediction can also be seen using the analogy of Woodworth and Schlosberg (1954) to colour perception. One of the striking features of colour perception is that, although wavelength varies continuously, the appearance of the circumference of the colour circle to a person with normal colour vision involves discrete regions of relatively consistent colour, with comparatively abrupt transitions between. Psychophysical studies have confirmed that we are relatively insensitive to changes in wavelength if they occur within a region belonging to a colour category, and more sensitive to changes of the same physical magnitude occurring across the boundary between two colours (Bornstein and Korda, 1984). Many other examples of such phenomena have been described for perceptual categories that vary on a single underlying physical dimension (Harnad, 1987a). Collectively, they are known as categorical perception. The phenomenon widely regarded as the hallmark of categorical perception is that linear physical changes in a stimulus can have non-linear perceptual effects, with changes which occur near to or across category boundaries being easier to detect.

Morphed images have also made it possible to explore categorical perception effects with multidimensional stimuli, such as faces (Beale and Keil, 1995; Calder et al., 1996a; Etcoff and Magee, 1992). The pioneering study was by Etcoff and Magee (1992), who took advantage of these technical developments to investigate categorical perception of facial expressions; they converted photographs from the Ekman and Friesen (1976) series of pictures of facial affect into line drawings, and used a computer program to create several series of drawings representing equal interpolated steps between two different facial expressions posed by the same individual.

With these line drawing stimuli, Etcoff and Magee (1992) measured *identification* of the emotion seen in individual stimuli falling along a particular expression

continuum (e.g., from happiness to sadness), and *discrimination* between pairs of these stimuli with an ABX task (in which stimuli A, B, and X were presented sequentially; subjects had to decide whether X was the same as A or B). For identification, Etcoff and Magee (1992) observed sharp boundaries between a region of each continuum perceived as corresponding to one expression, and a region corresponding to the other expression. In the ABX discrimination task, Etcoff and Magee (1992) found that people were more accurate at detecting the differences between pairs of drawings which crossed a subjective category boundary (such as between a drawing seen as happy in the identification task and a drawing seen as sad) than they were at detecting equal physical differences which lay within a category (i.e., between two drawings which would be identified as happy, or two drawings identified as sad). This was clear evidence of categorical perception of facial expressions.

Calder et al. (1996a) used photograph-quality morphed images of expression continua to replicate Etcoff and Magee's (1992) findings of categorical perception of facial expressions with photographic-quality images. Calder et al. (1996a) also found sharp boundaries between the region of each continuum perceived as corresponding to one expression, and the region corresponding to the other expression, and they showed that subjects' discrimination performance could be predicted from their identification performance with the same stimuli; discrimination was poorest for stimulus pairs in which identification of the emotion was most consistent, and discrimination was best when identification was least consistent.

Findings of categorical perception of facial expressions again challenge the view that the perception of facial expressions is determined by a small number of underlying dimensions, such as pleasant–unpleasant. On a dimensional account, the perceptual function should be continuous from one emotion to the next; the categorical perception results show that it is not. In Experiment 3, we further examined this phenomenon, introducing a much more severe test of categorical perception than has been used in previous publications, by examining within-category and between-category discriminabilities for stimuli drawn from the six possible continua forming the perimeter of the hexagon shown in Fig. 1b. In addition, we adopted a task (simultaneous perceptual matching) which eliminated the memory requirements inherent in the ABX discrimination task used in studies based on methods from the speech perception literature. Enhanced discriminability was still found for stimuli falling across category boundaries.

Although the results of Experiments 1–3 were inconsistent with a two-dimensional account of the perceptual classification of facial expressions of emotion, there were also aspects of the data which indicated that subjects could none the less often see differences between stimuli they assigned to a common perceptual category; reaction times for identification were affected by the closeness of a morphed image to the nearest prototype, and within-category discriminabilities were not always at chance level. In Experiment 4, we therefore examined systematically the extent to which subjects could recognise exactly which prototypes particular morphed images had been interpolated between; although not perfect, they did have some insight.

These findings provide data which must be explained by any adequate model of facial expression recognition, and which allow us to reject two-dimensional accounts as inadequate models of the perceptual mechanisms involved. An account treating facial expressions as corresponding to discrete categories of emotion is more readily reconciled with our findings, but they are also inconsistent with the strongest possible variants of the categorical perception hypothesis. In our General Discussion, we examine which forms of category model can provide plausible accounts of the data we present.

EXPERIMENT 1

Experiment 1 investigated identification of morphed facial expressions taken from all possible continua between pairs of the six universally recognised emotions in the Ekman and Friesen (1976) series. We were interested to determine whether these morphed expressions would still behave like discrete categories when they could come from many possible continua and a 6-way choice was available, or whether some continua would show properties predicted by two-dimensional accounts of emotion recognition, including indeterminate central regions (when the end-points of the continuum lie on opposite sides of 'emotion-space') or intrusions of other categories into the middle of the continuum (when the end-points are non-adjacent but not opposite in emotion-space).

2. Method

2.1. Stimuli

The Ekman and Friesen (1976) series contains examples of six different emotional facial expressions (happiness, surprise, fear, sadness, disgust and anger) which have been found to be accurately recognised from forced-choice between these possible emotions in most cultures of the world. Since these are slightly different from the list of Woodworth and Schlosberg (1954) of distinct emotions (happiness, surprise, fear, anger, disgust and contempt), we began by examining the confusabilities of the Ekman emotions, to create a representation analogous to that of Woodworth and Schlosberg (1954). From the control data of Ekman and Friesen (1976), we assembled an overall confusion matrix for all of the faces in the series, showing how likely a particular emotion was to be confused with each of the other emotions. This confusion matrix was then used to order the emotions by their maximum confusabilities, placing each adjacent to the one it was most likely to be confused with.

The result ran happiness–surprise–fear–sadness–disgust–anger, with mean percentage confusabilities for each pair of expressions in this sequence being happiness and surprise 0.8%, surprise and fear 5.8%, fear and sadness 2.4%, sadness and disgust 2.7%, disgust and anger 6.4%. The ends of the sequence, anger and happiness, were then joined to create the perimeter of a hexagon, as

illustrated in Fig. 1b. This gives a representation which is very like the emotion space resulting from a computer network model developed by Morishima and Harashima (1993). Fig. 1b is also similar to the Woodworth and Schlosberg (1954) circle shown in Fig. 1a, but with their disgust and contempt regions collapsed into the single emotion of disgust, with their single fear and suffering region separated into distinct emotions of fear and sadness, and with the relative positioning of disgust and anger reversed. As will become clear later, it turns out that the exact ordering of the emotions around the hexagon is not crucial; it simply provides a convenient means of exposition.

Fig. 1b shows that there are 15 possible continua between pairs of the 6 emotional expressions. Six continua lie around the perimeter of the hexagon (happiness–surprise, surprise–fear, fear–sadness, sadness–disgust, disgust–anger, anger–happiness). Six more continua form two inner triangles in Fig. 1b (happiness–fear, fear–disgust, disgust–happiness; surprise–sadness, sadness–anger, anger–surprise). The other 3 continua are diagonals in Fig. 1b (happiness–sadness, surprise–disgust, fear–anger).

Face JJ (Ekman and Friesen, 1976) was chosen for morphing because the photographs of all 6 emotional facial expressions (happiness, surprise, fear, sadness, disgust and anger) were of consistent quality, with reasonably standardised pose and lighting. The numbers and identifiers of the selected prototype images in the Ekman and Friesen (1976) series were happiness 34 JJ-4-07, surprise 39 JJ-4-13, fear 37 JJ-5-13, sadness 36 JJ-5-05, disgust 40 JJ-3-20, and anger 38 JJ-3-12. Photographic-quality continua were made, morphing between every possible pair from these 6 prototype expression images in the manner charted in Fig. 1b.

Five morphed images were created for each continuum. These were prepared by blending between two prototype expressions posed by JJ (e.g., happiness and sadness) in proportions 90:10 (i.e., 90% happy 10% sad for the happiness–sadness continuum), 70:30 (70% happy 30% sad), 50:50 (50% happy 50% sad), 30:70 (30% happy 70% sad), and 10:90 (10% happy 90% sad). We will refer to these as 90%, 70%, 50%, 30% and 10% morphs along the appropriate continuum (in our example, happiness–sadness).

The preparation of each continuum was done in the same way. Preparation of the happiness–sadness continuum is described in detail to illustrate the process. The procedure involved three stages.

2.1.1. Stage 1: Delineation

One hundred and eighty six points were positioned manually onto the photograph of JJ's face with a happy expression from the Ekman and Friesen (1976) series. The locations of these points were specified in terms of anatomical landmarks, with each facial feature represented by a set number of points; for example, the mouth was represented by 22 points, and each eyebrow by 8 points. These points were then joined to produce a delineated representation comprising 50 feature contours. Exactly the same method was applied to the photograph of the same model posing sadness. Hence, across the two prototype expressions

(happiness and sadness) there was conformity with respect to the anatomical positioning of the 186 points on each face, but not always their exact spatial positions; for example, the corners of the mouth were turned up in the happy face and down in the sad face, but the shape of the hairline was the same in both.

2.1.2. Stage 2: Shape interpolation

A continuum of face shapes was generated between the two delineated prototype face shapes (in our example, JJ happy and JJ sad). This was achieved by taking the delineation data for the two prototype images and calculating the vector difference for each landmark. For example, consider the point describing the tip of the nose; this has a location on the JJ happy face of $(x1, y1)$ and a location on the JJ sad face of $(x2, y2)$. Equations describing the vector from $(x1, y1)$ to $(x2, y2)$ were used to obtain positions for the point at the tip of the nose which moved 10%, 30%, 50%, 70% and 90% along a straight line from the location of that point in JJ's happy face $(x1, y1)$ to the location of that point in JJ's sad face $(x2, y2)$. This process was repeated for each of the 186 feature points, to generate the 5 face shapes which would interpolate at 10%, 30%, 50%, 70% and 90% distances between the two original face shapes.

2.1.3. Stage 3: Producing a continuous-tone image

The final stage created a continuous-tone (photographic quality) image for each of these interpolated face shapes. This was achieved by taking both of the prototype faces and 'warping' or 'stretching' them (as if they were printed on a rubber sheet) to the new shape, so that all points representing the same feature were aligned across images. The two faces, now with the same intermediary face shape, were then blended with the appropriate weight. For example, in the 90% happy 10% sad morph, the pixel intensities in each tessellation were arrived at by deforming the happy face 10% toward the sad prototype, and the sad face 90% toward the happy prototype, and then blending the grey levels in these two contributory images in the ratio nine parts from the happy prototype to one part from the sad prototype.

The resulting morphed faces are shown in Fig. 2. In total, there are 75 images (5 from each of 15 continua). Moving from left to right in Fig. 2, the columns show 90%, 70%, 50%, 30% and 10% morphs along each continuum. Note that a 90% morph on the happiness–sadness continuum would be the same as a 10% morph on a sadness–happiness continuum, and that the prototype expressions are not shown in Fig. 2. Fig. 2a shows morphs from the perimeter of the outer hexagon in Fig. 1b; happiness–surprise (top row), surprise–fear (second row), fear–sadness (third row), sadness–disgust (fourth row), disgust–anger (fifth row) and anger–happiness (bottom row). Fig. 2b shows morphs from the inner triangles in Fig. 1b. The first inner triangle runs happiness–fear (top row in Fig. 2b), fear–disgust (second row), disgust–happiness (third row); the second inner triangle runs surprise–sadness (fourth row in Fig. 2b), sadness–anger (fifth row), anger–surprise (bottom row). Fig. 2c shows morphs from the diagonals in Fig. 1b;

A.W. Young et al. / Cognition 63 (1997) 271–313

(a)

Fig. 2. (a)

Fig. 2. (b)

Fig. 2. Expression continua used in Experiment 1. (a) Outer hexagon in Fig. 1b (happiness–surprise–fear–sadness–disgust–anger–happiness). Going from left to right, the columns show 90%, 70%, 50%, 30% and 10% morphs along each continuum. From top to bottom, the continua shown in each row are happiness–surprise (top row), surprise–fear (second row), fear–sadness (third row), sadness–disgust (fourth row), disgust–anger (fifth row), anger–happiness (bottom row). (b) Inner triangles in Fig. 1b (happiness–fear–disgust–happiness; surprise–sadness–anger–surprise). Going from left to right, the columns show 90%, 70%, 50%, 30% and 10% morphs along each continuum. From top to bottom, the continua shown in each row are happiness–fear (top row), fear–disgust (second row), disgust–happiness (third row); surprise–sadness (fourth row), sadness–anger (fifth row), anger–surprise (bottom row). (c) Diagonals in Fig. 1b (happiness–sadness; surprise–disgust; fear–anger). Going from left to right, the columns show 90%, 70%, 50%, 30% and 10% morphs along each continuum. From top to bottom, the continua shown in each row are happiness–sadness (top row); surprise–disgust (middle row); fear–anger (bottom row).

happiness–sadness (top row), surprise–disgust (middle row), fear–anger (bottom row).

2.2. Subjects

Eight members of the MRC Applied Psychology Unit subject panel acted as subjects in the experiment. All were between the ages 30 and 45 years and had normal or corrected-to-normal vision. They were paid for participating.

2.3. Design and procedure

The 75 morphed faces shown in Fig. 2 were used in an identification task. They were presented one at a time, in random order, on a 256 grey scale 16″ Macintosh colour monitor using Psyscope software. Each morphed face subtended a horizontal visual angle of 3.6° and a vertical visual angle of 5.2°. The subjects' task was to decide whether each image was most like happiness, surprise, fear, sadness, disgust or anger. Responses were made using a box with 6 labelled buttons (one for each emotion), which was interfaced with the computer to record the choice of emotion and the reaction time (from stimulus onset) for each morphed image. On each trial, the morphed image remained in view until a button press was made. The ordering of the 6 emotion labels on the response buttons was randomised across subjects. No feedback was given as to the appropriateness of any responses.

Each subject undertook a total of 9 blocks of trials, across 3 testing sessions. In each of these blocks of trials, all of the 75 morphed faces were presented once. The first block of trials was discounted as practice, leaving 8 blocks of 75 trials from each subject for analysis. Note that the prototype face for each expression was never shown in the experiment; all the stimuli were morphs. Note too that the morphs were presented in random order; they were not grouped into the underlying continua.

3. Results

Data from Experiment 1 involved frequencies with which each morphed image was identified as happiness, surprise, fear, sadness, disgust or anger, and reaction times to make these classifications. We will consider each in turn.

First, frequency of identification. Data are presented as percentages of choices corresponding to each emotion category for each morphed face, in the upper graphs of Fig. 3. The continua belonging to the outer hexagon of Fig. 1b form Fig. 3a (happiness–surprise–fear–sadness–disgust–anger–happiness; these are the morphs shown in Fig. 2a), the first inner triangle of Fig. 1b (happiness–fear– disgust–happiness; the morphs shown in the upper half of Fig. 2b) forms Fig. 3b, the second inner triangle of Fig. 1b (surprise–sadness–anger–surprise; shown in the lower half of Fig. 2b) forms Fig. 3c, and the diagonals of Fig. 1b (happiness– sadness; surprise–disgust; fear–anger; shown in Fig. 2c) form Fig. 3d. For each continuum, the morphs are numbered in a series corresponding to the percentage of morphing from the prototype at the left end on the graph; thus, the morphs in the first continuum, happiness–surprise (top left in Fig. 3a) are numbered 90 (the 90% morph; i.e., 90% happiness 10% surprise), 70 (70% happiness 30% surprise), 50 (50% happiness 50% surprise), 30 (30% happiness 70% surprise), 10 (10% happiness 90% surprise), then the next continuum (surprise–fear) begins and is numbered 90 (90% surprise 10% fear), 70 (70% surprise 30% fear), and so on.
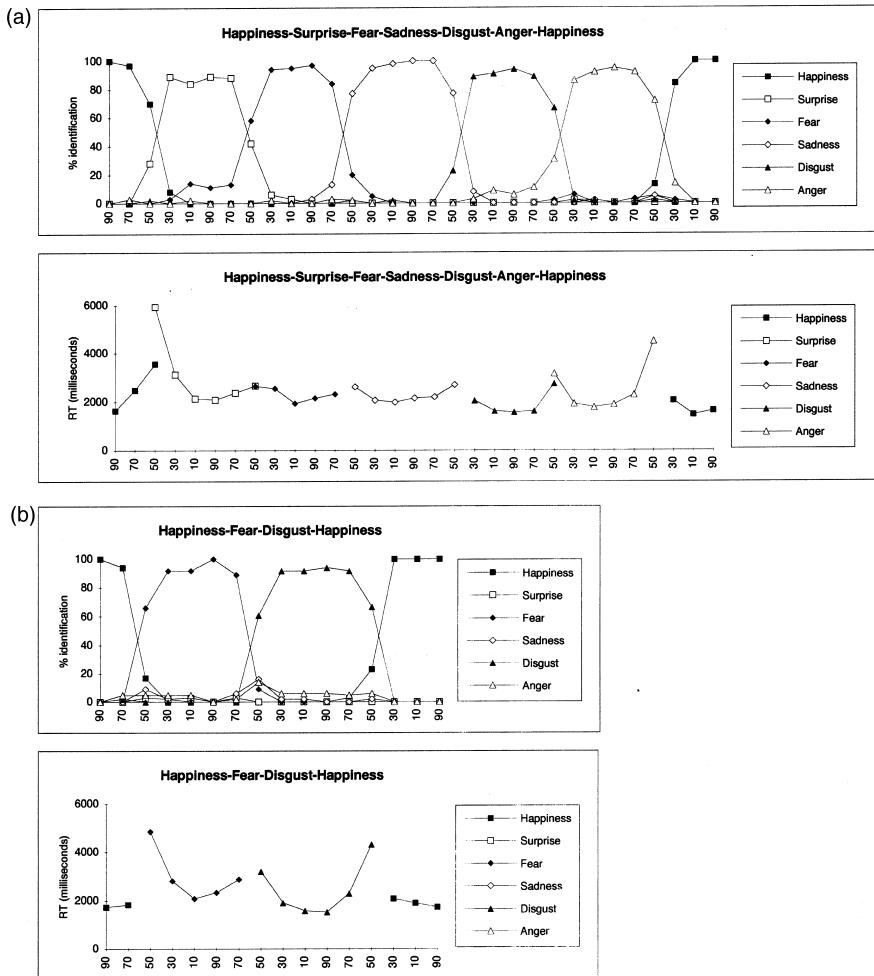
Fig. 3. Percentage identifications and reaction times (in ms) from Experiment 1. (a) Outer hexagon in Fig. 1b (happiness–surprise–fear–sadness–disgust–anger–happiness). (b) First inner triangle in Fig. 1b (happiness–fear–disgust–happiness). (c) Second inner triangle in Fig. 1b (surprise–sadness–anger–surprise). (d) Diagonals in Fig. 1b (happiness–sadness; surprise–disgust; fear–anger).

Had the prototype images been presented in the experiment, each would thus lie between the appropriate points numbered 10 and 90.

These identification data form a clear and highly consistent pattern. We can set upper and lower confidence limits at the 0.05 level using the binomial theorem. For every morphed image there are 64 data points, each involving a 6-way choice. Chance responding in the identification task would therefore lead to responses in the range 8–25% for choice of any particular emotion. The majority of data points in Fig. 3 are outside this range. Many are of course at 0%, showing that the morphed images do *not* look much like those emotions, and then there is a clear
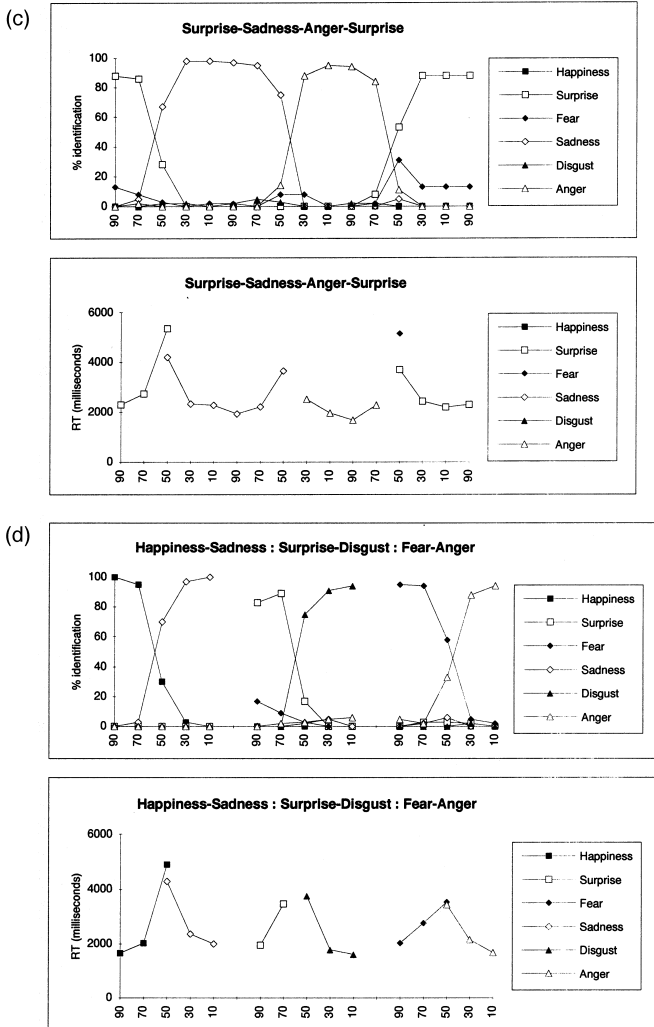
(c)

Surprise-Sadness-Anger-Surprise

Surprise-Sadness-Anger-Surprise

(d)

Happiness-Sadness : Surprise-Disgust : Fear-Anger

Happiness-Sadness : Surprise-Disgust : Fear-Anger

Fig. 3. (*continued*)

region corresponding to each perceived emotion where the choice of that emotion rises well beyond 25%. For every emotion, the 30% and 10% morphs approaching the prototype, and 90% and 70% going away from the prototype, are always classified as corresponding to that emotion more than 80% of times. This applies to *all* of the 15 continua plotted in Fig. 3, regardless of where they come from in the hexagonal representation used for Fig. 1b.

For identification, then, every continuum fell into two clear regions, each of which belonged primarily to the emotion category corresponding to the prototype at that end, with an abrupt shift from one category to the other at or near the centre

of the continuum. A particularly interesting feature of these regions where there were abrupt shifts in identification at the midpoint of the continuum is that there was little tendency for other categories to intrude. For example, we noted in Section 1 that in the schema of Woodworth and Schlosberg (1954) (Fig. 1a), a transition from a happy face to an angry face will pass close to the origin of the 'emotion space', where the expression should become indeterminate, and a transition from a happy to a frightened expression will enter a region where the expression should be seen as one of moderate surprise. However, the data for these continua simply show direct shifts from one prototype to the other; identification of images in the anger–happiness continuum forms the rightmost section of the top graph in Fig. 3a, and the happiness–fear continuum forms the leftmost section of the top graph in Fig. 3b. In other words, the half-way points on these continua were still seen as transitions between anger and happiness in the anger–happiness continuum, and between happiness and fear in the happiness–fear continuum. They were *not* perceived as regions of generalised uncertainty, or as any other emotion from the range used.

To test this property formally, we looked at identification of the 50% morphs for each of the 15 continua shown in Fig. 3. If these morphs lying at the mid-point of each continuum were still perceived as belonging to a particular continuum, identification as either of the two prototypes (e.g., as happiness or sadness for the happiness–sadness continuum) should be at higher rates than identification as any other emotion (e.g., as surprise, fear, disgust or anger for the happiness–sadness continuum). This held true for the 50% morphs from 13/15 of the continua shown in Fig. 3 ($p < 0.001$ on binomial theorem). The exceptions were fear–disgust (Fig. 3b; disgust was mostly seen, but sadness, anger and fear occurred at low rates) and anger–surprise (Fig. 3c; surprise was mostly seen, but then there was a clear intrusion of fear, followed by anger).

Although responses for most continua were largely in terms of the two prototype expressions, there were some cases where low but consistent rates of identification as an alternative expression arose. Inspection of Fig. 3 shows that surprise was sometimes identified as fear, and disgust as anger. When this happened, it was usually because one subject consistently did this. These alternative, non-modal identifications (surprise as fear, disgust as anger) are known to occur to prototype expressions in the Ekman and Friesen (1976) set, and they are known to be consistent across cultures (Ekman, 1972). Interestingly, when they occurred here, they were made throughout the range of morphs belonging to each prototype; for example, the regions around the surprise prototype in Fig. 3 (a, c and d) showed consistent alternative classification as fear.

We now turn to consider reaction times. These were calculated for all cases where identification rates were above the chance level of 25% given by the binomial theorem. Mean reaction times (in ms) are shown in the lower graphs of Fig. 3, with reaction time for each morph positioned directly below the corresponding identification rate (upper graphs in Fig. 3).

The reaction times showed a very clear scalloping. For every response category, RT was inversely related to distance from the prototype; for example, when a

morphed expression was identified as 'happiness', responses were fastest to 90% happy morphs, slower to 70% happy, and slower still to 50% happy. To confirm this, reaction times for each continuum were collapsed into those for morphs closest to the prototype (90% and 10% morphs in each continuum), morphs further from the prototype (70% and 30% morphs), and morphs furthest from the prototype (50% morphs). A one-factor Analysis of Variance showed a highly significant effect of distance from the prototype on reaction time ($F = 67.57$, df 2,28, $p < 0.001$); Tukey tests ($\alpha = 0.05$) confirmed that RTs were fastest for near-prototype morphs, and significantly slower with each increase in distance from the prototype.

Of course, identification rates also fell off with increasing distance from the prototype, and strong negative correlations were found between reaction time and identification rate. The statistical significance of these correlations was determined by converting each to a t-value (McNemar, 1962). For example, when RT and identification rate were correlated for the morphs identified as happiness (by pooling data across the happiness–surprise, happiness–fear, happiness–sadness, happiness–disgust and happiness–anger continua), a highly significant negative correlation was observed ($r = -0.95$, $t = 9.96$, df 10, $p < 0.001$). Significant negative correlations of reaction time and identification rate were also found for all of the other emotions (surprise, $r = -0.81$, $t = 4.83$, df 12, $p < 0.001$; fear, $r = -0.81$, $t = 4.83$, df 12, $p < 0.001$; sadness, $r = -0.90$, $t = 7.49$, df 13, $p < 0.001$; disgust, $r = -0.85$, $t = 5.68$, df 12, $p < 0.001$; anger, $r = -0.67$, $t = 3.02$, df 11, $p < 0.02$), and for an analysis in which the data were pooled across all 6 expressions ($r = -0.81$, $t = 12.41$, df 80, $p < 0.001$).

What the reaction time data show, then, is that distance from the prototype has a definite cost on ability to categorise the resulting morphed image.

4. Discussion

The results of Experiment 1 showed three main features. First, identifications of the morphed faces belonging to each continuum were mainly in terms of the prototypes at either end of the relevant continuum, with high identification rates at each end, and a sudden shift in the middle. This result is particularly impressive because the prototype images were never shown in Experiment 1; all of the stimuli were morphs. Second, in the regions where identification changed from one emotion category to another, there were few intrusions from categories which did not belong to that continuum. Hence, these regions were seen as involving transitions between categories, *not* as regions of generalised uncertainty or as regions where any other emotion from the range chosen was seen.

These features of the data are clearly inconsistent with the predictions of two-dimensional accounts of facial expression perception. These accounts vary in details of the exact geometric positioning of the 6 basic emotions we tested. However, they all concur in the view that at least some of the continua we tested

should lie at opposite points in emotion space, in which case the central region of the continuum should be of indeterminate emotion since it lies near the geometric origin, and some other continua should enter regions where the expression should be seen as an emotion other than the two prototypes used to create the continuum. Neither property was found. Instead, it is clear that a two-dimensional geometric interpretation of Fig. 1b will not hold; in effect, any continuum from one prototype to another passes directly between those two prototypes, and does not enter any other region. This is consistent with a category-based model, and inconsistent with a two-dimensional account.

However, the third main feature of the data from Experiment 1 is also inconsistent with the strongest possible variants of category-based model. In Experiment 1, the reaction time data showed that distance from the prototype had a definite cost on ability to classify the resulting morph; morphs which lay closest to the prototype expressions used to create them were most easily identified. This implies that, at some level, subjects were picking up differences from the category prototype, and that all morphed images assigned to the same emotion category were therefore not treated as equally good examples. We will explore the wider implications of this point later. For now, we simply note that in general the findings of Experiment 1 provide evidence inconsistent with a two-dimensional account of emotion perception.

However, we considered that a possible limitation on the results of Experiment 1 might be that our forced-choice identification task did not include 'neutral' as a possible alternative. Some dimensional theories explicitly state that the origin of the emotion-space is neutral (Russell, 1980), and, therefore, predict that the mid-point of continua passing near to or through the origin of the emotion space might be more readily seen as neutral rather than emotional. Therefore, we reasoned that the abrupt discontinuities and absence of mid-continuum intrusions noted in the identification curves for Experiment 1 might not be found if this (neutral) choice was available. This was investigated in Experiment 2.

EXPERIMENT 2

Experiment 2 included 'neutral' as a possible response alternative. As explained, we thought it possible that the mid-points of some continua might be more readily seen as neutral than emotional. On a dimensional account of emotion, this would apply especially to continua which lie along the hypothesised dimensions and, therefore, pass through the centre of the emotion space, such as happiness–anger in Fig. 1a.

In order to ensure that neutral was a valid choice of response for at least some stimuli, Experiment 2 included continua in which all of the 6 emotional expressions from the Ekman and Friesen (1976) series were morphed toward a neutral prototype. This also allowed us to explore whether morphs from continua including a neutral prototype would behave any differently to those belonging to continua created between any of the other emotions.

5. Method

5.1. Stimuli

The photograph of JJ's face with a neutral expression (photograph number 41, identifier JJ-3-04) from the Ekman and Friesen (1976) series was morphed with the 6 prototype expressions selected for Experiment 1, to create a further 6 continua running from each of the 6 emotional expressions to neutral. These emotion-to-neutral continua are shown in Fig. 4. From left to right, the columns show 90%, 70%, 50%, 30% and 10% morphs from an emotional to a neutral expression; happiness–neutral (top row), surprise–neutral (second row), fear–neutral (third row), sadness–neutral (fourth row), disgust–neutral (fifth row), anger–neutral (bottom row). Note again that the 90% morph on the happiness–neutral continuum is the same as a 10% morph on a neutral–happiness continuum, and that the prototype neutral expression is not shown in Fig. 4.

These 6 emotion-to-neutral continua were combined with the 15 emotion-to-emotion continua used in Experiment 1, to create a pool of 105 morphed faces from the 21 continua representing every pairwise combination of happiness, surprise, fear, sadness, disgust, anger and neutral.

5.2. Subjects

Eight members of the MRC Applied Psychology Unit subject panel participated in the experiment. All were between the ages 20 and 35 years and had normal or corrected-to-normal vision. They were paid for participating.

5.3. Design and procedure

The 105 morphed faces were used in an identification task. The general design and procedure were as for Experiment 1; the principal differences resulted from the inclusion of the emotion-to-neutral continua into every block of trials, giving 21 continua in total (105 stimuli), and the inclusion of neutral as a possible identification response. The subjects' task was thus to decide whether each image was most like happiness, surprise, fear, sadness, disgust, anger or neutral. They made their responses using a box with 7 labelled buttons (one for each possible choice), which was interfaced with the computer to record the choice of emotion and the reaction time for each morphed image. Each subject performed 6 blocks of trials with the 105 morphs, across 2 sessions. As in Experiment 1, no feedback was given but the first block was discounted as practice, leaving data for each subject consisting of identification choices and reaction times in 5 blocks of 105 trials.

6. Results

Data from Experiment 2 involved frequencies with which each morphed image

Fig. 4. Continua including a neutral expression used in Experiment 2. From left to right, the columns show 90%, 70%, 50%, 30% and 10% morphs along each continuum from an emotional to a neutral expression. From top to bottom, the continua shown in each row are happiness–neutral (top row), surprise–neutral (second row), fear–neutral (third row), sadness–neutral (fourth row), disgust–neutral (fifth row), anger–neutral (bottom row).
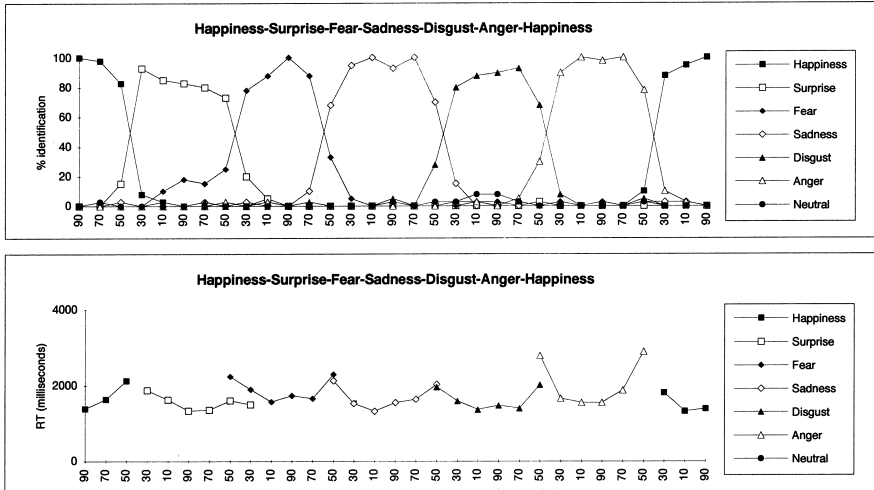
was identified as happiness, surprise, fear, sadness, disgust, anger or neutral, and reaction times to make these classifications. We will consider each in turn; analyses followed the procedures used for Experiment 1.

Data for frequency of identification are presented as percentages of choices corresponding to each emotion category for each morphed face, in the upper graphs of Fig. 5. The continua belonging to the outer hexagon of Fig. 1b are shown in Fig. 5a (happiness–surprise–fear–sadness–disgust–anger–happiness), the first inner triangle of Fig. 1b (happiness–fear–disgust–happiness) forms Fig. 5b, the second inner triangle of Fig. 1b (surprise–sadness–anger–surprise) forms Fig. 5c, and the diagonals of Fig. 1b (happiness–sadness; surprise–disgust; fear–anger) are shown in Fig. 5d. These, of course, correspond to the continua used in Experiment 1, shown in Fig. 3. What was new in Experiment 2 were the continua including neutral expression (morphs shown in Fig. 4); data for these are presented in Fig. 5e (happiness–neutral–surprise; fear–neutral–sadness; disgust–neutral–anger). As was done previously (in Fig. 3), the morphs are numbered in Fig. 5 according to the percentage of the prototype at the left end on the graph; thus, the first morph in the first continuum, happiness–surprise (top left in Fig. 5a) is numbered 90 (the 90% morph; i.e., 90% happiness 10% surprise). As before, had the prototype images been presented in the experiment, each would thus lie between the appropriate 10% and 90% morphs.

The data for the continua which were common to Experiments 1 and 2 (Figs. 5a–d) show a closely comparable pattern. For Experiment 2, we can set upper and lower confidence limits at the 0.05 level using the binomial theorem. For every morph there are 40 data points, each involving a 7-way choice. Chance responding in the identification task would therefore lead to responses in the range 5–25% for choice of any particular emotion. The majority of data points in Figs. 5a–d are outside this range. Many are at 0%, showing again that the morphs do not look much like those emotions, and then there is a clear region corresponding to each perceived emotion where the 30% and 10% morphs approaching the prototype, and 90% and 70% morphs going away from the prototype, are mostly classified as corresponding to that emotion.

As in Experiment 1, then, every one of the emotion-to-emotion continua (Figs. 5a–d) fell into two clear regions, each of which was identified as belonging primarily to the emotion category corresponding to the prototype at that end, with an abrupt shift from one category to the other at or near the centre of the continuum. Despite the inclusion of neutral as a possible response, there were still few intrusions into the regions where there were abrupt shifts in identification. To test this, we looked at identification of the 50% morphs in each of the 15 continua shown in Figs. 5a–d. For 13/15 of these continua, identification as either of the two prototypes was at higher rates than identification as any other emotion ($p < 0.001$ on binomial theorem). The exceptions were exactly the same as for Experiment 1; at the midpoints of the fear–disgust continuum (Fig. 5b; disgust was mostly seen, but sadness and fear occurred at low rates) and the anger–surprise continuum (Fig. 5c; surprise was mostly seen, but then there was a clear intrusion of fear, followed by anger).
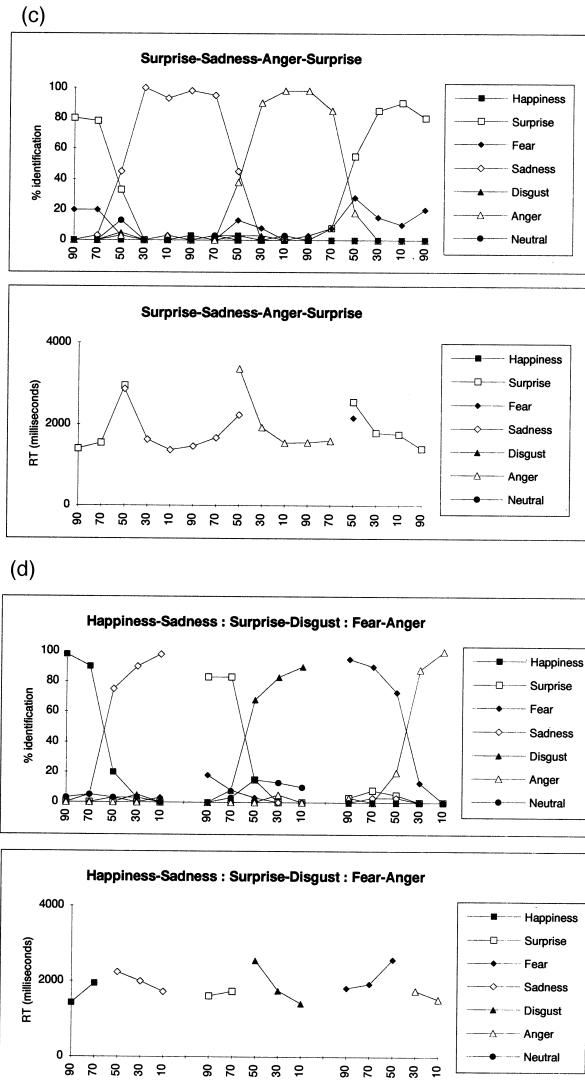
(a)

(b)

Fig. 5. Percentage identifications and reaction times (in ms) from Experiment 2. (a) Outer hexagon in Fig. 1b (happiness–surprise–fear–sadness–disgust–anger–happiness). (b) First inner triangle in Fig. 1b (happiness–fear–disgust–happiness). (c) Second inner triangle in Fig. 1b (surprise–sadness–anger–surprise). (d) Diagonals in Fig. 1b (happiness–sadness; surprise–disgust; fear–anger). (e) Continua including neutral expression shown in Fig. 4 (happiness–neutral–surprise; fear–neutral–sadness; disgust–neutral–anger).

As was noted in Experiment 1, surprise was sometimes consistently identified as fear, but in Experiment 2 no-one classified disgust as anger. In Experiment 2, however, disgust was sometimes identified as neutral (see Figs. 5a,b,d,e). As

(c)

(d)

Fig. 5. (*continued*)

before, then, these differences happen when a subject used an alternative classification of the emotions seen in JJ's prototype expressions.

In sum, the identification data for the 15 continua which were common to Experiments 1 and 2 showed a very high comparability across experiments. The inclusion of neutral as a possible response category in Experiment 2 had little discernible effect, and did not lead to morphs at the mid-points of emotion-to-emotion continua being identified as neutral.

(e)
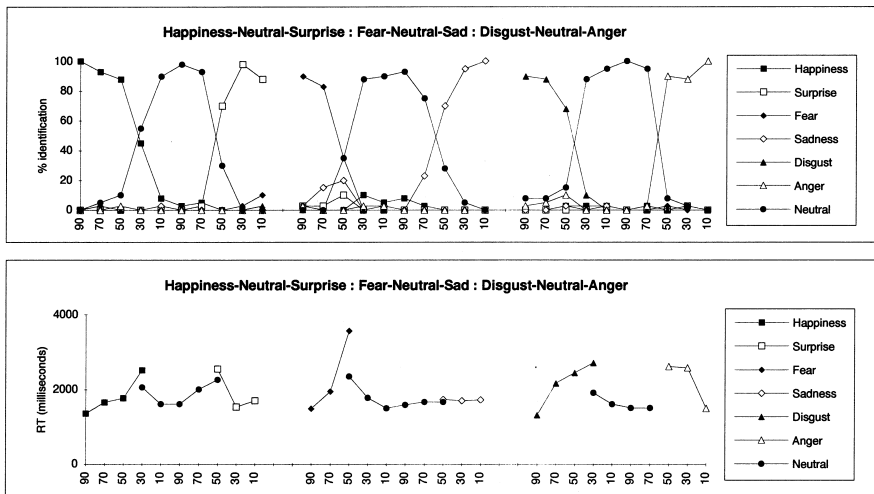
Fig. 5. (*continued*)

We now consider identification rates for the emotion-to-neutral continua, shown in the upper graph of Fig. 5e. They were similar to the emotion-to-emotion continua (Figs. 5a–d), except that the neutral region could be less broad; the discontinuity between identification rates occurred at the 30% morph for the happiness–neutral continuum. Hence, subjects see happiness as a broader category than neutral, which might well be expected as we are often looking for subtle cues to moods like happiness. Importantly, though, wherever the crossovers between neutral and emotional identifications occurred on each of the 6 emotion-to-neutral continua, there was again no tendency for other emotion categories to intrude ($p < 0.001$).

Reaction times were calculated for all cases where identification rates rose above the chance level of 23% given by the binomial theorem. These are presented in the lower graphs of Fig. 5.

As for Experiment 1, the reaction times showed a very clear scalloping; RT was inversely related to distance from the prototype for every response category. To confirm this, we again used the procedure of collapsing reaction times for each continuum into those for morphs closest to the prototype (90% and 10% morphs in each continuum), morphs further from the prototype (70% and 30% morphs), and morphs furthest from the prototype (50% morphs). A one-factor Analysis of Variance showed a highly significant effect of distance from the prototype on reaction time ($F = 73.34$, *df* 2,40, $p < 0.001$); Tukey tests ($\alpha = 0.05$) confirmed that RTs were fastest for near-prototype morphs, and significantly slower with each increase in distance from the prototype.

Again, strong negative correlations were found between reaction time and identification rate in Experiment 2. For example, when RT and identification rate

were correlated for the morphs identified as happiness (by pooling data across the happiness–surprise, happiness–fear, happiness–sadness, happiness–disgust, happiness–anger and happiness–neutral continua), a highly significant negative correlation was observed ($r = -0.84$, $t = 5.72$, df 14, $p < 0.001$). Significant negative correlations of reaction time and identification rate were also found for all of the other emotions (surprise, $r = -0.78$, $t = 4.70$, df 14, $p < 0.001$; fear, $r = -0.63$, $t = 3.21$, df 16, $p < 0.01$; sadness, $r = -0.91$, $t = 9.21$, df 17, $p < 0.001$; disgust, $r = -0.73$, $t = 4.36$, df 17, $p < 0.001$; anger, $r = -0.78$, $t = 4.64$, df 14, $p < 0.001$; neutral, $r = -0.70$, $t = 3.51$, df 13, $p < 0.01$), and for an analysis in which the data were pooled across all 7 expressions ($r = -0.71$, $t = 10.79$, df 119, $p < 0.001$).

7. Discussion

Even though Experiment 2 included neutral as a permissible response, results for the continua that were common to Experiment 1 and Experiment 2 were highly consistent across experiments, demonstrating unequivocally that the pattern found in Experiment 1 was not affected by the inclusion of neutral as a possible response in Experiment 2. In addition, the 6 emotion-to-neutral continua used for Experiment 2 each showed a marked discontinuity between a region identified as the appropriate emotion and a region identified as neutral. There was no tendency for other emotion categories to intrude into the crossover regions.

These findings again provide further evidence inconsistent with the predictions of a two-dimensional account of facial expression perception, for which emotion-to-emotion continua which lie close to or along the hypothesised dimensions (such as happiness–anger in Fig. 1a, or happiness–sadness in Fig. 1b), or at opposite points in the emotion space, should move through a region with no identifiable expression in the middle; in this region subjects should respond at random or perceive the emotion as neutral, yet this was never found.

In addition, Experiment 2 showed that neutral behaved in much the same way as the other emotion categories. Etcoff and Magee (1992) had also found sharp category boundaries between emotional and neutral faces. They suggested that emotions do not just dwindle down to nothing, but instead lose signal value abruptly at a certain point, and that there really is a state seen as 'neutral' which is its own category. Our data are entirely consistent with this position.

However, in both Experiment 1 and Experiment 2, the reaction time data showed that distance from the prototype had a definite cost on ability to classify the resulting morphed expression. This is inconsistent with the strongest possible form of categorical perception hypothesis, on which differences within a category should not be perceived at all. This strong version of the categorical perception hypothesis is therefore untenable for facial expressions, but it probably does not apply to other perceptual categories either; we can still see differences between different reds, or different yellows, it is just that we see the effect of an equivalent

difference in wavelength more easily if it crosses a boundary between colour categories (Bornstein and Korda, 1984).

For these reasons, the key indicator of categorical perception is nowadays taken to be the existence of better discrimination across category boundaries than within categories (Harnad, 1987b). This has been found in previous studies of categorical perception of facial expressions (Calder et al., 1996a; Etcoff and Magee, 1992). A potentially important procedural difference, though, is that these studies used morphs involving only 2 or 3 categories in any given task, whereas here we used 6 basic emotion categories in Experiment 1, and 7 categories for Experiment 2. We therefore decided to determine whether the hallmark categorical perception effect, enhanced discriminability of stimuli falling across the boundaries between perceptual categories, could be found when 6 emotion categories were used to generate stimuli. In addition, we made use of a simultaneous perceptual matching task which eliminated the memory requirements inherent in the more usual ABX discrimination task.

EXPERIMENT 3

Experiment 3 used sequential or simultaneous discrimination tasks. In the sequential (ABX) discrimination task, stimuli A, B and X were presented in sequence; subjects had to decide whether X was the same as A or B. The pairs to be compared (A, B) were taken from the morphs in continua belonging to the outer hexagon of Fig. 1b; these are the stimuli shown in Fig. 2a. This is the standard form of task used in the categorical perception literature; our aim was to determine whether better discrimination would be found for morphs falling across the boundaries between perceptual categories than for morphs that were close to the category prototypes.

The simultaneous discrimination task used a perceptual matching procedure in which pairs of stimuli involving identical (same) or adjacent (different) morphs from continua belonging to the outer hexagon of Fig. 1b (the images shown in Fig. 2b) were presented simultaneously, and subjects were asked to decide whether they were same or different. The advantage of this form of task is that it can be based on a genuine perceptual discrimination, with no necessary memory component; the aim was again to determine whether better discrimination would be found for morphs falling across the boundaries between perceptual categories than for morphs that were close to the category prototypes.

8. Method

8.1. Stimuli

The six continua forming the outer hexagon of Fig. 1b were used as stimuli; these are the 30 morphed images shown in Fig. 2a.

8.2. Subjects

Twenty-four members of the MRC Applied Psychology Unit subject panel participated in the experiment. None of these had taken part in Experiment 1 or Experiment 2. All were between the ages 25 and 50 years, and had normal or corrected-to-normal vision. They were paid for participating. Twelve were assigned to the sequential (ABX) discrimination task, and 12 to the simultaneous (same–different) matching task.

8.3. Design and procedure

First, we describe the ABX discrimination task. On each trial, a central fixation cross was presented for 250 ms, followed by a blank interval of 250 ms and then three successive face images. The first and second faces (A and B) were displayed for 750 ms each, and the third face (X) for 1 s; they were separated by blank intervals of 1 s. These times were the same as those used for ABX discrimination by Etcoff and Magee (1992) and Calder et al. (1996a). Faces A and B were always different, and on half of the trials face X was identical to face A and on the other half face B. Subjects were asked to make a button-press response to indicate whether X matched A or B. Choices of A or B, and reaction times from the onset of stimulus X were recorded.

Faces A and B differed by 1 step around the perimeter of the hexagon shown in Fig. 1b, using the images presented in Fig. 2a. All 30 of the possible face pairs were presented six times each during a single testing session, giving a total of 180 trials. The stimuli were presented on a 256 grey scale 16″ Macintosh colour monitor using Psyscope software. Each face subtended a horizontal visual angle of 3.6° and a vertical visual angle of 5.2°.

To acquaint subjects with the ABX procedure, the experiment began with 20 practice trials, selected at random. So that the subjects could assess their performance in this practice phase, an error was signalled by a bleep from the computer; there was no feedback in the experimental trials.

In the simultaneous (same–different) matching task, a central fixation cross was presented for 250 ms, followed by the two simultaneous face images, which remained in view until subjects responded, or for a maximum of 3500 ms if no response had been made at that time. The face images were aligned horizontally on the computer display screen, with one to the right and one left of centre, separated by a vertical gap of 1.4°. All 30 of the different face pairs created by taking adjacent images in Fig. 2a were presented six times each during a single testing session, randomly interleaved with an equivalent number of trials where the two images were exactly the same as each other, giving a total of 360 trials. Subjects were asked to make a button-press response to indicate whether each pair involved 'same' (identical) or slightly 'different' images. Choices and reaction times from stimulus onset were recorded. All other details of design and procedure were as for the ABX task.

9. Results

Both accuracy of discrimination and reaction times for correct choices were recorded. Our principal interest was in discrimination accuracy, and we will only use the RTs to establish that there was no overall speed–accuracy tradeoff. Note that subjects in Experiment 3 were never asked to identify any of the stimuli in terms of their emotion category, and that none had taken part in Experiment 1 or Experiment 2.

The upper graph in Fig. 6 shows a summary of identification data for the relevant continua taken from Experiment 1 (the perimeter of the hexagon in Fig. 1b: happiness–surprise–fear–sadness–disgust–anger–happiness), and the three lower graphs show summaries of discrimination accuracy for each pair of stimuli in Experiment 3.

First, there is a summary of sequential discrimination accuracy for the ABX task, giving the percentages of correct choices. In this graph, the labels along the x-axis represent the pairs of morphs being discriminated; for example, pair 90,70 in the happiness–surprise continuum (leftmost in the lower graphs of Fig. 6) involves the 90% morph (90% happiness 10% surprise) and the 70% morph (70% happiness 30% surprise). On this numbering convention, the 10,90 discriminations lie across the locations of the prototype images; for example, the 10,90 discrimination in the happiness–surprise–fear region involves the 10% morph on the happiness–surprise continuum (10% happiness 90% surprise) and the 90% morph on the surprise–fear continuum (90% surprise 10% fear).

The two lowest graphs in Fig. 6 show performance in the simultaneous discrimination task. This is represented in terms of the percentage of 'different' pairs correctly recognised as different, which is the obvious measure to take since subjects tend to look for differences between the stimulus pairs, responding 'same' whenever they fail to detect such a difference. Because there is thus a natural bias toward responding 'same' whenever differences were not detected (which explains why the percentages of correct responses to 'different' pairs can sometimes be less than 50%), we also calculated sensitivity values for the signal detection statistic d', which is uncontaminated by response biases.

To analyze these data, we began by using the same procedure as Calder et al. (1996a), which is based on a method which is widely adopted in the categorical perception literature. Subjects' performance in each discrimination task was predicted from their identification rates, and these predictions were compared with the observed discrimination results (Liberman et al., 1957).

The formula used to calculate subjects' predicted performance was identical to that used by Calder et al. (1996a). This formula is based on the view that discrimination will depend on two contributory factors; first, subjects' ability to make use of the physical differences between each pair of morphed images in a continuum, regardless of their expressions, and, second, any additional contribution from categorical perception of the facial expressions. As an estimate of the first factor (ability to make use of the physical differences between each pair of morphed images) we used the mean of the discriminabilities for the pairs at the
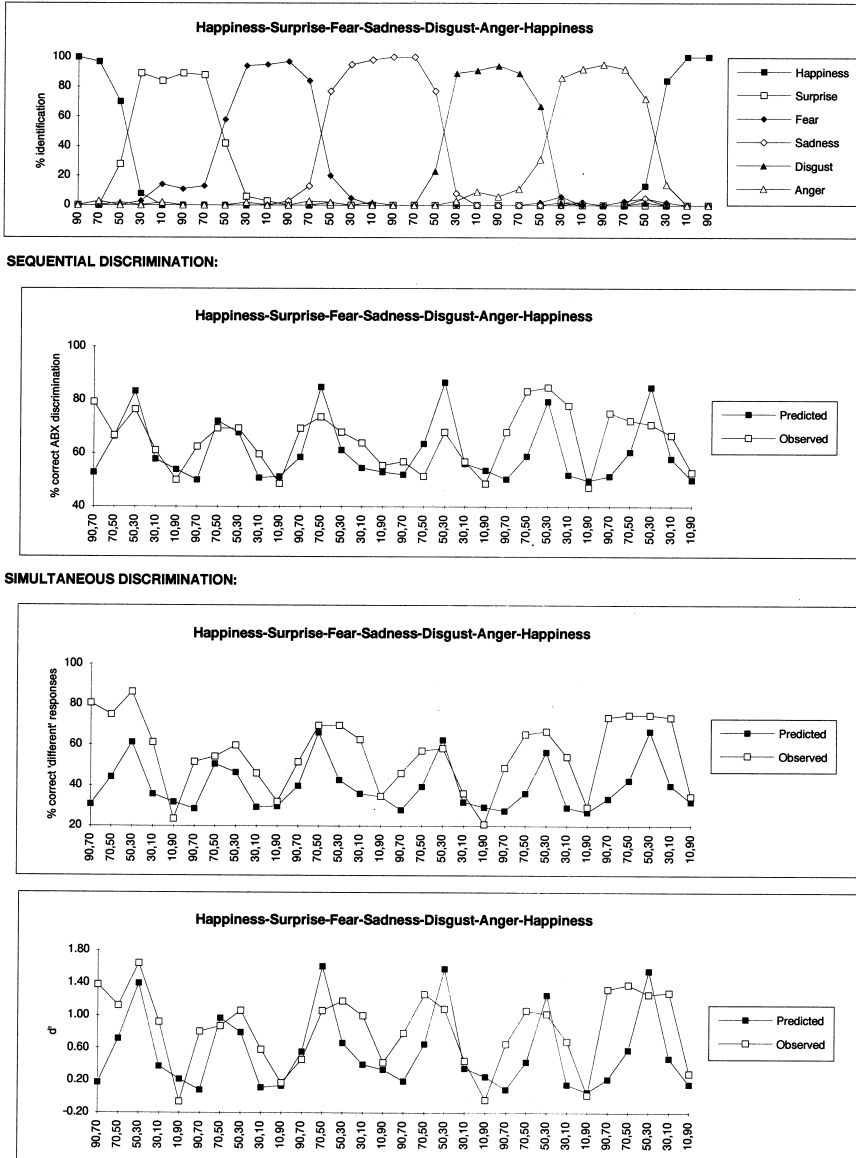
Fig. 6. Percentage identifications from Experiment 1 (top graph) and predicted and observed discrimination performance (lower graphs) in Experiment 3 (using the outer hexagon in Fig. 1b: happiness–surprise–fear–sadness–disgust–anger–happiness). Discrimination performance is shown in terms of percent correct sequential (ABX) discrimination, and for simultaneous discrimination as percent correct 'different' responses and as d'.

ends of each continuum (i.e., for the happiness–surprise continuum, we used the average discrimination of morphs 10,90 lying at each end). To estimate the second factor (contribution from categorical perception of the facial expressions), we used the sum of the differences between the percentage rates at which the members of each pair were assigned to each expression category in the identification task (i.e., using the identification data from Experiment 1). These two factors were then combined to yield the predicted performance curve for each continuum; for the percent correct measures this was done by adding 0.25 of the identification difference for the relevant pair to the estimate of the overall physical difference. Of course, this particular way of combining the estimates of the two factors is arbitrary, since we have no a priori way of knowing which will have the greater influence. However, the 1:4 ratio works reasonably well not only for Experiment 3 here, but for all of the experiments reported by Calder et al. (1996a). For the d' measure, the same formula was used, but with the second factor (sum of identification differences) weighted at 0.01 instead of 0.25, to allow for the different numerical scale of d' as compared to percent correct. Again, the same procedure was used for d' by Calder et al. (1996a).

The three lower graphs in Fig. 6 show subjects' mean predicted performance (filled squares) and their observed performance (open squares) for each measure. The fit between predicted and observed performance in each graph was assessed by correlating predicted and observed scores, and converting this to a t-value (McNemar, 1962). This showed a significant correlation of predicted and observed performance in each case (sequential discrimination, $r = 0.46$, $t = 2.76$, df 28, $p < 0.02$; simultaneous discrimination, percent correct 'different' responses, $r = 0.57$, $t = 3.69$, df 28, $p < 0.001$; simultaneous discrimination, d', $r = 0.53$, $t = 3.29$, df 28, $p < 0.01$). As can be seen, peak discriminabilities were generally found for morphs which were near the midpoint of each section of the continuum, whereas morphs lying closer to the prototype expressions (which were never shown in the experiment) were less discriminable.

An intriguing feature of this finding is that it appears to be the morphed images which subjects were slowest to assign to emotion categories in the identification task used in Experiment 1 which proved easiest to discriminate in Experiment 3. We therefore decided to find out whether the reaction times for identification of each morph from Experiment 1 could also predict discrimination accuracy in Experiment 3. This was done with the same procedure as above, but taking the contribution from categorical perception of the facial expressions to predicted performance to be one-fiftieth of the difference between the Experiment 1 reaction times (in ms) for identification of the morphs in each stimulus pair. This analysis of the relation between reaction time for identification of each morphed image and discrimination accuracy also showed significant correlations (sequential discrimination, $r = 0.40$, $t = 2.33$, df 28, $p < 0.05$; successive discrimination, $r = 0.65$, $t = 4.49$, df 28, $p < 0.001$), confirming that morphs which subjects were slower at assigning to emotion categories in Experiment 1 were indeed easier to discriminate in Experiment 3.

We now turn to consider the reaction times from the discrimination tasks. These

cannot be used as a principal measure, because discrimination accuracies were at chance level for some pairs. Also, in the particular case of ABX discrimination, response times will probably be affected by the order of the stimuli; for example, subjects may be faster to respond when face B = face X than when face A = face X, and it is quite likely that they are doing different things in making these comparisons to the immediate or one-before-last items. However, despite these problems affecting the interpretability of discrimination RTs, we considered it important to establish that there was no overall speed–accuracy tradeoff; i.e., that poor discrimination did not arise from subjects responding very quickly to these pairs, without due consideration. Analyses of the relation between reaction time and discrimination accuracy for Experiment 3 showed highly significant negative correlations (sequential discrimination, $r = -0.73$, $t = 5.58$, df 28, $p < 0.001$; successive discrimination, $r = -0.80$, $t = 7.13$, df 28, $p < 0.001$). In other words, RTs were fastest when discrimination was good, and slowest when discrimination was poor. Hence, there was no sign of subjects trading speed against accuracy.

Although the design used here, in which faces could differ by 1 step around the perimeter of the hexagon shown in Fig. 1b, has certain powerful features, there is one problem which can limit the interpretation made. This is that, whilst the morphing process creates equal shape intervals for any one of the 6 constituent continua in the hexagon, these intervals need not be equal across continua. Instead, the size of interval for each continuum will be dependent on the similarity of the two prototype expressions involved; the more similar the prototypes are in the first place, the smaller will be the intervals between the interpolated images. The consequence of this is that the size of the 1 step intervals which pass a prototype (i.e., the 10,90 discriminations in Fig. 6) need not be exactly the same as the sizes of adjacent intervals.

For these reasons, we decided to confirm the critical categorical perception finding of better discrimination for pairs of stimuli drawn from regions where there were marked differences in identification rates than for pairs of stimuli lying close to the prototypes, using a procedure which did not involve data from the 10,90 pairs. If one removes the 10,90 pairs from consideration, then each set of discrimination results involves 6 continua running from pairs numbered 90,70 to 30,10 using the convention outlined for Fig. 6. Of these, the 90,70 and 30,10 pairs in each continuum lie close to one of the two prototypes, and the 70,50 and 50,30 pairs are more distant from the prototypes. We, therefore, calculated the mean discriminabilities of near prototype (90,70 and 30,10) and more distant from prototype (70,50 and 50,30) pairs in each continuum. If categorical perception applies, then we would expect better discrimination of pairs that are more distant from the prototypes, even though the members of near-prototype and far-from-prototype pairs are an equal number of steps apart in the morphed continua. This was found; mean discriminability was poorer for near-prototype pairs in all 6 continua for each of the lower 3 graphs in Fig. 6 (Wilcoxon matched-pairs signed-ranks test: sequential discrimination, $T = 0$, $N = 6$, $p < 0.05$; simultaneous discrimination, percent correct 'different' responses, $T = 0$, $N = 6$, $p < 0.05$; simultaneous discrimination, d', $T = 0$, $N = 6$, $p < 0.05$).

Finally, it seems likely from these analyses that the sequential and simultaneous discrimination tasks tap the same phenomenon, since results were always consistent across these variants of the discrimination task, even though different groups of subjects were involved. There was a high correlation between observed percent correct performance for sequential and successive discrimination ($r = 0.82$, $t = 7.45$, df 28, $p < 0.001$).

10. Discussion

Experiment 3 showed significant correlations of predicted and observed discrimination performance, and the critical categorical perception finding of better discrimination for pairs of stimuli drawn from regions where there were marked differences in identification rates than for pairs of stimuli lying close to the prototypes. Better discrimination of far-from-prototype stimuli was found even though the members of near-prototype and far-from-prototype pairs were an equal number of steps apart in the morphed continua. These results held both for sequential (ABX) and for simultaneous discrimination, and the findings were achieved using a set of stimuli consisting entirely of morphed faces made by combining prototype images from 6 possible emotion categories.

These findings provide a strong test of categorical perception, and rule out two important sources of potential artifact which are often neglected.

First, the typical categorical perception experiment uses a single continuum ranging between two prototype expressions. A problem with this sort of design is that during the course of the experiment subjects can readily learn the prototypes and the range of stimuli in between; the consequence might be to bias subjects' correct discrimination responses toward the middle of each continuum (Poulton, 1975). Here, we eliminated such problems by drawing stimuli from six possible continua forming the perimeter of an emotion hexagon, and by never showing any of the prototypes from which these continua were derived.

Second, the standard ABX discrimination procedure incorporates a memory load; in order to decide whether face X matches face A or face B, one must hold representations of A and B in memory. Hence, although the ABX discrimination task is usually interpreted as if it were a test of perceptual processing, it also taps short-term memory function. It was possible then that the pattern of discrimination data (heightened discrimination at the category boundary) found with the ABX task might have reflected a short-term memory phenomenon rather than a perceptual effect. Here, we replicated the results of ABX discrimination with a same–different matching task that did not incorporate a memory load as an essential component. Of course, it remains (just) *possible* that people still make use of some form of short-term storage even in the same–different matching task; however, the important point is that same–different matching with simultaneously presented pairs does not in itself *demand* that they do so, and it thus gives a rather better estimate of perceptual discrimination than does the sequential ABX task. Given the high degree of concordance between the data from these two tasks, it

seems reasonable to infer that categorical perception of facial expressions is more likely a perceptual than a short-term memory phenomenon.

We can conclude, then, that categorical perception (in the sense of enhanced discrimination across boundaries and poorer within-category discrimination) is found even when stimuli are drawn simultaneously from a large set of emotion categories.

EXPERIMENT 4

Although Experiments 1–3 produced evidence consistent with a category-based rather than a dimensional account of facial expression recognition, they also showed that people can see differences between morphed expressions, even when they assign them to a common category. This was implied by our findings that distance from the prototype affects ability to classify the resulting morphed expression (Experiment 1 and Experiment 2), and that discrimination need not be at chance level for within-category pairs (90,70 and 30,10 pairs in Experiment 3).

In Experiment 4, we were interested to explore further the nature of this within-category discriminability, and especially to determine whether people can perceive the direction in which an image has been morphed. For example, can they see the anger in a 90% happy 10% angry morph, or in a 70% happy 30% angry morph, even though they would classify both of these as 'happy'? To investigate this, we asked subjects to rank order the emotions which each morph mostly closely approximated.

11. Method

11.1. Stimuli

Stimuli were the 75 morphed images from Experiment 1, and JJ's 6 prototype expressions from the Ekman and Friesen (1976) series.

11.2. Subjects

Eight members of the MRC Applied Psychology Unit subject panel participated in the experiment. All were between the ages 25 and 40 years and had normal or corrected-to-normal vision. None had participated in Experiments 1–3. They were paid for participating.

11.3. Design and procedure

The experiment involved 2 blocks of trials, carried out in a single testing session. In each block, the 75 morphed images from Experiment 1 were presented one at a time in random order. Interspersed among these trials with morphed faces

were 2 sets of trials with JJ's 6 prototype expressions (happiness, surprise, fear, sadness, disgust, anger) from the Ekman and Friesen (1976) series.

Subjects were asked to decide for each face image which emotion (happiness, surprise, fear, sadness, disgust or anger) it was most like, which it was second-most like, and which it was third-most like. Responses were made via sequential button presses for first, second and third choices, using a box with 6 labelled buttons (one for each emotion), which was interfaced with the computer to record ratings and begin the next trial once the third choice had been made. On each trial, the morphed image remained in view until the three button press choices were completed. The ordering of the 6 emotion labels on the response buttons was randomised across subjects.

To ensure that subjects were used to this procedure, the first block of trials was preceded by 12 practice trials with the 90% and 70% morphs from the 6 emotion-to-neutral continua used in Experiment 2; these images form the two left-most columns in Fig. 4.

As for Experiment 1, stimuli were presented on a 256 grey scale 16″ Macintosh colour monitor using Psyscope software. Each face subtended a horizontal visual angle of 3.6° and a vertical visual angle of 5.2°.

12. Results

The issue of interest was whether subjects could see that a morphed image which was close to prototype A (e.g., one which is 90% happy) had been combined with prototype B (e.g., 10% anger). To investigate this, we created an arbitrary scale to indicate which emotions were seen in each morphed image. On this scale, the first choice emotion for each stimulus was given a score of 3, the second choice a score of 2, the third choice scored 1, and the other (unchosen) emotions scored 0. What we were looking for was whether people could see the combinations of expressions *within* the region consistently assigned to a particular emotion category; i.e., whether a morph lying close to prototype A was rated most like expression A, but next-most like B (where B is the expression it was blended with), and unlike C, D, E or F (expressions not used in that particular morph). The critical question then becomes that of whether the mean score for emotion B was higher than the mean scores for emotions C–F.

A complicating factor is that some of the prototype facial expressions are more similar than others; hence, it is important to separate out cases where emotion B is seen in a morph simply because the prototype expression for emotion A resembles emotion B anyway. For this reason, the prototype images were included in Experiment 4, to give an estimate of the intrinsic similarities between JJ's facial expressions regardless of morphing. The scores for the relevant prototype were then subtracted from the scores for each of the morphed images, so that only *differences* between the perception of each emotion in a morphed image and the appropriate prototype were considered.

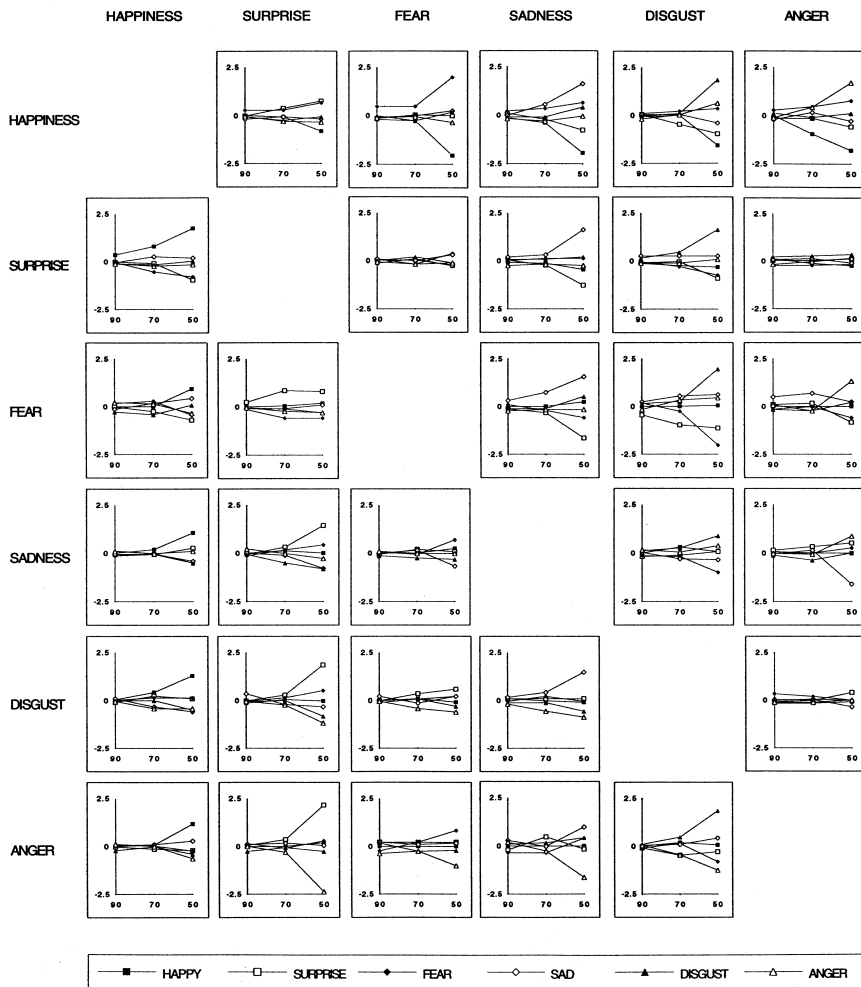These difference scores are presented in Fig. 7. Each of the graphs in Fig. 7

Fig. 7. Scores representing differences between the perception of each emotion in a morphed image and the appropriate prototype. Each graph shows difference scores (*y* axis) for the perception of each emotion in 90%, 70% and 50% morphs (*x* axis). Each of the near prototype expressions (i.e., the ones contributing 90%, 70% or 50% to each morph) is shown as a separate row. Moving from left to right along each row, the graphs show data for morphs moving toward each of the 6 different far prototype expressions (i.e., those contributing 10% to 90% morphs, 30% to 70% morphs, or 50% to 50% morphs).

shows difference scores (*y*-axis) for the perception of each emotion in 90%, 70% and 50% morphs (*x*-axis). Each of the near-prototype expressions (i.e., the ones contributing 90%, 70% or 50% to each morph) is shown as a separate row. Moving from left to right along each row, the graphs show data for morphs moving toward each of the 6 different far-prototype expressions (i.e., those contributing 10% to 90% morphs, 30% to 70% morphs, or 50% to 50% morphs).

Because these scores represent the differences from the relevant near prototype, the score for the perception of the near-prototype emotion quickly becomes negative as the morph moves away from the prototype. What is of interest, however, is what happens to the difference scores for the far-prototype emotions. Inspection of Fig. 7 shows that these are generally indistinguishable from the scores for other emotions at the 90% morph level (where the far prototype contributes 10%), start to rise above the scores for other emotions at the 70% morph level (where the far prototype contributes 30%), and are usually well clear of the rest at the 50% point.

To analyze this statistically, we compared the difference scores for far prototypes to the mean difference score for the 4 unrelated emotions which were not involved in that continuum (i.e., excluding the near and far prototypes). A two-factor Analysis of Variance was used to determine the effects of distance from the prototype (90%, 70%, 50% morphs; repeated measure) and type of emotion (far prototype versus unrelated; repeated measure). This showed significant main effects of both factors (distance from prototype, $F = 104.77$, df 2,58, $p < 0.001$; type of emotion, $F = 87.63$, df 1,29, $p < 0.001$). Both of these main effects were qualified by the significant interaction shown in Fig. 8 (distance from prototype X type of emotion, $F = 111.64$, df 2,58, $p < 0.001$). Analysis of this interaction with post hoc Tukey tests ($\alpha = 0.05$) showed that perception of the far-prototype emotion was significantly different from the perception of unrelated emotions at the 50% level and at the 70% level, but not at the 90% level (where the far prototype only contributes 10% to each morphed image).

Note that in Fig. 8, where any overall similarity between prototype expressions has already been removed by subtraction, the mean score for perception of unrelated expressions is effectively zero at all levels of morphing. This is exactly as would be expected if the emotions which do not come from one of the prototypes are not perceived in the morphed images. The finding that this holds for
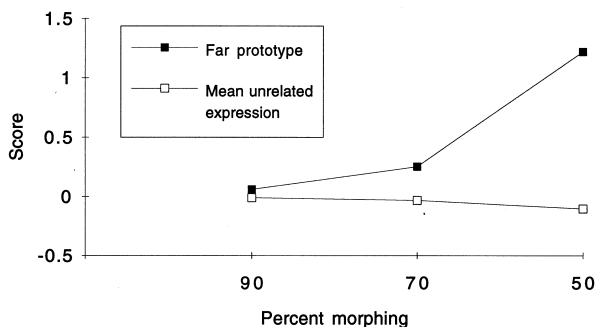

Fig. 8. Interaction of distance from prototype with type of emotion; the scores represent differences between the perception of an emotion in a morphed image and the appropriate prototype. Ability to perceive the far prototype emotion was significantly different from the perception of unrelated emotions (i.e., the 4 emotions which were not used in the creation of a particular morph) at the 50% level and at the 70% level, but not at the 90% level.

50% morphs is consistent with the identification results from Experiments 1 and 2, where we noted that there were few intrusions from unrelated categories at the mid-point of each continuum.

13. Discussion

The results of Experiment 4 show that when a prototype only contributes 10% to a particular morph, subjects cannot reliably identify which prototype that 10% has come from, but that for prototypes contributing 30% they begin to be able to do this. This is interesting because in Experiment 1 all 30% morphs were consistently identified as belonging to the category contributing 70%. However, we noted that distance from the prototype affected reaction times for classifying morphed expressions in Experiment 1, which implied that people can see differences between these morphed expressions, even when they assign them to a common category. Experiment 4 has taken this point a step further, by showing not only that people are sensitive to differences between expressions which they assign to a common overall category, but that they can have some insight into the nature of these differences.

However, one should not overestimate the extent of this insight, which was elicited by a forced-choice procedure, where subjects were made to select 3 labels for each face. Etcoff and Magee (1992) had included a 'free naming' task in which subjects were shown each picture and encouraged to describe the expression using whatever words or labels they wished. They noted that in these circumstances ''descriptions in which two emotions were mentioned were virtually non-existent, even though use of multiple labels of the subjects' choosing was encouraged'' (Etcoff and Magee, 1992, p. 236).

Nonetheless, the ability to see some within-category differences revealed by our experiments makes sense, given that in everyday life we are often interested not only in the emotion *category* to which a perceived facial expression belongs, but the *intensity* with which it is expressed. Moreover, more subtle emotion states may involve transitions from one emotion to another, or even blends, and again some ability to pick these up perceptually is desirable. Consistent with our findings, Ekman et al. noted that for faces which had some action units relevant to a second emotion, subjects could often identify the second emotion, though with a certain amount of difficulty (Ekman et al., 1987).

GENERAL DISCUSSION

Our experiments examined perception and recognition of facial expressions, using morphs of the full set of emotions from the Ekman and Friesen (1976) series. Although the very substantial computational requirements of rendering and morphing such a large set of continua meant that only one person's face (JJ's) was used to create the morphed images, his prototype expressions were well recognised in the control data of Ekman and Friesen (1976), and the results proved highly

consistent across nearly all of the morphed continua. There is thus no reason to believe that our findings are constrained by the use of a single face, and we would expect they will generalise to other models.

In Experiment 1 and Experiment 2 we showed that, for identification, morphed facial expressions still behaved like discrete categories when stimuli could come from any possible continuum and a 6-way choice of all emotions or a 7-way choice including neutral was available. In all cases, identifications of the morphed faces belonging to each continuum were mainly in terms of the prototypes at either end of the relevant continuum, with high identification rates at each end, and a sudden shift in the middle. Moreover, in the regions where identification changed from one emotion category to another, there were few intrusions from categories which did not belong to that continuum. Even in Experiment 4, where subjects were asked to list in order the 3 most appropriate emotion labels for each morph, they showed no tendency to perceive emotions which did not come from one of the prototypes used to generate each morphed image. Hence, these regions with identification shifts were seen as involving orderly transitions between categories, not as regions of generalised uncertainty. These findings held even when 'neutral' was allowed as a response, in Experiment 2; in effect, neutral behaved in the same way as the other emotion categories.

These results provide strong evidence that facial expressions are perceived as belonging to discrete categories. Note especially that the findings held for all expression continua, regardless of their positioning with respect to confusabilities or hypothesised dimensions of emotion. This is inconsistent with any account of emotion perception in terms of a two-dimensional model. On such accounts, continua which lie close to or actually run along the hypothesised dimensions should move through a region with no identifiable expression in the middle; in this region subjects should respond at random or perceive the emotion as neutral, yet this was never observed. Since we explored *every* possible pairwise continuum in Experiments 1 and 2, we might reasonably expect that some should have corresponded to underlying dimensions of emotion perception, if these exist. No such evidence was forthcoming. Similarly, on an account based on two underlying dimensions, continua which cross these dimensions should have a discrete region in the middle which belongs to neither prototype. For example, on the schema of Woodworth and Schlosberg (1954) (see Fig. 1a), the surprise–anger continuum should run through a central region perceived as fear. Again, this type of pattern was never found.

It is clear, then, that the colour circle analogy used by Woodworth and Schlosberg (1954) gives a misleading account of emotion recognition. Although the types of representation used in Fig. 1(a and b) may capture certain interesting and important aspects of the similarities (and hence confusabilities) between emotions perceived in the face, there is no acceptable two-dimensional solution to the whole problem of classifying facial expressions. The two-dimensional circle representation works for colour because hue is determined by a single underlying physical variable (wavelength), and coded in the visual system by opponent processes operating along two dimensions (red–green; blue–yellow). The analogy

of facial expression perception to the colour circle probably breaks down because facial expressions arise from the interaction of several distinct physical cues.

Of course, we may well be able to perceive continuous variation in each of these physical cues to emotion; if you look at each continuum in Fig. 2, you can see the shape of the lips change, the mouth open or close, the eyes widen or narrow, the eyebrows raise or lower, and so on. Our point is simply that the visual system uses these continuously variable cues in combination, to create perceived categories. There is nothing in our results to encourage the view that the visual system codes facial expressions solely in terms of their positions with respect to a small number of underlying dimensions.

Hence, the morphed facial expressions from the Ekman and Friesen (1976) series are best considered as lying in a multi-dimensional perceptual space, in which the continua linking any particular prototype to any of the other prototypes pass through entirely distinct regions. Dimensions such as pleasant–unpleasant thus correspond to intellectual, not perceptual constructs.

To understand the implications of this, two closely related theoretical issues need to be distinguished. The first issue concerns whether facial expressions of emotion are analyzed as discrete categories, or in terms of underlying dimensions (Ekman, 1982; Ekman et al., 1972). As we have noted, our results are in line with the analysis as categories hypothesis, since morphed facial expressions were always seen as belonging to discrete categories, with no evidence of any underlying dimensions. This consistency is all the more impressive because the prototype images were never shown in Experiment 1 or Experiment 2; all of the stimuli were morphs.

The second issue concerns the level at which assignment to categories is made; whether (as we have argued here) perceptual classification mechanisms are specifically tuned to the combinations of facial features that represent each particular emotion, and we assign a seen facial expression perceptually to the category exemplifying the nearest emotion, or whether category membership is assigned at a post-perceptual stage by higher conceptual and linguistic systems (Etcoff and Magee, 1992). As Etcoff and Magee (1992) pointed out, the hypothesis that emotions are analyzed as discrete *perceptual* categories also predicts that facial expressions will show categorical perception effects. Experiment 3 therefore used discrimination tasks, in which the pairs to be compared were taken from the morphs in continua belonging to the perimeter of the hexagon drawn in Fig. 1b. Results showed categorical perception in terms of better discrimination for pairs of stimuli taken from regions belonging to different perceptual categories, and poorer discrimination of pairs of stimuli lying close to the prototypes. This was achieved using a set of stimuli consisting entirely of morphed faces, which were made by combining prototype images from 6 possible emotion categories. We can thus conclude that categorical perception (in the sense of enhanced discrimination across boundaries and poorer within-category discrimination) is found even when stimuli are drawn simultaneously from a large set of emotion categories.

When placed end to end, the 6 continua used in Experiment 3 had no fixed

beginning or end. Experiment 3 was thus an extension of the technique used by Calder et al. (1996a), who used a fear–happiness–anger–fear triangle to show that categorical perception of facial expressions is not created by range effects, anchor effects, or learning of prototypes during the experiment (in fact, prototypes were never shown). The present Experiment 3 therefore provides further strong support for the view that categorical perception of facial expressions reflects what is in subjects' heads at the start of the experiment, rather than artefacts created by reliance on single continua. In addition, Experiment 3 used a simultaneous discrimination task that did not incorporate a memory load as an essential component as well as the more conventional sequential (ABX) discrimination; closely comparable results were found with these tasks. In this way, we were able to show that categorical perception of facial expressions is more likely a perceptual than a short-term memory phenomenon.

Despite these clear demonstrations that facial expression recognition shows properties consistent with the idea that recognition involves assigning expressions to discrete categories, reaction time data from the identification tasks used in Experiment 1 and Experiment 2 showed that distance from the prototype had a definite cost on ability to classify the resulting morphed image; morphs which lay closest to the prototype expressions used to create them were most easily identified. This finding provides evidence against the view that categorical perception effects are an epiphenomenon of verbal labelling. Instead, the best-discriminated pairs were those which were actually the most difficult to label. However, the finding also shows that people are sensitive to differences between particular exemplars of facial expressions and some idealised category prototype.

Experiment 4 took this a step further, and showed not only that people are sensitive to differences between expressions which they assign to a common overall category, but that they can have some insight into the nature of these differences. Hence, the categorisation of facial expressions into basic emotions is carried out by a system which remains sensitive to physical differences within a category; for instance, the differences between two different faces which may signal fear to an observer may still be seen, even though they are assigned to a common category. But there are limits to our ability to see within-category differences; Experiment 3 showed the classical categorical perception effect of better discrimination for pairs of stimuli drawn from regions belonging to different perceptual categories, and poorer discrimination of pairs of stimuli lying close to the prototypes.

We have shown, then, that categorical perception effects of enhanced discrimination at category boundaries are found even though subjects can be sensitive to differences between morphed expressions in identification tasks. We therefore need to consider which forms of category model can provide plausible accounts of the data we have presented. Our view is that an account treating facial expressions as corresponding to discrete categories of emotion fits our findings, but we accept they are inconsistent with the strongest possible variants of the categorical perception hypothesis; it is not the case that assignment to discrete categories

means that there is nothing else to account for. In fact, though, findings from other domains of research are also inconsistent with exaggerated and oversimplified versions of categorical perception; the overall pattern of findings reported here for facial expressions is no different than for other perceptual categories.

Consider colour perception. Although we can discriminate a wavelength we see as blue from a wavelength we see as green more easily than we can discriminate an equivalent difference in wavelength falling within the blue or within the green category, we can nonetheless still see differences between different blues and different greens (Bornstein and Korda, 1984). Similarly, studies of categorical perception effects for speech show that stimuli falling within the same phonetic category can be discriminable from each other, even though within-category discriminability is generally lower than between-category discriminability (Miller, 1994). Such findings exactly parallel our results for perception of facial expressions of emotion.

Our experiments show that, although facial expressions are recognised as discrete categories, there are clear effects on recognition of distance from the prototype; the closer a seen face approximates one of the category prototypes, the more easily it is recognised. But this is also true for all known forms of human categorisation. In perceptual categorisation, for example, studies of speech recognition show that it is possible to construct good or poor exemplars of each phoneme (Miller, 1994). Or consider semantic categorisation. Again, closeness to category prototypes is well-known to exert influences on our ability to assign exemplars to the correct semantic category; we can classify robins as belonging to the category 'birds' more quickly than we can classify atypical exemplars such as penguins (Shoben, 1982; Young et al., 1989).

Facial expressions, then, show effects which are found for other types of human classificatory ability. This is in some ways counterintuitive. Because of the strong biological background evident in facial expressions of emotion (Darwin, 1872; Ekman, 1973), it might have been thought that the evidence of categorical perception we have found reflected a relatively fixed influence from a dedicated underlying biology. Although this assumption is often made, in part because other categorical perception effects (e.g., for speech or colour perception) also seem to be linked to underlying biological predispositions, we see no compelling reason to accept it. Instead, it seems to us more likely that categorical perception will prove to be an emergent property of population coding in the nervous system; in other words, we suspect that categorical perception effects of the type investigated here will arise whenever populations of cells become tuned to distinct categories. Findings that categorical perception effects can be created by learning, such as those observed for recognition of identity from the face (Beale and Keil, 1995), are consistent with this view.

It seems, then, that rapid classification of prototypes and better across-boundary discriminability reflect the underlying organisation of human categorisation abilities. If so, then Harnad (1987b) was right to suggest that categorical perception reflects fundamental properties of human cognition.

Acknowledgments

References

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

Benson, P.J. and Perrett, D.I. (1991). Synthesising continuous-tone caricatures. *Image and Vision Computing*, *9*, 123–129.

Bornstein, M.H. and Korda, N.O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research*, *46*, 207–222.

Burt, D.M. and Perrett, D.I. (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society, London*, *B259*, 137–143.

Calder, A.J., Young, A.W., Perrett, D.I., Etcoff, N.L. and Rowland, D. (1996a). Categorical perception of morphed facial expressions. *Visual Cognition*, *3*, 81–117.

Calder, A.J., Young, A.W., Rowland, D., Perrett, D.I., Hodges, J.R. and Etcoff, N.L. (1996b). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology*, *13*, 699–745.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: John Murray.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, *3*, 1–8.

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

Ekman, P. (Ed.). (1973). *Darwin and facial expression: A century of research in review*. New York: Academic Press.

Ekman, P. (Ed.). (1982). *Emotion in the human face*. Cambridge: Cambridge University Press.

Ekman, P. (1992). Facial expressions of emotion: An old controversy and new findings. *Philosophical Transactions of the Royal Society, London*, *B335*, 63–69.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, *115*, 268–287.

Ekman, P. and Friesen, W.V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W.V. and Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. New York: Pergamon.

Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., Lecompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., Scherer, K.R., Tomita, M. and Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*, 712–717.

Etcoff, N.L. (1984). Selective attention to facial identity and facial emotion. *Neuropsychologia*, *22*, 281–295.

Etcoff, N.L. and Magee, J.J. (1992). Categorical perception of facial expressions. *Cognition*, *44*, 227–240.

Field, T.M., Woodson, R., Greenberg, R. and Cohen, D. (1982). Discrimination and imitation of facial expressions by neonates. *Science*, *218*, 179–181.

Harnad, S. (Ed.). (1987a). *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

Harnad, S. (1987b). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1–25). Cambridge: Cambridge University Press.

Hasselmo, M.E., Rolls, E.T. and Baylis, G.C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*, 203–218.

Johnson, M.H., Dziurawiec, S., Ellis, H. and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*, 1–19.

Liberman, A.M., Harris, K.S., Hoffman, H.S. and Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

McNemar, Q. (1962). *Psychological statistics* (Third edition). New York: Wiley.

Miller, J.L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, *50*, 271–285.

Morishima, S. and Harashima, H. (1993). Emotion space for analysis and synthesis of facial expression. *IEEE International Workshop on Robot and Human Communication*, 188–193.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J. and Dolan, R.J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*, 812–815.

Poulton, E.C. (1975). Range effects in experiments on people. *American Journal of Psychology*, *88*, 3–32.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Sergent, J., Ohta, S., MacDonald, B. and Zuck, E. (1994). Segregated processing of facial identity and emotion in the human brain: A PET study. *Visual Cognition*, *1*, 349–369.

Shoben, E.J. (1982). Semantic and lexical decisions. In C.R. Puff, (Ed.), *Handbook of research methods in human memory and cognition* (pp. 287–314). New York: Academic Press.

Sprengelmeyer, R., Young, A.W., Calder, A.J., Karnat, A., Lange, H.W., Hömberg, V., Perrett, D.I. and Rowland, D. (1996). Loss of disgust: Perception of faces and emotions in Huntington's disease. *Brain*, *119*, 1647–1665.

Woodworth, R.S. and Schlosberg, H. (1954). *Experimental psychology: Revised edition*. New York: Henry Holt.

Young, A.W., Newcombe, F., de Haan, E.H.F., Small, M. and Hay, D.C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain*, *116*, 941–959.

Young, A.W., Newcombe, F., Hellawell, D. and de Haan, E.H.F. (1989). Implicit access to semantic information. *Brain and Cognition*, *11*, 186–209.