COGNITION AND EMOTION, 2001, *15* (5), 633–658

Categorical perception of facial expressions of emotion: Evidence from multidimensional scaling

David Bimler and John Kirkland

Massey University, New Zealand

Recent studies have shown that the perception of facial expressions of emotion fits the criteria of categorical perception (CP). The present paper tests whether a pattern of categories emerges when facial expressions are examined within the framework of multidimensional scaling. Blends of five ‘‘pure’’ expressions (*Angry*, *Sad*, *Surprised*, *Happy*, *Neutral*) were created using computerised ‘‘morphing’’, providing the stimuli for four experiments. Instead of attempting to identify these stimuli, subjects described the proximities between them, using two quite different forms of data: similarity comparisons, and sorting partitions. Multidimensional scaling techniques were applied to integrate the resulting ordinal-level data into models which represent the interstimulus similarities at ratio level. All four experiments yielded strong evidence that the expressions were perceived in distinct categories. Adjacent pairs in the models were not spaced at equal intervals, but were clustered together as if drawn towards a ‘‘perceptual magnet’’ within each category. We argue that spatial representations are compatible with CP effects, and indeed are a useful tool for investigating them.

One attempt to systematise the diversity of human facial expressivity centres on *prototype* expressions corresponding to a small number of discrete basic-emotion categories. These prototypes are configurations of facial-feature displacements, produced by well-defined combinations of muscle contractions. Tomkins (1962) argued that certain facial behaviours have evolved to meet different survival needs and warrant the status of ‘‘primary affects’’. Evidence has since accumulated that these configurations maximise the rate of correct responses in forced-choice identification (Izard, 1971; Ekman, Friesen, &

Correspondence should be addressed to John Kirkland, PhD, Department of Health and Human Development, Massey University, Private Bag 11-222 Palmerston North, New Zealand.

We are especially grateful to Professor P. Ekman who granted permission for use of facial expression photographs, and who generously performed the FACS analyses described in the text. Robyn Surville and Sheryn Lane performed the arduous task of morphing the stimuli. We are grateful for the cooperation of the local high schools which allowed their students to participate in Experiment 4, and of course to the participants themselves. We would also like to thank David Matsumoto for his very helpful comments on previous versions of this article.

http://www.tandf.co.uk/journals/pp/02699931.html DOI:10.1080/02699930143000077

Ellsworth, 1972). They convey similar emotions in a range of cultures, supporting the thesis that they are innate (Ekman, 1994). These widely replicated findings suggest a categorical structure to the domain of facial expressions (FEs). Neurological conditions that selectively disrupt recognition of different FEs suggest that the neural pathways subserving FE recognition are not the same for all emotion categories (Adolphs, Tranel, Damasio, & Damasio, 1994), so one might speak of detection mechanisms ''tuned'' to each prototype.

There is also evidence for an alternative approach, in which the affective content of FEs are analysed into continuously varying attributes that can be treated as the dimensions of a spatial model (e.g., Nummenmaa, 1992; Russell & Bullock, 1986). Woodworth (1938) located expression categories along a continuum to reflect the systematic nature of the misidentifications among them. The techniques of factor analysis and multidimensional scaling (MDS) are often used to reconstruct such models from the pattern of similarities among FE stimuli, with misidentifications (Osgood, 1966), dissimilarity judgements (Paramey, Schneider, Josephs, & Slusarek, 1994), or triadic ''odd-one-out'' judgements (Alvarado, 1996; Gladstones, 1962) providing the information. This dimensional account is compatible with ''relativity effects'', in which the way a stimulus is perceived changes when viewed in the context of expressions from different categories (Russell & Fehr, 1987). Such effects can also be accommodated within the categorical framework as adaptation or fatigue of specific detection mechanisms (Prkachin & Prkachin, 1994). In another context effect, significant changes to the emotion seen in a FE are brought about by a brief description of what the poser has just experienced (Carroll & Russell, 1996). This is hard to reconcile with a strongly categorical model in which stimuli are assigned to categories at an early stage of perception.

Much of the evidence for the categorical perspective was acquired with forced-choice identification, in which subjects use a restricted list of emotion terms (for instance, *Anger*, *Sadness*, *Surprise*, *Happiness*, *Fear*, *Disgust*, and perhaps *Contempt*) to identify FEs. The procedure is open to methodological criticism (Russell, 1994). For instance, the apparent universality of the categories across cultures may merely demonstrate that equivalents for the emotion terms can be found in different languages, and that informants concur in the use of those equivalents; it does not show them to be universally ''basic''. The procedure has been defended (Ekman, 1994) and extended (Haidt & Keltner, 1999). However, it would be helpful to test the claims of the categorical perspective using the MDS technique more familiar within the dimensional tradition, and that is the approach taken in the research reported here.

The debate reflects a parallel disagreement whether human affective states are better characterised in terms of certain irreducible ''basic emotions'', or a smaller number of affective dimensions. For our purposes it is enough to note the striking resemblance between MDS models of expressions, and the analogous maps of ''emotion space'' derived from the semantic similarities among emotion terms (Roberts & Wedell, 1994) and from the correlations within mood self-reports (Feldman, 1995). One way to account for this correspondence is to regard FEs as a communication channel, encoding information about the signaller's emotional state (Osgood, 1966). Much of the interest in FEs is as ''the mind's construction in the face'': they are a way to study by proxy the domain of emotions, the latter being less tractable to examination, for it is hard to present observers with standardised emotion stimuli. Any evidence as to whether the differences among expression categories are qualitative or quantitative carries a corollary that the situation among emotions is similar.

There is no doubt that we categorise FEs; the moot point is whether assignment to categories is a feature of perception, or of conceptual processing. A natural extension of the categorical perspective is to search for categorical perception (CP) effects in the perception of FEs. The normal definitions of categorical perception entail two criteria: ''A CP effect occurs when (1) a set of stimuli ranging along a physical continuum is given one label on one side of a category boundary and another label on the other side and (2) the subject can discriminate smaller physical differences between pairs of stimuli that straddle that boundary than between pairs which are within one category or the other'' (Harnad, 1987). As a corollary, a pair of stimuli that lie near the middle of a category is harder to discriminate than another pair (objectively differing by the same amount) near the boundary. In particular, imperfect instances of a perceptual category are perceived as better than they are, because of their similarity to a good instance. One can imagine a ''focus'' in each category—a prototypal, best-possible stimulus; less-good stimuli are displaced in its direction in the perceptual space. This is the concept of a ''perceptual magnet'' (Iverson & Kuhl, 1995), which differs from the CP formulation in placing more emphasis on this reduced within-category discrimination, and less on the boundaries between categories.

The classic examples of categorical perception involve auditory domains: musical intervals, and phonemic stimuli such as vowels and consonants (Harnad, 1987). The feature they have in common is a continuous perceptual ''space'' which listeners perceive as a discontinuous series of categories. In musical intervals and stop consonants, each stimulus is described by a single parameter—the ratio between two tones, and the voice-onset time, respectively—defining a one-dimensional space, in which better or worse category members are easily synthesised. In particular, one can create within-category and cross-category pairs, separated by the same interval.

The techniques of computerised ''morphing'' allow researchers the same fine control over facial expressions of emotion. Any two photographs of different ''pure'' facial expressions (i.e., ones that viewers consistently identify as representing a specific emotion) define a continuum: one can interpolate between them, creating a sequence of intermediate steps, each one differing by the same amount from the adjacent step. If CP applies to facial expressions, then a blended stimulus (e.g., 30% of one expression plus 70% of the other) will be primarily perceived as one or the other extreme. The dominant category will obscure the contribution of the minor component, reducing the perceived differences between adjacent steps at the ends of the sequence as the stimuli bunch up towards the endpoints, and increasing the differences across the category boundary.

To test the above criteria of CP, two types of tasks are typically applied: identifying each stimulus with an appropriate label, and making Same/Different or forced-choice ABX judgements about pairs of items to measure their discriminability. Some studies have carried over from auditory precedents the procedure of presenting in sequence the stimuli to be discriminated, or judged as ''same'' or ''different''. This leaves open the possibility that the appearance of categories is a feature of short-term memory load rather than of perception: Sequential presentation strengthens CP, by forcing subjects to encode stimuli before comparing them, losing potential intra-category distinctions in the process. Thus, Iverson and Kuhl (1995) found that CP effects for vowels increased when longer intervals separated the stimuli. Parallel presentation allows stronger conclusions to be drawn.

Etcoff and Magee (1992) applied this approach to line drawings of expressions. Eight continua were used, each continuum comprising 11 stimuli (two pure expressions plus blends interpolated at 10% intervals). Labelling and ABX tasks provided evidence for greater discrimination across category boundaries. Subsequent studies have used photographic-quality computer-manipulated (morphed) images for stimuli, and measured reaction times as well as accuracy for discrimination of each pair. Calder, Young, Perrett, Etcoff, and Rowland (1996) replicated those results with six continua. In their Experiment 4, Calder et al. used a same/different judgement task in which pairs of stimuli were presented simultaneously. De Gelder, Teunisse, and Benson (1997) took additional precautions to exclude non-CP explanations. For children as well as adult observers, stimuli were most easily discriminated around category boundaries (i.e., the point on each continuum where identification reached 50%). Finally, Young et al. (1997) collected data for an exhaustive set of 21 continua.

In a departure from those meticulous studies, the present study tests for CP effects by importing quite different procedures for data collection and analysis, from a different research tradition. Morphed facial expressions were examined using MDS to analyse the relative similarities of different pairs of stimuli.[1] MDS provides a spatial model of the data in which each stimulus is represented by a

---

[1] For an earlier application of MDS to impure, blended expressions, see Nummenmaa (1992). That study tested whether more than one emotion could be expressed simultaneously, and which combinations are mutually compatible. It is not clear whether the actors' expressions could be described as interpolations between extremes (in the sense that an increase in the contribution of expression A was accompanied by a decrease in the intensity of expression B).

point $\mathbf{x}_i$. If CP effects contribute to the perception of facial affect, then the subjective distances reconstructed between morphed expressions in the MDS solution will differ from the objective differences used in creating them, with the $\mathbf{x}_i$ corresponding to the i-th morphed expression displaced towards the focus of the category it belongs to; this would manifest as a clumping of points around the prototypes.[2]

In four experiments, we collected data about the dissimilarities perceived among expressions. If the values of δ_{ij} (the perceived dissimilarities) were measured to sufficient accuracy, they could be compared directly with the underlying objective differences. But in these experiments the observations were: (a) ordinal-level; (b) incomplete (not every comparison between one dissimilarity and another was made); and (c) limited in reliability by the discernment of the observers. A triadic procedure was followed in Experiment 1, and a ‘‘quartet’’ procedure in Experiment 2, using two separate sets of stimuli, 31 in each. In Experiment 3, data were collected in a different form: a sorting procedure was applied to both sets of stimuli. Experiment 4 extended the sorting procedure by applying it to the two sets together.

The role of MDS is to integrate these imperfect data. Each d_{ij} (we write d_{ij} for the reconstructed distance between the i-th and j-th stimulus points) is a ratio-level estimate of the subjective δ_{ij} for those stimuli, determined by *all* the proximity data, not just the judgements for that particular pair of i, j. MDS rules out combinations of d_{ij} which are consistent with the data but cannot be embedded within a space of the specified dimensionality. As well as spatial models, the data were modelled with hierarchical ‘‘trees’’. A tree representation of similarity data consists of nodes corresponding to the stimuli E_i (these comprising the ‘‘leaves’’ of the tree), and a branching structure of links. The distance d_{ij} is defined as the total length of the links one traverses in moving between the nodes for E_i and E_j. A tree model is evidence for CP if the nodes are arranged in a structure of distinct branches, corresponding to the categories.

If CP effects do emerge within the multidimensional framework, this is strong evidence that categories play a key rôle at some stage in the processing of FE stimuli. It is worth emphasising that the existence of categories or perceptual magnets is fully compatible with an analysis of expressions in dimensional terms. An analogy is often drawn with the domain of colour perception (Etcoff & Magee, 1992; Young et al., 1997), and here the evidence for CP effects (Davies & Corbett, 1997; Kay & Kempton, 1984) and the existence across cultures of consistent ‘‘focal colours’’ supplement rather than supplant the

---

[2] The category foci do not necessarily coincide with the ‘‘pure’’ expressions used here as prototypes. It may be that a ‘‘pure’’ stimulus can only approximate the focus of its category with the latter lying outside the range which actors can express. Calder, Young, Rowland, and Perrett (1997) found that computer-exaggerated exemplars of expressions were recognised more rapidly, and rated as more intense, than the unexaggerated images.

widely accepted opponent-process theory (featuring a red/green and a blue/yellow dimension).

Morphed stimuli offer the potential to improve models of ‘‘expression space’’, and a secondary purpose of this research is to exploit that potential. When observers compare expression prototypes (Paramey et al., 1994), the accuracy of their judgements is arguably limited by the size of the dissimilarities involved. Here, they judge the small dissimilarities between adjacent morphs as well. Note also that these stimuli can rule out models of insufficient dimensionality, for two-dimensional spatial models make the specific, falsifiable prediction (Young et al., 1997) that at least one continuum exists, such that particular morphs along it fall into some third category, rather than the categories of either of the end-point ‘‘pure’’ FEs.

EXPERIMENT 1: F-SERIES. TRIADIC DATA

Subjects

Thirteen subjects were recruited informally (eight women, five men). All were aged 30 to 45.

Stimuli

Five ‘‘pure’’ expressions were selected: items e57, e58, e61, e63, e65 from Ekman and Friesen (1976). These expressions exemplify the emotional states *Happy*, *Surprised*, *Angry*, *Sad*—providing two each with a positive and a negative valence—and *Neutral* (we are in the process of adding *Afraid* and *Disgusted* to this assortment). The five exemplars are portrayed by the same woman. In addition, the items are similar in terms of lighting, angle, etc., leaving the expressions themselves (in the form of displacements of easily located facial landmarks) as the only way they differ.

These pure stimuli can be paired in ten ways. Equally spaced sequences, comprising 26 morphs in total, were interpolated between these pairs of end-points. Using E_m to indicate a morphed stimulus, composed from pure stimuli E_a and E_b, we can write

$$E_m = p_m \ E_a + (1 - p_m) \ E_b. \tag{1}$$

Labels were given to each E_m, indicating the respective E_a, E_b, $(1 - p_m)$ (see Table 1). Figure 1 illustrates six of these stimuli: the pure expressions for *Sad* and *Surprised*, and four images interpolated between them. An independent description of the images (P. Ekman, personal communication) using the FACS system of Facial Action Units is as follows:

Figure 1. Example of the continua used in these experiments: *Sad* expression (L), *Surprised* (R), and four images interpolated at intervals of 20%.

Sa	1b+15a	sad
SaSu20	1b+5a	sad/surprised
SaSu40	1a+5a+25	surprised/sad
SaSu60	1+2+5a+26	surprised
SaSu80	1+2+5b+26	surprised, stronger than previous
Su	1+2+5c+26	surprised, even stronger than previous.

The pure and morphed images together, 31 stimuli in total, comprise the F-series. The same procedure was followed with a second set of five exemplars from Ekman and Friesen, portrayed by a man (items e105, e101, e103, e110, e107). This yielded the M-series of images, which have the same values for E_a, E_b, p_m as the F-series.

The morphing procedure described by Calder et al. (1996) was followed as a guideline. Morphing was undertaken by first scanning to digitise monochrome slide images and subsequently applying suitable image-interpolation software (‘‘Design Studio’’), applying as many corresponding points as necessary to make smooth transitions. Each printed image measured 5 × 8 cm.

Table 1 shows that the number of intermediate blends varies between sequences. Only a single blend was interpolated for three pairs of pure expressions: *Happy-Neutral*, *Surprised-Happy*, *Angry-Surprised*. When the expression prototypes are arranged according to the circumplex model (Russell, Lewicka, & Niit, 1989), as a first approximation to an ‘‘expression space’’, these pairs are immediate neighbours around a circle. They are known to be relatively similar, in the sense that an example of one member of a pair is likely to be misidentified as the other. These three sequences are unhelpful for testing CP, although the single intermediate helps triangulate the MDS solution and contributes to the reliability of the measured dissimilarities for the other sequences. More intermediates were created for pairs of pure expressions that lie further apart in the circumplex: The *Sad-Happy*, *Sad-Surprise*, and *Angry-Neutral* sequences each contain four.

TABLE 1
Codes for five pure expressions and 26 morphs, showing number of morphs in each sequence (above diagonal) and descriptive codes for each morph (below diagonal)

	A	Sa	Su	H	N
Angry (A)	–	2	1	3	4
Sad (Sa)	ASa33, ASa66	–	4	4	3
Surprised (Su)	ASu50	SaSu20, SaSu40, SaSu60, SaSu80	–	1	3
Happy (Ha)	AH25, AH50, AH75	SaH20, SaH40, SaH60, SaH80	SuH50	–	1
Neutral (N)	AN20, AN40, AN60, AN80	SaN25, SaN50, SaN75	SuN25, SuN50, SuN75	HN50	–

Method

Subjects used a form of the triadic method (previously applied to facial expressions by Alvarado, 1996; Gladstones, 1962). Subjects viewed items in groups of three and indicated which of each three was the ''odd-one-out'': ''For each set, choose the photograph which is *least similar* to the other two photos''. The items were described as ''photographs'' throughout, to disarm any suspicions as to their artificial nature. The wording of the instructions was designed to encourage subjects to reach a judgement on the basis of underlying emotion: '' 'Similarity' in this case has the sense of 'How similar are the *emotions* expressed in the photos?' or 'How similar are the person's feelings?' ''

Triads were selected randomly. The subject created and judged 10 triads at a time, by shuffling the items and dealing them out into piles of 3 (leaving a single card, to be held over until the next shuffle). This procedure was repeated 12 times so each subject made 120 judgements. Ignoring duplicated triads, the data thus consist of 13 × 120 triadic comparisons, out of the 31.30.29 / 6 = 4995 possible combinations of 3 out of 31 items. A series of pilot studies revealed no difference between random selection of triads, versus a Balanced Incomplete Design, this being the usual procedure when a large number of items generates too many triads for each subject to view exhaustively.

Subjects found the task fatiguing. Some reported that dissimilarities soon became difficult to assess, so that their odd-one-out judgements became a process of assigning emotion labels to each item and choosing the cognitively least-similar emotion. They were encouraged to take as many breaks as they required for unstereotyped responses. The self-randomising procedure allowed subjects to take the items home where they could respond to the triads at their own pace. Some took weeks before returning the completed data-entry forms, while others performed the task in a single session (taking about 45 minutes).

In addition, the same subjects went through a ‘‘pairing-up’’ procedure. The object here was to provide similarity information by arranging the items into most-similar pairs.[3] To achieve the most-similar pairs, the subject shuffled the items and dealt them out into two rows of 15, one row above the other (again, with a single item left over). Then the subject matched each item in the top row with the one most similar to it in the bottom row, by rearranging items within the rows (and by exchanging them between rows if this improved the matching).

Analysis. The triadic judgements can be regarded as relationships between similarities: Choosing E_i as the odd-one-out of items E_i, E_j, E_k is equivalent to asserting that E_j and E_k are the most similar pair, that is, the subject informs us that $\delta_{ij} > \delta_{jk}$, $\delta_{ik} > \delta_{jk}$ (where δ_{ij} is the perceived dissimilarity between E_i and E_j). To recover the actual dissimilarities, we subjected these ordinal-level data to non-metric multidimensional scaling.

The MDS program used is an implementation of a Maximum Likelihood algorithm (ML), very similar to MAXSCAL (Takane, 1978). It calculates the likelihood that those particular comparisons between similarities, applied to a given configuration, would result in the greater-than/less-than judgements which were in fact observed. The essence of the ML approach is to iteratively adjust the configuration until this likelihood reaches a maximum.

Subjects reported that some triads were problematical: The dissimilarities were all of comparable magnitude, so none of the expressions stood out as least similar. We interpret such reports as ‘‘equidistance’’ judgements—$\delta_{ij} = \delta_{jk} = \delta_{ik}$—and incorporate them by expanding each equality into a pair of relationships: $\delta_{ij} \geq \delta_{jk}$ and $\delta_{jk} \geq \delta_{ij}$, etc. In effect, any departure from equality between the reconstructed inter-point distances d_{ij}, d_{jk}, d_{ik} is penalised (by reducing the overall likelihood of the configuration); the corresponding points $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$ are pushed towards forming an equilateral triangle.

In addition, estimated dissimilarities δ^*_{ij} were obtained. Each δ^*_{ij} is b_{ij}/c_{ij}, where c_{ij} is the total number of comparisons between δ_{ij} and another dissimilarity δ_{ik} or δ_{jk} (involving some third stimulus E_k), and b_{ij} is the number of such comparisons in which δ_{ij} was larger. In other words, δ^*_{ij} is the fraction of triads in which E_i or E_j was the odd-one-out. Processing these estimates with the group-mean hierarchical-clustering algorithm arranged items in a tree model.

---

[3] This is similar to the ‘‘nonserial matching’’ or Pick 1/*N* procedure applied to the Frois–Wittman facial expression photographs by Andrews and Muldoon (1954).

Results

A three-dimensional solution was chosen. This requires 88 parameters (compared to 59 for a two-dimensional solution). Adding a third dimension leads to an improvement in the overall log-likelihood from -1344 to -1214. A very similar solution results from following the more usual approach of applying MDS to the δ^*_{ij}, rather than directly to the comparisons. However, that approach does not allow the use of the likelihood-ratio test. When comparing rival solutions, twice the difference in their log-likelihoods (i.e., 2×130) follows an asymptotic χ^2 distribution with 29 degrees of freedom: The difference in the numbers of parameters (Takane, 1978). This is significant at the $p < .001$ level.

To bring order to the scatter of points, *constrained* MDS was then applied. Specifically, we imposed the constraint that the points corresponding to pure and blended expressions should be aligned so that each morphed point lies somewhere along a line connecting its pure-expression points: $\mathbf{x}_m = q_m\ \mathbf{x}_a + (1 - q_m)\ \mathbf{x}_b$. This requires a single parameter q_m to locate each morphed item, saving 52 degrees of freedom. Any distortions introduced by the constraint are minor. The improved visual clarity and the reduction in the number of parameters comes at the cost of decreasing the log-likelihood from -1214 to -1252: according to a likelihood-ratio test, this is not significant. Figure 2 is the result.

Only the points corresponding to pure expressions need to be labelled, since the identity of each morphed item is obvious from the position of $\mathbf{x}_m$ on lines between its respective $\mathbf{x}_a$, $\mathbf{x}_b$; the E_m are represented by symbols. The symbols indicate which branch each E_m belongs to in the tree described below.

A salient feature of Figure 2 is the ‘‘clumping’’ of the morph points $\mathbf{x}_m$. They are not spaced out at even intervals; they have gravitated towards one endpoint or the other, as predicted by the CP hypothesis. To illustrate this, crosses are marked at even intervals along the dotted lines between pairs of pure-expression endpoints, each cross linked to the corresponding $\mathbf{x}_m$ by a solid line. Clumping is also present in unconstrained and in two-dimensional solutions. A simple way to test whether the effect is more than a coincidence is to constrain the solution further, setting $q_m = p_m$ so that each $\mathbf{x}_m$ is anchored in place between $\mathbf{x}_a$ and $\mathbf{x}_b$; no longer free to slide back and forth like a bead on a wire. This saves a further 26 degrees of freedom but it reduces the log-likelihood to -1376: sufficient deterioration to reject the hypothesis (at the $p < .001$ level) that the subjective intervals between the morphs in each continuum are equal.

The horizontal axis appears to be a happy-sad dimension, distinguishing positively valenced expressions from negative ones. The vertical axis lends itself to a description such as ‘‘intensity’’ or ‘‘arousal’’: Interpreted in this way, Figure 2 has much in common with the circumplex found in a number of previous MDS studies of facial expression (Nummenmaa, 1992; Russell & Bullock, 1986; Russell, Lewicka, & Niit, 1989) and emotion (Feldman, 1995; Roberts & Wedell, 1994).

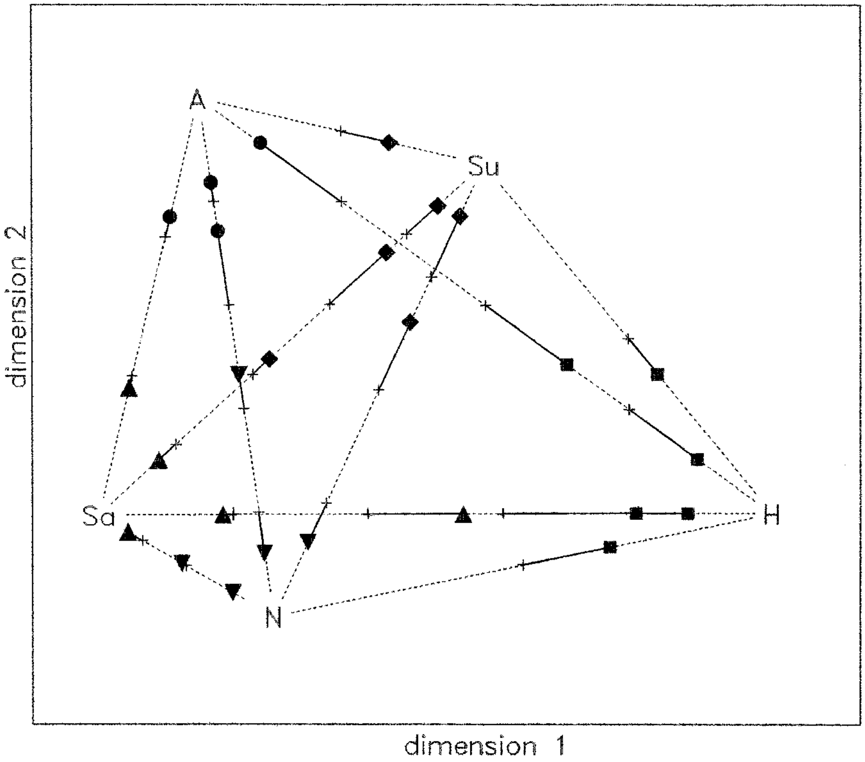


Figure 2. Two-dimensional projection of *constrained* three-dimensional configuration for 31 facial expression items (F-series), derived from triadic data from 13 observers. Each morphed expression is anchored to a line drawn between its constituent pure expressions, and labelled according to the branch it belongs to in a dendrogram, as follows: ●, *Angry*; ▲, *Sad*; ▲, *Neutral*; ◆, *Surprised*; ■, *Happy*. The branches in turn are identified by the pure expression each contains.

However, the third dimension is required to accommodate the large perceived separations between pairs of pure expressions around the circumplex perimeter (*Happy-Neutral*, *Surprised-Happy*, *Angry-Surprised*, *Angry-Sad*). Each pair is too dissimilar to be flattened into two dimensions without distortion (equivalently: the dissimilarities between *Neutral* and the "active" pure expressions are too small). Unblended *Anger* lies at one extreme of this third dimension, and unblended *Surprise* at the other. A previous MDS study of facial affect (Paramey et al., 1994) found a similar axis, distinguishing pure expressions of Anger and Disgust at one extreme from Surprise and Fear at the other. Paramey et al. suggested the interpretation "active-reactive", but attaching a label to each dimension is not essential: Here, the MDS solutions are merely means to an end.

Deriving a tree from the triads confirms that the clumping is not an artefact of a spatial model. One cannot speak of items being spaced between two extremes at intervals, equal or unequal, for the spatial concept of ‘‘continua’’ has no equivalent in a tree structure. However, the crucial claim of CP can be tested: if intra-category similarities do indeed exceed similarities *between* categories, the tree should consist of relatively discrete subtrees or branches.

There are five subtrees, each consisting of a pure-expression stimulus plus the morphed items clustered around it. The members of each subtree are listed in Table 2, and shown in Figure 2 by the symbols representing the items. Note from Figure 2 that *Neutral* and *Sad* expressions are not as spatially distant as we believed when deciding to interpolate three morphs between them. An anonymous reviewer has reminded us that neutral expressions are often labelled as ‘‘Sad’’ when ‘‘Neutral’’ is not among the response options (Ekman & Friesen, 1976). This amounts to a higher sampling density in this region of expression space, producing very similar items, which act as a ‘‘bridge’’ between the two clusters and cause them to merge. In consequence, the subtrees for *Neutral* and *Sad* join at a relatively low dissimilarity.

Pairing-up data. MDS was applied to the subjects’ pairing-up decisions by treating them as comparisons between similarities. Specifically, we interpreted them as Pick 1/15 data: by matching E_i from the top row with E_j from the bottom row—rather than matching E_i with some other bottom-row E_k or E_j with some other top-row E_l—the subject informs us that $\delta_{ij} < \delta_{ik}$, $\delta_{ij} < \delta_{jl}$ for all $k, l \neq i$ or j.

A deficiency of pairing-up judgements is that they provide information about the configuration’s short-range details, but not about its global structure. This creates a possibility that the MDS analysis may become stuck in a local maximum, and fail to find the solution with the best possible goodness-of-fit. The initial configuration becomes a factor. To remedy this, we used the configuration derived from triadic data to initialise the MDS process. Items appear to be clumped in the result, but it is not possible to ascribe a significance to this, since a condition for applying the likelihood-ratio test is that all the comparisons are independent. This is not the case for pairing-up judgements: once E_i, E_j are paired, neither are available for pairing with another item E_k, even if $\delta_{ik} < \delta_{kl}$ for all other E_l.

The configuration from constrained MDS is similar to Figure 2. For reasons of space it is not shown here, but the product-moment correlation between the respective d_{ij} is $r = .72$. A tree (summarised in Table 2) shows the items clustering in the same fashion as in that produced from triadic data, with minor differences.

EXPERIMENT 2: M-SERIES. QUARTET DATA

Although unlikely, the possibility remains that the observed clustering of items is an objective feature of the particular stimuli used, rather than a CP phenomenon. Perhaps the morphs were not spaced equally along the 10

TABLE 2
Summary of six hierarchical trees produced for the two series of 31 stimuli, in four experiments

	F-series				*M-series*		
	1a	1b	3	4	2	3	4
A	A	A	A	A	A	A	A
AH25	A	A	A	A	A	A	A
ASa33	A	A	A	A	A	A	A
AN20	A	A	A	A	A	A	A
AN40	A	A	A	A	A	A	A
Sa	Sa	Sa	Sa	Sa	Sa	Sa	Sa
SaN25	Sa	Sa	Sa	Sa	Sa	Sa	Sa
SaN50	*N*	Sa	Sa	*N*	Sa	Sa	*N*
SaH20	Sa	Sa	Sa	Sa	Sa	Sa	Sa
ASa66	Sa	*N*	Sa	Sa	Sa	Sa	Sa
Sa20Su	Sa	Sa	Sa	Sa	Sa	Sa	Sa
Sa40Su	*Su*	*Su*	Sa	Sa	Sa	Sa	Sa
N	N	N	N	N	N	N	N
AN60	N	N	N	N	N	N	N
AN80	N	N	N	N	N	N	N
Sa75N	N	N	N	N	N	N	N
SuN75	N	N	N	N	*Su*	N	N
Su	Su	Su	Su	Su	Su	Su	Su
SuN25	Su	Su	Su	Su	Su	Su	Su
SuN50	Su	Su	Su	Su	Su	Su	Su
SaSu60	Su	Su	Su	Su	Su	Su	Su
SaSu80	Su	Su	Su	Su	Su	Su	Su
ASu50	Su	Su	Su	Su	Su	Su	Su
H	H	H	H	H	H	H	H
AH50	H	*A*	H	H	H	H	H
AH75	H	H	H	H	H	H	H
SuH50	H	H	H	H	H	H	H
SaH40	*Sa*	*Sa*	H	H	H	H	H
SaH60	H	H	H	H	H	H	H
SaH80	H	H	H	H	H	H	H
HN50	H	H	H	H	H	H	H

Note: Each column represents a tree. Each stimulus (row) is labelled A, Sa, N, Su, H, according to the branch it belongs to in that tree, where each branch iin turn is identified by the pure expression it contains. Columns 1a, 1b for the F-series stimuli refer to the trees derived from triadic and pairing-up data from Experiment 1. For Experiment 4, combining both stimulus series in a single tree, each stimulus is labelled according to the nearest pure expression from the same series as itself.

Six stimuli are borderline cases (SaN50, ASa66, Sa40Su, SuN75, AH50, SaH40), not grouped in the same category in every tree.

continua, despite the best efforts of the morphing-software operator. To guard against this, a second stimulus set was generated (the M-series), as described in Experiment 1 above, under ‘‘Stimuli’’. Moreover, a different procedure was applied: a ‘‘quartet task’’.

Subjects

Ten adult informants were recruited informally (two men, four women, four unrecorded).

Method

In this procedure the subjects were instructed to shuffle the items and deal them out into 7 groups of four ‘‘quartets’’, plus a leftover group of three. The task was to indicate the pattern of similarities within each quartet by grouping it as (a) two pairs of similar items, or as (b) a group of three similar items, plus an odd-one-out. Subjects were free to apply (a) or (b), depending on which arrangement fitted each quartet best.[4] ‘‘Similarity’’ was explained in the same way as in Experiment 1. The leftover group of three was treated as a triad. Subjects were also invited to make supplementary decisions: for case (a), they could indicate which pair was more similar; for case (b), they could provide a triadic odd-one-out judgement for the relatively similar group of three.

This whole process of creating and analysing quartets was repeated 10 times. As before, the task was designed to be taken home and completed at the subject’s own pace. To compensate for the greater complexity of the task, the number of quartets is less than the number of triads in Experiment 1.

Analysis. Quartet judgements can be understood and analysed in terms of comparisons between pairs of distances. Consider a quartet E_i, E_j, E_k, E_l: if a subject pairs up the items as E_i, E_j and E_k, E_l, we interpret this as a judgement that the sum of dissimilarities $\delta_{ij} + \delta_{kl}$ is less than the sums that correspond to the two alternative ways of pairing the quartet ($\delta_{ik} + \delta_{jl}$, and $\delta_{il} + \delta_{jk}$). If the subject selects an odd one out, E_l, this is equivalent to asserting that E_i, E_j, E_k form a tighter cluster (i.e. are more similar to each other) than the clusters left by the alternative choice of E_i or E_j or E_k as odd one out. This simplifies to the judgements that:

$$\delta_{ik} + \delta_{ij} < \delta_{kl} + \delta_{jl}, \quad \delta_{ij} + \delta_{jk} < \delta_{il} + \delta_{kl}, \quad \delta_{ik} + \delta_{jk} < \delta_{il} + \delta_{jl}.$$

---

[4] Quartet data were collected by Isaac (1970), and by Tantam, Monaghan, Nicholson, and Stirling (1989), respectively using subsets of Frois–Wittman facial expressions and of Ekman–Friesen photographs. In both studies, the data were used for assessing the accuracy of the subject’s perception, rather than for MDS purposes.

Scaling quartet data is an iterative process of changing the modelled distances d_{ij} (by moving points $\mathbf{x}_i$) to progressively improve the fit between the observations and the model, at each stage comparing these judgements against the corresponding sums of distances ($d_{ij} + d_{kl}$, etc.). For purposes of constructing a tree, the estimated dissimilarities are $\delta^*_{ij} = b_{ij} / c_{ij}$, where c_{ij} is the number of comparisons (involving other items E_k, E_l) between $\delta_{ij} + \delta_{kl}$ and some other sum of dissimilarities, and b_i is the number of such comparisons in which $\delta_{ij} + \delta_{kl}$ was the larger.

Results

For a spatial solution, three dimensions gave a significantly better fit than two dimensions—87 degrees of freedom instead of 59—the difference in log-likelihood being 134 ($p < .001$). Constraining the $\mathbf{x}_m$ to lie on lines between their respective $\mathbf{x}_a$, $\mathbf{x}_b$ reduced the log-likelihood by 61 while saving 52 degrees of freedom. The difference borders on significance, but for the sake of clarity, this constrained solution—highlighting the displacement of morphs towards pure expressions—is the one shown in Figure 3. Imposing the additional constraint that $q_m = p_m$ (saving 26 degrees of freedom) reduced the log-likelihood by a further 45; this *is* significant ($p < .05$), demonstrating that the clumping of items is not merely coincidental.

The contrast between Figures 2 and 3 is intriguing. Repeating the MDS using sorting data produces solutions which differ in the same way (Experiments 3 and 4, below). Within each category, there seems to be considerable latitude for expressions to vary while still considered to be good exemplars of that emotion. For example, the pure expression of *Anger* in the F-series of stimuli (e61) can be described in the FACS system as composed of Action Units 4+5+23; while its counterpart in the M-series (e103) is composed of AUs 4+5+7+2+26 (P. Ekman, personal communication): The latter incorporates two AUs absent in the former, ‘‘lids tight [squint]’’ and ‘‘jaw drop’’. This agrees with the finding (Alvarado, 1996) that more than one distinguishable expression can come under the rubric of a single verbal label. In that study, three expressions which differed in terms of facial action (i.e. the contraction of particular muscles), and were seen as dissimilar in triadic comparisons, were consistently identified as *Surprised* (or as *Angry*, for another three). It becomes possible to ask whether the ‘‘focal expression’’ in such cases is a single point in expression space, or a set of points. Any attempt at a general model of facial expressions in spatial terms must allow for these ‘‘actor differences’’.

The location of unblended *Neutral* in Figure 3 deviates from the expected circumplex. The third dimension again serves to distinguish *Anger* from *Surprise*, but in this case the *Neutral* expressions are displaced towards the *Surprise* extreme of the axis. This renders any identification of the dimensions in Figure 3 problematical, apart from the ‘‘pleasant-unpleasant’’ valence of *D1*.

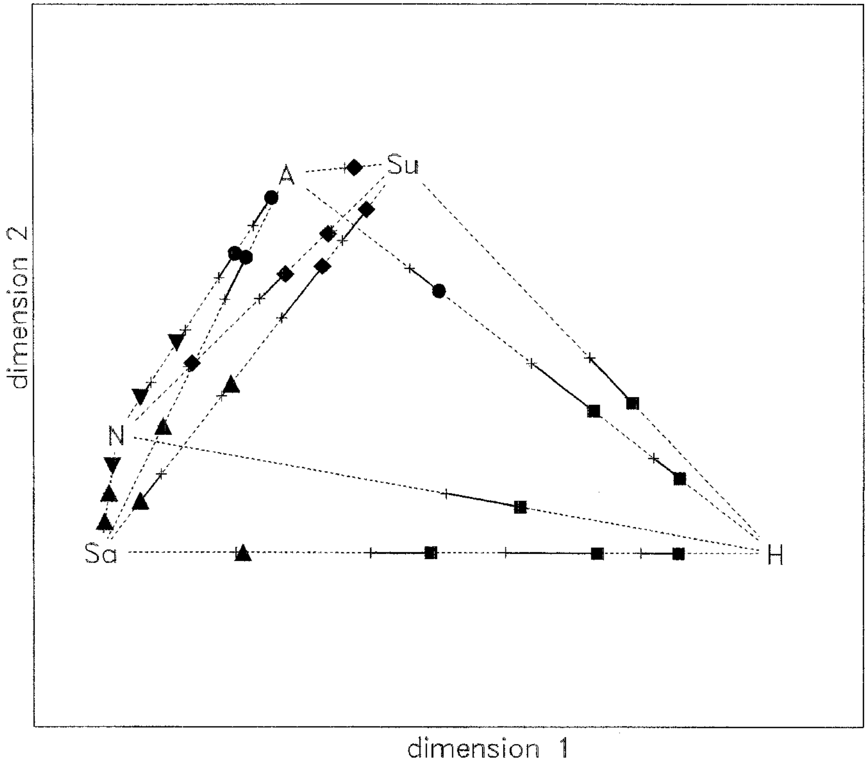


Figure 3. First two dimensions of *constrained* three-dimensional configuration for 31 facial expression items (M-series), derived from quartet data from 10 observers. Each morphed expression is labelled according to the branch it belongs to in a dendrogram, as follows: ●, *Angry*; ▲, *Sad*; ▲, *Neutral*; ◆, *Surprised*; ■, *Happy*.

Further evidence that the clustering of M-series items into categories is more than a coincidence comes when the δ^*_{ij} derived from the quartet comparisons are converted into a tree. The tree consists of five subtrees or category branches, each containing a pure expression and the morphs most similar to it. The membership of each category is substantially the same as for the trees derived for F-series items in Experiment 1 (Table 2).

EXPERIMENT 3: M- AND F-SERIES. SORTING DATA

At this stage the objection could be made that although observers found the triad and quartet procedures straightforward, they are not natural activities. Whether they have any relevance outside the laboratory is a moot point. Each item is viewed in the context of two or three other items, whereas judgements made about an expression have more ecological validity when it is seen in the context of the portrayer's entire expressive gamut (Ekman, O'Sullivan, & Matsumoto, 1991). This objection can be countered by applying the Method of Sorting.

This procedure has a long history in the study of facial expressions of emotion, starting with Hulin and Katz (1935), who asked 22 observers to arrange the 72 Frois–Wittman pictures into groups representing distinct expressions. To cite but a few examples, it has been applied to pure and mixed expressions (Nummenmaa, 1992), and to pure expressions, sorted by preschool children (Russell & Bullock, 1986), and by adults of several cultures (Russell et al., 1989). However, sorting data are unspecific about the global structure of a configuration: If all or most subjects agree in placing E_i and E_j in separate groups, we know that they are dissimilar—but not *how* dissimilar. To elicit more information, this study uses ''Additive sorting'', an extension of the basic procedure (Bimler, Kirkland, & Chen, 1998).

Subjects and stimuli

The stimuli were the two sets used in Experiments 1 and 2, each containing 31 items. Twenty-three adult subjects (sexes not recorded) sorted the F-series and 23 sorted the M-series.

Procedure

Additive sorting involves several steps. In the first step, each subject was requested to group together items which ''belonged together'' or were most similar. The number of groups and the number of items in each group were left up to the subjects (single-item groups were permitted). Subjects were also left to make their own interpretation of ''similarity''. If clarification was requested, ''similarity'' was explained as in Experiment 1 (i.e., '' 'How similar are the *emotions* expressed in the photos?' or 'How similar are the person's feelings?' '').

In the following steps, subjects were invited to reduce the number of groups, by selecting the two ''most similar'' groups and merging them into one. They repeated this merging until only two groups remained, or until the remaining groups had so little in common that nominating two of them as most similar was not possible.

Results

In the first stage of sorting, subjects created an average of 8.8 groups (F-series) and 8.5 groups (M-series). This is large enough to make it unlikely that subjects were simply grouping the items into familiar expression categories.

We applied ''reconstructed dyads analysis'' to the data. Essentially this is a way of rendering each subject's sorting decisions into a form amenable to Maximum Likelihood MDS, by decomposing them into a sequence of comparisons between similarities. This approach has previously been applied to sorting data for photographs of facial affect (Bimler & Kirkland, 1997; Bimler et al., 1998) and found to have advantages over the standard approach of applying MDS to the aggregated sorting ‘‘co-occurrences’’.

Again, blended expressions become polarised, gravitating towards the dominant component. Tree models for the F-series and M-series both contain five clearly resolved branches, each containing a ‘‘pure’’ expression and thus identifiable as an emotion category. The membership of these branches is the same as in the trees from Experiments 1 and 2 (Table 2).

Spatial solutions for these data exhibited a comparable degree of clumping. Three-dimensional configurations were chosen, on the basis of the third dimension agreeing with those derived for the triadic and quartet data. For both stimulus sets, saving 52 degrees of freedom by constraining the morphed items to lie on lines between the prototype items reduced the total log-likelihood by a relatively small amount, while saving a further 26 degrees of freedom by setting $q_m = p_m$ produced a larger reduction. However, sorting data do not meet the condition of independence of comparisons, so it is not possible to apply the χ^2-test to these differences. The positions of the prototype points in the constrained three-dimensional solutions are very similar to those shown in Figures 2 and 3, as is the general extent of clumping. Comparing the d_{ij} of the F-series solution with those derived from triadic data, the product-moment correlation $r = .90$. Comparing the d_{ij} of the M-series solution with those derived from quartet data, $r = .89$.

EXPERIMENT 4: M- AND F-SERIES COMBINED. SORTING DATA

Alternative explanations, not involving CP, can be offered for the clumping observed in Experiment 3. It is possible that the five original items stood out in some way, providing the subjects with natural nuclei to cluster the 26 synthesised items around (the large number of groups made in the Grouping stage of sorting militates against this). For instance, it may be that despite the skill and best efforts of the software operator, artefacts appear in the morphed expressions which distinguish them from the pure ones. It is also worth remembering that each pure item lies at the end of four continua, with several slight variations in its neighbourhood, whereas each morph lies somewhere on a single continuum, with two neighbours. One can imagine the slight variants calling attention to the pure item, which epitomises their shared qualities.

A test of this explanation would be to repeat Experiment 3—without all of the pure items. Even when absent, do the perceptual magnets still exert an attractive force? We applied a test along similar lines. The stimulus sets for the subjects were randomised, so that some pure expressions were not available for sorting for half the subjects. Moreover, each subject saw a different selection of each pure item's immediate neighbours. This was done by combining the Male and Female stimuli, shuffling them, and splitting the deck, repeating this process for every pair of subjects.

Subjects

One group of subjects consisted of 21 adults. A second group consisted of 70 students (36 males, 32 females, two unrecorded; aged 13 to 15) recruited from local high schools. Observers in this age range have matured enough to perform close to adult level in tasks of facial expression recognition (Kolb, Wilson, & Taylor, 1992). De Gelder et al. (1997) found that 9- to 10-year-old children identified and discriminated morphed FE stimuli in a manner qualitatively similar to adults.

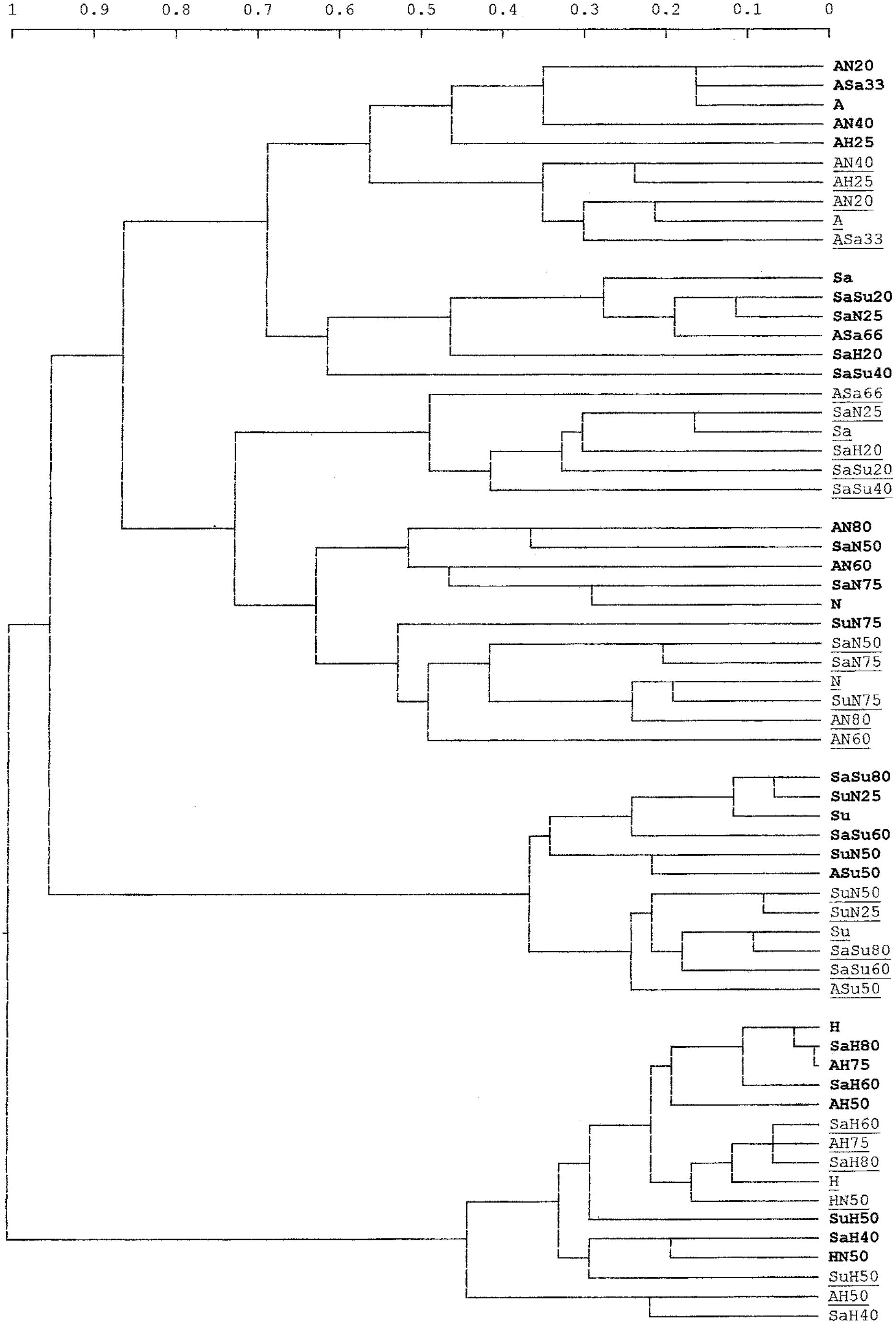
Procedure

The additive sorting method was used. Before applying it to the facial expression items, the students practised sorting a set of 21 animal names (analysis of these data, and comparison with previous MDS studies of the animal-name domain, confirmed that they understood the task). The instructions to subjects urged them to overlook (as far as possible) any differences in facial *appearance* between the male and female poser, and to sort the pictures into groups on the basis of similarities between the emotions they portrayed. Observers are capable of distinguishing these two sources of overall dissimilarity as the recognition of faces and facial affect are processed separately in the brain (Etcoff, 1984). Adult subjects created an average of 9.9 groups in the first stage of sorting. The students arranged the items in an average of 6.1 groups, suggesting that they were applying looser criteria of similarity.

Results

For the purposes of reconstructed dyad analysis, the split-deck incomplete design is useful. It ensures that a given item appears in different contexts, and that every subject is forced to make different sorting decisions, providing information which might not be available if everyone had sorted the entire 62 items of the combined sets (Bimler & Kirkland, 1997).

Many observers found that the male and female posers' portrayals of some expression categories were sufficiently dissimilar to be sorted into separate groups. For example, of the 26 observers who were presented with both portrayals of pure *Angry*, only 11 grouped them together at step 1. This was not a matter of ignoring instructions and sorting by gender, as subjects showed no reluctance about grouping both portrayals of *Happy* together; some of the expressions genuinely differ between posers, in intensity or quality. This is shown quite clearly in the dendrogram from these data (Figure 4). The expected


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
AN20
ASa33
A
AN40
AH25
AN40
AH25
AN20
A
ASa33
Sa
SaSu20
SaN25
ASa66
SaH20
SaSu40
ASa66
SaN25
Sa
SaH20
SaSu20
SaSu40
AN80
SaN50
AN60
SaN75
N
SuN75
SaN50
SaN75
N
SuN75
AN80
AN60
SaSu80
SuN25
Su
SaSu60
SuN50
ASu50
SuN50
SuN25
Su
SaSu80
SaSu60
ASu50
H
SaH80
AH75
SaH60
AH50
SaH60
AH75
SaH80
H
HN50
SuH50
SaH40
HN50
SuH50
AH50
SaH40

five-way branching of the tree is present: the strength of the perceptual magnet effect was not diminished by the split-deck manipulation of the items. However, the F-series and M-series members of the *Angry* cluster (five each) are sufficiently dissimilar to form distinct subclusters. In contrast, the F-series and M-series members of the *Happy* cluster (five each) are mingled together. Other clusters show intermediate degrees of subclustering.

Spatial MDS was also performed, and a three-dimensional solution was chosen. As before, the constraint was imposed that each point indicating a morphed stimulus should lie somewhere on a line between the appropriate endpoints: 10 endpoints in this case, and 20 continua. The items congregate into clumps to a comparable extent as in Figures 2 and 3. Unfortunately, the number of points and lines detract from the clarity of a two-dimensional projection of the solution, so it is not included here.

DISCUSSION

The techniques of multidimensional scaling (MDS) were used to produce spatial models (incorporating constraints) and nonspatial models (trees), summarising the similarities perceived among expressions of facial affect. At least three dimensions were required for an adequate representation of the data. Tempting though it is, we are loath to speculate about the interpretation of these axes, given the extent of ‘‘actor variation’’ between the F-series and M-series stimuli. And as with any MDS application, it is important to remember that the axes arise from geometrical assumptions impinging on the data; they may or may not correlate with phenomena outside the models.

Every effort has been made to avoid the possible artefacts of MDS. In particular, the common procedure of estimating δ^*_{ij} from triadic data by ‘‘vote-counting’’ the triads (e.g., Alvarado, 1996) can produce spurious circumplex solutions in which items which belong near the centre are displaced outward toward the periphery of the perceptual space (Gladstones, 1962)—a kind of centrifugal force. Similar distortions appear when estimating dissimilarities from sorting data. ‘‘Scaling methods [...] seem often to give circular models, possibly in the form of a ring. This happens whatever the stimuli are’’ (Nummenmaa, 1992). In the constrained solution, this would have been enough to force the morphed items towards the pure items, which occupy extremal positions in the configuration, marking the vertices of a polyhedron (but note that clumping is also present in the unconstrained solution). As well as the centrifugal effect, aggregated sorting data (co-occurrences) have a propensity to exaggerate any clustering of the items (Bimler & Kirkland, 1997). To counter this, the

Figure 4 (opposite). Hierarchical tree derived from sorting data for 62 facial expression items. Codes in bold (**A50H**) indicate M-series expressions; underlined (A50H) indicates F-series.

intermediate vote-counting stage is bypassed by the MDS algorithm we applied, which operates directly on the comparisons.

To summarise the Results sections above: Perceived dissimilarities are decreased in the neighbourhoods of the unblended expressions, in an analogous manner to the ‘‘perceptual magnet’’ effect. The data are too scanty to try pinpointing the focus of each category—the prototypal, most representative expression—or to tell whether these foci coincide with the unblended items. The evidence for this within-category shrinkage of dissimilarities persists across four different forms of data, and across spatial versus tree representations. Subject to the following caveats, it seems that categorical perception (CP) effects are not merely artefacts of the particular procedures used in previous experiments with morphed expressions.

We have concentrated on the within-category shrinkage, placing less emphasis on the discontinuities *between* categories. Judging from the agreement about the membership of each categories (Table 2), the boundaries remain stable for two sets of stimuli, and across different forms of data. However, the spacing of the interpolated morphs is too coarse to locate these boundaries precisely; they may vary between subjects (indeed, if there is a projective component, they may vary for a single subject, depending on mood and circumstances). Note that a boundary is a region of quite specific uncertainty. One might expect that an item midway between *Happy* and *Angry* (for instance)—making it ambiguous or hard to categorise—would therefore cross over to the *Neutral* cluster, but there is no evidence of this happening (in agreement with Calder et al., 1996). Young et al. (1997) noted that this absence of FE ‘‘metamers’’ is enough to exclude a two-dimensional spatial model: Normally, MDS applications rely on goodness-of-fit measures to decide on the appropriate dimensionality.

Criteria for membership of the *Neutral* cluster appear to be more specific than that. We have replicated earlier findings (Etcoff & Magee, 1992) that a *neutral* expression has the properties of a category. When blended with an ‘‘active’’ expression, the proportion of the latter must reach some threshold before the combination loses its neutral status. In addition, *Surprised* acts as a perceptual magnet, along the *Surprised-Neutral* and *Sad-Surprised* continua. This was also observed by Young et al. (1997). This raises the interesting possibility that a categorical status for *Surprised* expressions relies on certain features of shading or skin texture which are lost in line drawings, as Etcoff and Magee (1992) who used such stimuli, found no peaks in discrimination that would indicate category boundaries when they compared *Surprised* with *Angry* and *Afraid*.

Note that instead of measuring subjects’ discrimination between stimuli (just-noticeable differences)—the second criterion for CP (Harnad, 1987)—we have focused on the perceived dissimilarities between stimuli with larger differences. MDS looks for structure in the relationships between stimuli, as opposed to labelling each stimulus or analysing its qualities in isolation. Thus, this study ignores the first criterion. These caveats mean that our data do not prove the existence of CP for facial expressions; they can more modestly be described as compatible with it. Our results should ideally be repeated with a larger pool of individuals displaying the expressions.

Data from four experiments displayed the signs of CP when tree models were derived from them, but these signs remained when dimensional models were derived instead. Etcoff and Magee (1992) state that the CP effects they demonstrated rule out a spatial representation of expressions of affect; (i.e., that CP is inherently tree-like and nonspatial). We argue, to the contrary, that the two are compatible (see also Ekman et al., 1972, chapter 12), and that CP can be easily accommodated within the framework of spatial models: Although it might be more appropriate to talk of ''perceptual magnets'' in this situation.

The analysis with MDS is one point of difference between this study and earlier research with morphed faces. Another is the parallel presentation of stimuli. Here, subjects had the entire gamut of expressions in front of them as they sorted the items or matched them into pairs. Setting an expression in this kind of context increases the ecological validity of judgements made about it (Ekman et al., 1991): Expressions are categorised more accurately when subjects have prior experience of that actor's expressive range. In the case of the triadic data (direct comparisons of dissimilarities), subjects viewed each three stimuli concurrently. After the first 10 triads, they were acquainted with the full range of stimuli (many felt overacquainted with it, after 120 triads).

It is possible that subjects were mentally pigeon-holing the expressions—attaching verbal labels to them and grouping items which they had labelled the same. If this was the case then the tasks used here were still identification tasks and their results confound CP with the necessarily categorical nature of those verbal labels. Unlike sequential-presentation tasks, the triadic and sorting procedures do not *enforce* the use of an encoding strategy, but they do not exclude it either. Verbal encoding has been shown to contribute to CP of colours. In a cross-cultural comparison, using a triadic method to measure the dissimilarities between coloured chips, Kay and Kempton (1984) found evidence for a category boundary between *blue* and *green* for English-speaking informants, whereas no such boundary emerged for observers whose native language does not verbalise a *blue-green* distinction (replicated by Davies, Sowden, Jerrett, Jerrett, & Corbett, 1998). This category boundary dissolved when the experiment design was manipulated to discourage observers from basing their decisions on a verbal-encoding strategy. However, the large average number of groups in the sorting tasks of Experiments 3 and 4 militates against a verbal-encoding explanation for the present results, as does the distinction found in Experiment 4 between F- and M-series exemplars of some pure emotions. Alvarado (1996) presents similar contrary evidence from multiple exemplars.

Recall that the instructions in all four experiments were worded so as to focus the attention of the observers on the emotions expressed in the stimuli, and away from physical characteristics, such as the identity of the poser, in Experiment 4.

This leaves open the possibility that the evidence for categoricity relates specifically to emotion concepts, rather than to the visual representation of stimuli (we are indebted to two anonymous reviewers for pointing this out). Different results and a shift in the balance between dimensional and categorical perception might arise from differently worded instructions (for instance, inviting observers to judge similarity by comparing stimuli as abstract patterns of light and shade, or as abstract combinations of displacements of facial landmarks).

The wider context of this study is the perennial issue of whether the variations among facial-affect expressions are categorical or dimensional. The question is of interest, not merely in itself, but also for the implications it carries for the structure of emotions. Research within the categorical approach has long been dominated by the use of prototypal stimuli, and forced-choice identification. The advent of morphing techniques allow the possible limitations in this procedure to be avoided. Our observation of CP effects implies that the category to which a FE belongs is an important element in its description, at least.

There are limits, however, in how far this can be extrapolated to emotions. The latter seem to co-occur in various combinations—indeed, such co-occurrence is presupposed for studies of mood structure using self-report data (Feldman, 1995). It must be remembered that FEs, as a channel for communicating emotional signals, have a limited capacity: A single bit, according to Osgood's (1966) information-theory analysis. Under normal conditions the channel is combined with information from situation and tone of voice. Our ability to extract emotional content from degraded signals such as line drawings (Etcoff & Magee, 1992; Paramey et al., 1994), or static, monochrome photographs speaks of a system evolved to maximise redundancy and robustness rather than information capacity, and it may be that the categories are a product of these communication constraints.

Manuscript received 10 September 1999
Revised manuscript received 7 February 2001

REFERENCES

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*, 669–672.

Alvarado, N. (1996). Congruence of meaning between facial expressions of emotion and selected emotion terms. *Motivation and Emotion*, *20*, 33–61.

Andrews, T.G., & Muldoon, J. F. (1954). Factorial study of Schlosberg's analysis of facial expressions. *American Psychologist* [Abstract], *9*, 321.

Bimler, D.L., & Kirkland, J. (1997). Multidimensional scaling of hierarchical sorting data applied to facial expressions. *Scandinavian Journal of Psychology*, *38*, 349–357.

Bimler, D.L., Kirkland, J., & Chen, S.-J. (1998). *A cross-cultural study of facial expressions of emotion using multidimensional scaling*. Annual Report, Research and Clinical Center for Child Development (Hokkaido University), *21*, 35–42.

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

Calder, A.J., Young, A.W., Rowland, D., & Perrett, D.I. (1997). Computer-enhanced emotion in facial expressions. *Proceedings of the Royal Society of London B*, *264*, 919–925.

Carroll, J.M., & Russell, J.A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70*, 205–218.

Davies, I.R.L., & Corbett, G. G. (1997). A cross-cultural study of colour-grouping: Evidence for weak linguistic relativity. *British Journal of Psychology*, *88*, 493–517.

Davies, I.R.L., Sowden, P.T., Jerrett, D.T., Jerrett, T., & Corbett, G.G. (1998). A cross-cultural study of English and Setswana speakers on a colour-triads task: A test of the Sapir-Whorf hypothesis. *British Journal of Psychology*, *89*, 1–15.

de Gelder, B., Teunisse, J.-P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion*, *11*, 1–23.

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

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

Ekman, P., Friesen, W.V., & Ellsworth, P. (1972). *Emotion in the human face*. New York: Pergamon.

Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991). Confusions about context in the judgment of facial expression: A reply to 'The contempt expression and the relativity thesis'. *Motivation and Emotion*, *15*, 169–176.

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

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

Feldman, L.A. (1995). Variations in the circumplex structure of mood. *Personality and Social Psychology Bulletin*, *21*, 806–817.

Gladstones, W.H. (1962). Empirical comparisons of three multidimensional scaling methods. *Australian Journal of Psychology*, *14*, 197–201.

Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition and Emotion*, *13*, 225–266.

Harnad, S. (1987). *Categorical perception*. Cambridge, UK: Cambridge University Press.

Hulin, W.S., & Katz, D. (1935). The Frois–Wittmann pictures of facial expression. *Journal of Experimental Psychology*, *18*, 482–498.

Isaac, P.D. (1970). Dissimilarities as indices of individual perceptual structure. *Perception and Psychophysics*, *7*, 229–233.

Iverson, P., & Kuhl, P.K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*, 553–562.

Izard, C.E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*, 65–79.

Kolb, B., Wilson, B., & Taylor, L. (1992). Developmental changes in the recognition and comprehension of facial expressions: Implications for frontal lobe function. *Brain and Cognition*, *20*, 74–84.

Nummenmaa, T. (1992). *Pure and blended emotion in the human face*. Helsinki: Suomaleinen Tiedakatemia.

Osgood, C.E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology*, *7*, 1–30.

Paramey, G.V., Schneider, K., Josephs, I., & Slusarek, M. (1994). Identification of emotional meaning in line drawings of faces. Report 44, Faculty of Psychology, Ruhr-University Bochum.

Prkachin, G.C., & Prkachin, K.M. (1994). Adaption to facial expressions of emotion. *Cognition and Emotion*, *8*, 55–64.

Roberts, J.S., & Wedell, D.H. (1994). Context effects on similarity judgments on multidimensional stimuli: Inferring the structure of the emotion space. *Journal of Experimental Social Psychology*, *30*, 1–38.

Russell, J.A. (1994). Is there universal recognition of emotions from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, *115*, 102–141.

Russell, J.A., & Bullock, M. (1986). On the dimensional pre-schoolers use to interpret facial expressions of emotion. *Developmental Psychology*, *22*, 97–102.

Russell, J.A., & Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *Journal of Experimental Psychology: General*, *116* (3), 223–237.

Russell, J.A., Lewicka, M., & Niit, M. (1989). A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Research*, *57*, 848–856.

Takane, Y. (1978). A Maximum Likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—Theory. *Japanese Psychological Research*, *20*, 7–17.

Tantam, D., Monaghan, L., Nicholson, H., & Stirling, J. (1989). Autistic children's ability to interpret faces: A research note. *Journal of Child Psychology and Psychiatry*, *30*, 623–630.

Tomkins, S.S. (1962). *Consciousness, imagery and affect: Vol. 1*. New York: Springer.

Woodworth, R.S. (1938). *Experimental psychology*. New York: Holt.

Young, A.W., Rowland, D., Calder, A.J., Etcoff, N.L., Seth A., & Perrett, D.I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*, *63*, 271–313.