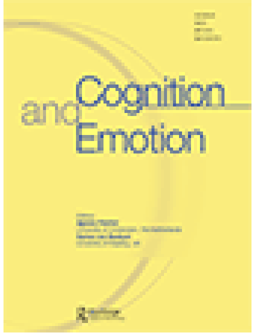

Cognition and Emotion

Routledge
Taylor & Francis Group

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: http://www.tandfonline.com/loi/pcem20

Categorical Perception of Facial Expressions: Categories and their Internal Structure

Beatrice de Gelder , Jan-Pieter Teunisse & Philip J. Benson

To cite this article: Beatrice de Gelder , Jan-Pieter Teunisse & Philip J. Benson (1997) Categorical Perception of Facial Expressions: Categories and their Internal Structure, Cognition and Emotion, 11:1, 1-23, DOI: 10.1080/026999397380005

To link to this article: http://dx.doi.org/10.1080/026999397380005

COGNITION AND EMOTION, 1997, *11* (1), 1–23

Categorical Perception of Facial Expressions: Categories and their Internal Structure

Beatrice de Gelder
Tilburg University, The Netherlands and *Université Libre de Bruxelles, Belgium*

Jan-Pieter Teunisse
Tilburg University, The Netherlands

Philip J. Benson
Oxford University, UK

The question whether judgements of facial expression show the typical pattern of categorical perception was examined using three sets of 11 photographs, each constituting an 11-step continuum extending between two extreme protypical exemplars: angry-sad, happy-sad and angry-afraid, respectively. For each continuum, intermediate exemplars were created using a morphing procedure. Subjects first identified all faces in each continuum in terms of the extreme expressions, and then performed an ABX discrimination task on pairs of faces two steps (Experiments 1 and 2) or three steps (Experiment 3) apart. The classical categorical perception prediction that discrimination performance must peak around the point on the continuum at which identification reaches 50% was tested not on group means, as in earlier

---

Requests for reprints should be sent to Dr B.L.M.F. de Gelder, Tilburg University, Postbus 90153, 5000 LE Tilburg, The Netherlands.

Research presented in this paper was partly supported by a grant to the first author (Ministry of Education and Scientific Research of the Belgian French Community, Action de Recherche Concertée ''Language processes in different modalities''). The studies were presented at the 30th Meeting of the Psychonomic Society, St. Louis, November 1994, at Haskins Laboratories and at the Psychology Department, State University of Arizona, Tempe in January 1995. We thank P. Piree and E. Mulder for assistance in testing the subjects. We are indebted to Nico Frijda, Paul Bertelson, and Jean Vroomen for valuable discussion.

This project was partly supported by the Ministry of Education of the Belgian French-speaking Community (Concerted Research Action 91/96–148). Stimuli were manufactured by D. Rowlands (St. Andrews, Scotland) under the sponsorship on UK ESRC grant No. R000234003 to A. Young (MRC-APU Cambridge, England) and D. Perrett (St. Andrews, Scotland). Thanks should go to P. Bertelson for suggestions with respect to data analysis and for insightful comments on the manuscript.

studies, but on a subject-by-subject basis. It was supported by the results for both adults (Experiment 1) and 9- to 10-year-children (Experiment 3). For adults, two noncategorical interpretations of the main finding were discarded by showing that it was not replicated with the same material presented upside down (Experiment 2).

INTRODUCTION

Like a wide range of other perceptual objects, some major facial expressions of emotion are easily recognised and distinguished from one another by normal subjects even across different cultures (see Ekman, 1994, for a recent overview). Among the many questions raised by this perceptual fluency a major one concerns its functional basis. Expression perception might be based on an acquired mental lexicon of facial expressions built up in the course of communicative development and under the control of general learning principles. It might also reflect a biologically endowed ability for expression recognition. In the latter case, perception of facial expressions would proceed by assigning tokens of facial expressions to cognitive categories representing the basis elements or the primitives of the human facial emotion lexicon.

This latter approach to categorisation ability is well illustrated by research on speech perception. Liberman (1996) presents a comprehensive discussion of the early rationale for categories in speech perception. The basic methodology of speech experiments investigating categorical perception (CP) took advantage of speech synthesis that became available in the 1960s. A typical experiment presented listeners with a synthetically created multistep continuum inserted between two end-points made up of two natural tokens. The goal of speech researchers was to study the subjective perception of a linguistically meaningful boundary emerging from the acoustic continuum.

The original CP experiments (see Liberman, 1996; Liberman, Harris, Hoffman, & Griffith, 1957, for a historical overview) challenged a common intuition that categorisation of speech sounds would vary monotonically with the physical stimulus value. Instead, a clear discontinuity in subjective judgements of phonetic information was observed. Although discussions on the theoretical interpretation of CP have continued, it is generally agreed that the basic phenomenon of CP has the following profile: (a) stimulus identification has a steep slope with a somewhat abrupt change in labelling probabilities somewhere along the continuum corresponding to the hypothetical category boundary; (b) a peak in sensitivity is found in the discrimination function at the category boundary at a location corresponding to the change in labelling probability; (c) within-category discrimination is at chance level; (d) discrimination is predicted by the labelling functions.

In the classical analysis of the CP paradigms these requirements should be met jointly (see Repp, 1984, for a more detailed analysis and critical approval). Although each of these requirements has been criticised, there is a clear baseline requirement resulting from those criticisms. The pattern of evidence that is desirable for CP is that there is a maximum in the discrimination performance and that this is found at that point in the continuum where, in the identification task, the two responses are equiprobable (the 50% identification point). In addition, another categorical characteristic might be a slowing down of the identification decision at about the same point.

It is important to clarify the analogy between CP in speech and in facial expressions and this for theoretical as well as for methodological reasons. The basic notion underlying the analogy is that expression categories emerge in the perception of a continuous stimulus domain. In contrast to CP experiments, where one dimension, such as voice onset time (VOT), is manipulated to synthesise a continuum, in the stimuli used here multiple properties of the face are covaried. Although very little is known about how expressive facial information is conveyed, it seems reasonable that the face, as a configuration, plays an important role here. There is evidence that some semantic aspects of face perception rest critically on the presence of a configuration, as illustrated by the inversion effect (Valentine, 1988; Yin, 1969), and a composition effect (Young, Hellawell, & Hay, 1987). A method of facial synthesis, where a large set of facial parameters are being covaried simultaneously, converges with this notion that configuration is important for conveying expressive information. But covariation of different stimulus aspects may create problems of its own as we will discuss later.

The present study considers four emotions that are generally taken to be basic (anger, fear, happiness, and sadness) arranged in three continua (angry-sad, happy-sad, and happy-afraid). Indeed, it has been argued that these four emotions are those which can best be defined along three separate dimensions of positive-negative, approach-avoidance, and level of arousal (Frijda, 1986). These emotions also appear to be independent of the commonly used verbal labels (Ekman, Friesen, & Ellsworth, 1972) and might be interpreted as reflecting the Darwinian view of the basis of emotion in action patterns (Frijda, 1986). Concentrating on these emotions we chose to omit other candidates, such as neutral, as it is not clear how a neutral expression relates to perception of emotions in faces (in production terms, this is the resting state). The issue of what constitutes a basic emotion, as contrasted with complex or composite emotions, is a vexed one (see Frijda, 1986, for an overview of various ways of defining basic emotions). It is to be expected that present taxonomies of even so-called basic emotions may still represent a heterogeneous set. For example, anger and fear, the two most explored emotions, are very different with respect to what the perception of these facial expressions means for the perceiver. The study of CP for emotions is in its infancy and these are among the crucial issues that future research will have to address.

Etcott and Magee (1992) observed CP of facial expressions using line-drawings. In accordance with the standard formulation of CP, subjects showed dichotomous identification behaviour and discriminated better between stimuli on different sides of the category boundary, than within the same category for the same physical distance between the members of a pair. The present study pursues the issue of CP by using real photographs and a sophisticated digital image morphing technique to create intermediate stimuli between two prototypical facial expressions (Benson & Perrett, 1991). Experiment 2 presents the same expression continua used in Experiment 1, but using inverted presentation. Experiment 3 investigates the existence of CP in children but with the inclusion of a task, seeking a goodness rating for the stimuli from adults as well as from children.

EXPERIMENT 1

The discrimination and identification of facial emotions from photographs of faces were examined in adult subjects for evidence of categorical perception. Three continua, each obtained through a morphing procedure between two posed prototypes, were used. As explained in the Introduction, the most generally agreed manifestation of categorical perception is the occurrence of a peak in discrimination performance around the point on the continuum at which identification reaches 50%. One problem in devising a test of that notion is the existence of individual differences in the location of that point. The solution we adopted consisted of reducing the discrimination data of each subject to two values, one corresponding to the predicted peak and the other to regions of the continuum on either side of the peak. The same form of analysis was applied to reaction times which were measured in both the discrimination and the identification task.

Method

Subjects

A total of 24 undergraduate students from Tilburg University, 12 of each sex, aged 18–26, were paid for participation in one experimental session.

Materials

Three sets of 11 computer-generated continuous tone black and white photographs of faces were used (see Appendix). Each set constituted a continuum of facial expressions between two natural posed exemplars taken from standardised material (Ekman & Friesen, 1976). The continua extended respectively from angry to sad, from happy to afraid, and from happy to sad. The original photographs were digitised and intermediate faces were created using a morphing program (Benson & Perrett, 1991).

The morph transformation started with delineation of the digitised images of the two original images. Landmarks were placed manually on corresponding critical points of the two faces, using a graphical tool. The transformation from one extreme expression to the other was effected in nine intermediate steps. The transformation affected both the feature configuration (warping) and the tonal information (skin texture or pigmentation). The weighted blending procedure was based on linear interpolation between point-to-point pixel intensity values.

Procedure

Subjects were tested individually in a sound-attenuated cabin in the laboratory. A monitor (Commodore 369SX at 256 grey levels) was placed at a distance of 1.5 metres from the subject, and the 9.5cm × 6.5cm photographs subtended a 3.6 × 2.5 degree visual angle. Each subject worked successively on the three sets of photographs, in counterbalanced order. For each set, two tasks were administered: (1) the ABX discrimination task; then (2) the identification task.

In the ABX task, each trial began with an auditory warning signal, followed after 800msec by successive presentation of three photographs for 1sec each, separated by 1sec intervals. The first two pictures, A and B, were always different, and the third, X, was identical either to A or B. The subject was instructed to indicate which picture, A or B, picture X was identical to by pressing one of two buttons, labelled A and B, with a finger of either the left or right hand. Reaction time (RT) was measured from the appearance of the X stimulus to the response. The warning signal of the next trial appeared 2sec after the subject's response. A and B were always two steps apart on the continuum, so nine comparisons were possible. For each comparison, the four possible orders of presentation (ABA, ABB, BAA, BAB) were each presented three times. The resulting 108 trials were presented in random order. The task began with 10 practice trials with faces showing several expressions.

In the identification task, the 11 photographs of the set were presented one-by-one, and the subject identified each stimulus by pressing one of two buttons, with labels bearing the Dutch names of the two end-emotions of the continuum. A trial was announced by the same auditory warning signal, and was followed after 800msec by the face which remained for 1sec. RT was measured from the appearance of the stimulus to the start of the response. The warning signal of the next trial appeared 2sec after the subjects' response. Each of the 11 stimuli was presented three times in random order, giving a total of 33 trials.

Results

The total identification function (percent responses) for each continuum is shown in Fig. 1 (left). All three curves have the usual sigmoidal shape. For each continuum, the identification responses of each subject were submitted to a logit transformation (Finney, 1964), which provided estimations of the 50% point and of the slope of the identification function at that point. The means of these two variables are given for each continuum in Table 1.

The ABX discrimination data were compared subject-by-subject to the prediction from the identification data, following the principles described in the Introduction. Given that the AB points used in the ABX task were two steps apart, each subject's 50% identification point falls into two successive AB intervals, and the peak discrimination is supposed to fall into one of these. For instance, for a subject whose 50% point is 4.2, the peak must occur in one of the two intervals; 3–5 and 4–6. We chose to

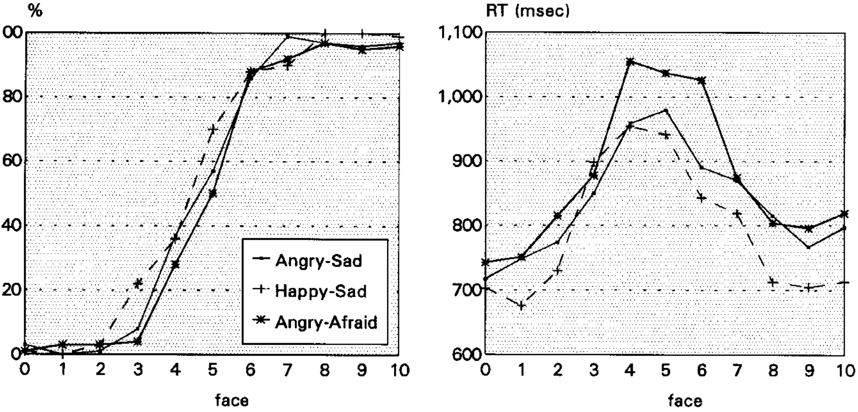

FIG. 1. Experiment 1 (upright faces). *Left*: Mean response percentages for the 3 continua in the identification task (% ''sad'' responses in the angry-sad and happy-sad continua, % ''afraid'' responses in the angry-afraid continuum). *Right*: Mean RTs for the 3 continua in the identification task.

consider the two intervals as containing the predicted peak. Our test consisted in calculating, for each subject, two measures of discrimination performance: (1) a ‘‘peak performance’’ value, which is the mean of the observed percentage of correct responses over the two peak intervals; and (2) a ‘‘nonpeak performance’’ value, which is the mean of the same percentage correct over the seven remaining intervals. Significance was assessed using the *t*-test

The results for each continuum appear in Table 1. The prediction of higher performance in the peak intervals is supported for each continuum. However, it can be noted that discrimination is definitely above chance in the nonpeak intervals, contrary to predictions of early formulations of categorical perception (Repp, 1984, but see Harnad, 1987). Given our decision to test the discrimination data on a subject-by-subject basis, traditional figures showing mean performance across subjects for each whole continuum, similar to those presented by Etcoff and Magee (1992), will not be provided, because they do not show the critical contrast on which the present analysis is based.

RTs in both the identification task and the ABX task were submitted to subject-by-subject analysis following the same principle as the correct discrimination data. For identification, the prediction for categorical perception was a slower RT for the two photographs on either side of the 50% point (the ‘‘peak RTs’’) than for the other items. The results that appear in Table 2 support the prediction for all three continua.

For the ABX task, the prediction was shorter mean RTs for the two ‘‘peak’’ intervals than for ‘‘non-peak’’ ones; this was not supported for any of the three continua.

TABLE 1
Experiment 1 (Upright Faces): Identification and Discrimination Results

	Angry-Sad	*Happy-Sad*	*Angry-Afraid*
Identification			
50% point	4.65	4.42	4.94
Slope	61.28	62.24	57.12
Discrimination			
Peak	68.90	73.50	81.70
Nonpeak	64.00	68.00	76.00
t-test	2.28	2.28	3.27
df	23.00	23.00	23.00
2-tailed *P*	< .032	< .032	< .003

TABLE 2
Experiment 1 (Upright Faces): Mean Identification RTs for 2 Photographs Around the 50% Identification Point and All Other Points

	Angry-Sad	*Happy-Sad*	*Angry-Afraid*
50% point	942.00	982.00	1043.00
Rest	809.00	748.00	837.00
t-test	5.12	7.31	6.57
df	23.00	23.00	23.00
2-tailed *P*	< .001	< .001	< .001

Discussion

Two aspects of the results seem to indicate some degree of categorical processing of the material we have been using. The first aspect is higher discrimination performance in the ABX task for pairs of items falling across the 50% identification point. The second is slower identification RTs for items bordering the 50% point than for items further away. Each of these effects was tested on predictions for individual subjects and was found significant for each of our three continua.

Before drawing any firm conclusions from these results, one must consider the possible role of alternative and noncategorical factors. The morphing procedure produces equal physical steps between adjacent items on each continuum, at least for those points on the faces to which it was applied. But equal physical distances can give rise to unequal perceptual distances, and it was suggested that one should measure the latter using one of the many available psychophysical scaling methods. The problem, however, is that category boundaries are factors that can create distortions between physical and perceived distance. How to separate the contribution of hypothetical noncategorical factors from categorical factors (to avoid throwing the baby out with the bathwater) is not immediately apparent. On the other hand, for noncategorical variations in psychophysical step sizes to produce the present results, larger steps should happen in order to coincide with the 50% identification point of the majority of subjects on each of our three continua, and this is rather unlikely.

A potentially more serious possibility is that more accurate discrimination towards the middle of the continuum, combined with slower identification, could happen on any continuum which, like the present continua or those explored by Etcoff and Magee (1992), extends between exemplars of only two categories. In this situation, only two responses are proposed in the identification task and are probably available to the subject for short-term retention in the discrimination task. The problem does not arise for other domains in which CP has been studied with continua extending to several categories, like the /ba/-/da/-/ga/ continuum for speech (Miller, 1994), or the hue-wheel for colour.

It appeared to us that a useful control of the possible influence of the dichotomisation artefact might be to run the same two tasks with the same material presented upside down. Inversion has frequently been used in research on face perception as a control for the role of nonface-specific properties of the material (see Valentine, 1988, for a review). In the present context it not only could shed light on the dichotomisation issue, but also meet a remaining concern with that of step sizes.

EXPERIMENT 2

For the reasons explained in the discussion of Experiment 1, a new group of adult subjects performed the same tasks, ABX discrimination and identification, but the same material was presented inverted.

Method

A total of 12 undergraduate subjects from Tilburg University, 6 of each sex, participated in one session. They were paid for their participation and none had been exposed to our material before. In the happy-sad continuum one subject had a 50% identification point that fell out of the 1–9 range, and was excluded from further analyses.

The materials and procedure were the same as in Experiment 1, the only difference being that the pictures were presented inverted throughout.

Results

The total identification function (percent responses) for each continuum is given in Fig. 2 (left). It appears that the curves have more or less the customary sigmoidal shape for the happy-sad and angry-afraid continua, although with a shallower slope than in Experiment 1. But for the angry-sad continuum, the curve is approximately bell-shaped. Examination of individual data shows that about half the subjects have a negative slope, indicating a tendency to call inverted angry faces sad and vice versa. Of course, these observations preclude further analyses of the data concerning this particular set of pictures, which would require valid estimations of the category boundary.

For the remaining sets of photographs, happy-sad and angry-afraid, the ABX data were submitted to the same analysis as in Experiment 1. The results are shown in Table 3. For neither of the two continua is there a significant difference between discrimination performance on peak and nonpeak comparisons, respectively. However, overall discrimination performance is above chance level for the two continua.

Identification RTs for the same two continua were similar to those in Experiment 1, analysed according to proximity to the estimated category boundary. The results in Table 4 are similar to those in Experiment 1: RT is slowed down for the photographs bordering the 50% point, compared to the other points. With 2-tailed *t*-tests, the effect is significant for the happy-sad continuum, and fails just short of the .05 level for angry-afraid. But one

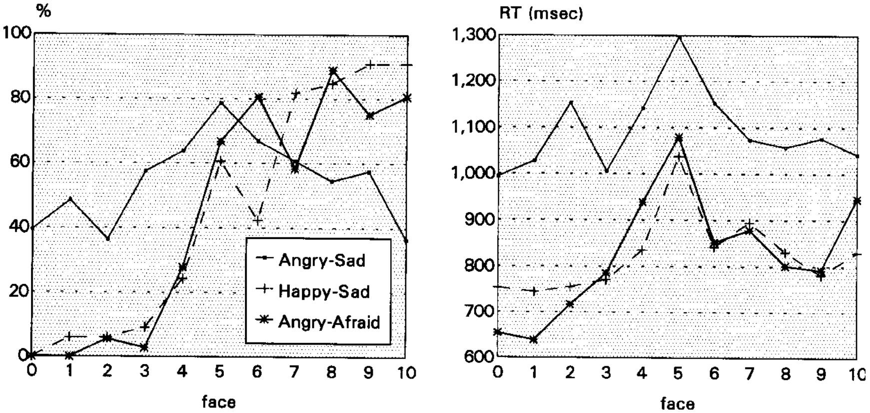

FIG. 2. Experiment 2 (inverted faces). *Left*: Mean response percentages for the 3 continua in the identification task (% ‘‘sad’’ responses in the angry-sad and happy-sad continua, % ‘‘afraid’’ responses in the angry-afraid continuum). *Right*: Mean RTs for the 3 continua in the identification task.

TABLE 3
Experiment 2 (Inverted Faces): Identification and Discrimination Results

	Angry-Sad	*Happy-Sad*	*Angry-Afraid*
Identification			
50% point	–	5.58	5.72
Slope	–	43.64	40.07
Discrimination			
Peak	–	62.60	76.30
Nonpeak	–	61.50	70.30
t-test	–	0.30	1.41
df	–	10.00	11.00
2-tailed *P*	–	< .773	< .185

TABLE 4
Experiment 2 (Inverted Faces): Mean Identification RTs for 2 Photographs Around the 50% Identification Point and All Other Points

	Angry-Sad	*Happy-Sad*	*Angry-Afraid*
50% point	–	910.00	930.00
Rest	–	806.00	802.00
t-test	–	2.77	2.13
df	–	10.00	11.00
2-tailed *P*	–	< .020	< .056

could argue that as the difference was predicted, a 1-tailed *t*-test should be applied. In this case, the difference would be significant at $P < .028$.

For the ABX task, there was again no difference in speed between comparisons across and those on one side of the boundary.

Discussion

For the angry-sad photographs, the subjects had an erratic identification performance, which made analysis of other data concerning these stimuli impossible. Apparently, any decision as to whether these pictures are perceived as belonging to the end-points associated with either of the two emotions becomes inaccessible because of inversion. For the other two sets of photographs, which were used in Experiment 1, contrasting results were obtained for ABX discrimination performance and identification RTs, respectively.

The discrimination performance peak on comparisons straddling the category boundary is no longer observed. Thus, the notion of a dichotomy artefact, suggesting that such a peak should occur toward the middle of any continuum extending between two different exemplars, is not supported and the positive result obtained in Experiment 1 cannot be explained by that particular mechanism. The present finding, it may be noted, would also create difficulties for an interpretation based on psychophysically unequal steps: How inversion of the pictures would reinstate equality is not evident.

However, in the identification task, the slower RTs around the category boundary are again observed. Thus, this particular effect is not specific to emotion processing and may result from a dichotomy artefact or some other general mechanism.

Expression information and the inversion paradigm have not, to our knowledge, been used to explore the issue of the configural basis of facial expressions. Configurational information is more critical for making face judgements than for judging other perceptual objects (Diamond & Carey, 1986). Because inversion drastically reduces the CP effect, we can infer that the information relevant to CP might be carried by facial configuration. The issue of relative versus absolute inaccessibility of expressive information following inversion clearly requires further exploration.

Experiment 3 was set up to generalise the findings to young children, and a new task was also added. Studies by Miller and collaborators have shown systematic variations in goodness judgements within the same category and have argued for a graded internal structure of phonetic categories (see Miller 1994, for overview and discussion). Miller and colleagues have also found that this prototype structure is reflected in RTs, in the sense that the higher the goodness ratings, the less time it takes listeners to identify a stimulus. To what extent might this situation carry over to the domain of facial expressions? In our case, this would mean that two different expressions, both identified as ''anger'', can still be clearly discriminated. The reason for this good within-category discrimination could be that one stimulus is perceived as a better expression of anger than another. Note that goodness of category membership may matter comparatively less for speech stimuli, where categorical decisions are only one step in a whole process of word recognition. In contrast, a judgement about goodness of emotion expression is likely to affect perception and action much more directly.

EXPERIMENT 3

The main objective of this experiment was to investigate if the evidence for categorical perception obtained in Experiment 1 (with adults) could be replicated with children. Because one could expect children to have lower discrimination capacity than adults, the inter-item space for the ABX task was set at three steps. In consequence, new adult subjects had to be tested for comparison purposes. Another difference in Experiment 1 was that the subjects were asked to rate the photographs' goodness as exemplars of their category.

Method

Subjects

A total of 24 9- to 10-year old children, 12 of each sex, from a public school in Tilburg, were tested individually in a quiet room at their school; 59 undergraduate students, aged 17–21, participated in a short session each, 24 with the angry-sad continuum, 18 with happy-sad, and 17 with angry-afraid. No subjects had seen the material before.

Procedure

For each continuum, three tasks were administered successively: the ABX discrimination task, the identification task, and a goodness rating task. The ABX task was run with three-step intervals instead of two. Apart from that, the administration procedure for both the ABX task and the identification task was the same as in Experiment 1.

For the goodness rating task, each set of photographs was divided into two subsets, one containing the first six items and the other one the last six items. The pictures were presented on the screen under the same conditions as in the identification task. The subjects expressed their ratings on an answer sheet that indicated the expression to be rated and gave a 10-point scale. The instructions specified that a rating of 1 meant no expression of the relevant emotion and a rating of 10 an excellent expression.

Results

The total identification functions of children and of adults are given for each continuum in Fig. 3. The children's functions have somewhat shallower slopes than those of adults, which is confirmed by the values shown in Table 5. The adults' functions are comparable to those obtained in Experiment 1.

In the ABX task, because three-step intervals were used, each subject's 50% identification point falls in three successive intervals. Therefore, the predicted peak discrimination is measured by the mean accuracy over these three intervals, and the nonpeak discrimination by the mean over the remaining five intervals. The data from one child could not be used in the analysis of ABX performance on the happy-sad and angry-afraid continua, because his 50% identification points fell outside the 2–8 range. The results are shown in Table 5.

Before considering the peak-nonpeak contrast, it can be seen that the performance of the children on nonpeak discrimination is lower than that of adults. The latter is, as expected, superior to that observed in Experiment 1 for two-step discrimination. In the children, peak interval superiority is significant for happy-sad and angry-afraid continua. On the angry-sad continuum, the peak-nonpeak difference is in the expected direction but falls short of significance. It can be noted that the angry-sad continuum is where nonpeak discrimination is lowest, not only in the children but also in adults of both this experiment and Experiment 1. Thus, a floor effect may have prevented the manifestation of the peak-nonpeak difference in some of the children. In the adults, the pattern of statistical significance is the opposite of that of the children. The peak interval superiority is significant for angry-sad, but not for happy-sad or for angry-afraid. For the last two

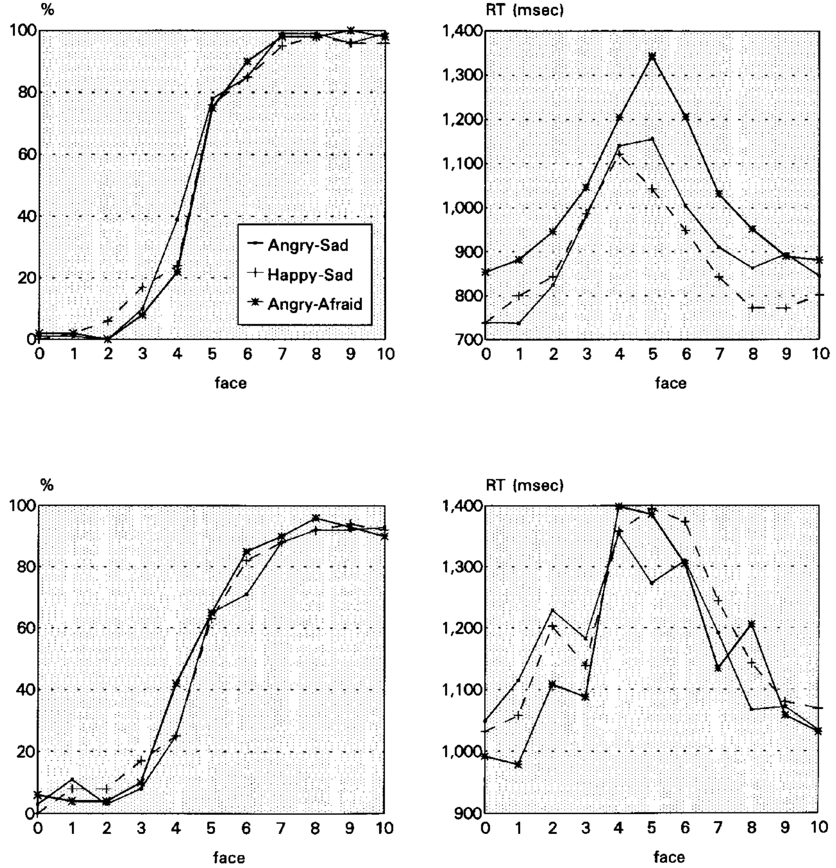

FIG. 3. Experiment 3 (upright faces). *Adults*: *Top left*: Mean response percentages for the 3 continua in the identification task (% ''sad'' responses in the angry-sad and happy-sad continua, % ''afraid'' responses in the angry-afraid continuum). *Top right*: Mean RTs for the 3 continua in the identification task. *Children*: *Bottom left*: Mean response percentages. *Bottom right*: Mean RTs.

continua, nonpeak discrimination is at a high level, 85% and 91%, respectively. Therefore, one can consider the possibility of a ceiling effect. Finally, it can be noted that in the case of a happy-sad application of a 1-tailed test, which the existence of a prediction would justify, would provide a probability level of .054, very close to significance.

In identification RTs (Table 6) the significant slowing down obtained in other experiments is again observed for each continuum, in both the children and adults. No peak-nonpeak difference is obtained for discrimination RT.

TABLE 5
Experiment 3 (Upright Faces): Identification and Discrimination Results

	Angry-Sad	*Happy-Sad*	*Angry-Afraid*
A. Adults			
Identification			
50% point	4.42	4.47	4.56
Slope	63.56	63.48	68.23
Discrimination			
Peak	81.10	88.00	90.60
Nonpeak	74.60	85.09	90.90
t-test	3.16	1.70	− 0.28
df	23.00	17.00	16.00
2-tailed *P*	< .004	< .108	< .783
B. Children			
Identification			
50% point	4.92	4.77	4.67
Slope	51.28	55.54	54.17
Discrimination			
Peak	63.10	69.70	82.60
Nonpeak	60.80	64.30	78.40
t-test	0.98	2.42	2.89
df	23.00	22.00	22.00
2-tailed *P*	< .335	< .024	< .008

TABLE 6
Experiment 3 (Upright Faces): Mean Identification RTs for 2 Photographs Around the 50% Identification Point and All Other Points

	Angry-Sad	*Happy-Sad*	*Angry-Afraid*
A. Adults			
50% point	1164.00	1122.00	1282.00
Rest	866.00	825.00	964.00
t-test	5.91	5.88	4.99
df	23.00	17.00	16.00
2-tailed *P*	< .001	< .001	< .001
B. Children			
50% point	1295.00	1404.00	1352.00
Rest	1143.00	1143.00	1109.00
t-test	3.55	6.27	5.31
df	23.00	23.00	23.00
2-tailed *P*	< .002	< .001	< .001

The results of the goodness rating task appear in Fig. 4. In all categories the mean rating increases monotonically with closeness to the prototype. This pattern is observed in both the children and adults. As a comparison of Figs. 3 and 4 indicates, the rating data correlate strongly with the identification RTs for the same photographs, in both the children and adults. Of 12 product-moment correlations (2 groups × 3 continua × 2 prototypes) 10 are significant at $P = .05$. The only nonsignificant correlations are those for sadness on the angry-sad continuum for adults ($r = .76$), and for anger on the same continuum for children ($r = .79$).

Discussion

Experiment 3 was devised in order to investigate whether evidence for categorical perception obtained in Experiment 1 would be replicated with children under the age of 10 years. The main criterion for categorical processing, a peak in discrimination performance around the category boundary, was observed for all three continua, but failed to reach significance for the more difficult continuum, angry-sad. It might be the case that a floor effect played a role in this case. As a comparison, adults were tested on the same tasks: The peak superiority effect was only partially replicated in this easier discrimination task, and typically with the two most difficult continua. The absence of effect with the easier angry-afraid material may be partly due to a ceiling effect. Indeed, one can speculate whether the resort to adult control subjects was a sound decision, given that absolute performance level is of little relevance to the main theoretical question. The direct comparison of the children of Experiment 3 with the adults of Experiment 1 at the level of performance profile is probably most instructive. The profile of discrimination performance of the two groups is remarkably similar and allows the same conclusions.

Experiment 3 shows that children categorise facial expressions in a very similar way to adults. The only difference concerns speed of response, a variable that is known to increase with age. Facial expression perception as examined here does not seem to be affected by developmental differences in the processing style of younger children, such as the ‘‘developmental dip’’ (Carey, 1981).

Finally, adults as well as children provide clear goodness ratings for all the expressions studied. Rated goodness decreases as distance from each natural unmanipulated expression increases. Also, there appears to be a systematic relationship between RTs in the identification task and stimulus ratings. Further research is needed to explore this relationship and its importance for understanding the discrimination results.

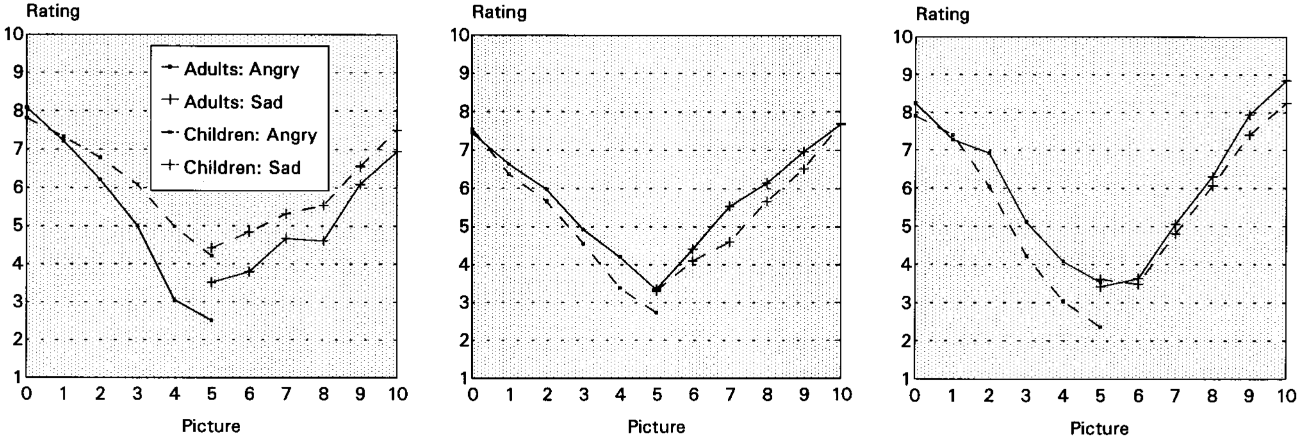

FIG. 4. Experiment 3 (upright faces): Ratings for angry-sad (*left*), happy-sad (*centre*), and angry-afraid (*right*). *Left*: Mean response percentages for the 3 continua in the identification task (% ''sad'' responses in the angry-sad and happy-sad continua, % ''afraid'' responses in the angry-afraid continuum). *Centre*: Mean RTs for the 3 continua in the identification task.

GENERAL DISCUSSION

This study addressed the question of whether facial expressions are perceived categorically, and if so, whether this is the case for adults as well as for children. Combining the different results from the three tasks and the three experiments, it appears that, generally speaking, the facial expressions studied here are clearly perceived categorically by adults as well as by children. At the same time, subjects appear to judge accurately whether a given facial expression is a good or a poor example of the facial expression represented. We comment on these results and discuss whether inferences about underlying categories of facial expression perception can be based on these behavioural results.

The possibility that categorisation behaviour results from learning was first argued for by Lane (1965). A learning explanation stresses the conceptual as opposed to the perceptual status of categories and focuses on the role of labels and their role in memory to explain categoricity. As understood in this context, learning explanations tend to reduce the power of inferences about underlying biological functional categories. Artificial neural networks using a back-propagation learning rule have been used to simulate CP effects (Harnad, Hanson, & Lubin, 1994). Indeed, such networks can be configured to exhibit categorical behaviour very much in the manner of human and animal performance. These implementations are not a substitute for human subjects because, currently, they cannot be directly related to high-level categorical behaviour. Networks operate heuristically and algorithmically and can be sensitivity-tuned to yield categorical perception (CP), but they are not capable of operating at the cognitive level as a human subject might during an expression task. Although examination of hidden unit weights at the learning stage may give some indication of the magnitude and consistency of just noticeable differences, one cannot assume these will reflect conscious or unconscious human discrimination grounded in terms of category symbols or categories in the mind.

We have already noted that the possibility of CP findings in the identification task might be based on task-specific factors. Note that a methodology where bipolar continua are combined and the anchor point A of a continuum a-b is also that of another continuum a-c, etc, does not present a full answer to the objection of possible task artefacts. In the present study, Experiment 2 addresses that issue satisfactorily and serves as a control for the effect of such factors. Improving CP methodology would require a study of categories in a multidemensional space. It has been illustrated that just noticeable differences play a less significant role when more than two categories are present, as in a high-dimensional expression space (each expression represents one dimension: Benson, 1992, 1995). Facial expressions are highly interactive in nature, and the presence of blended emotions is well known (Ekman & Friesen, 1976; Katsikitis & Benson, submitted; Nummenmaa, 1988). Only in a high-dimensional space can such phenomena be properly assessed (Benson, 1995).

There are two important and potentially confounding factors in similarity assessments and ABX discriminanda for CP research and are especially relevant to a discussion of higher-order stimuli interpretations. First, a given facial expression may be signalled by only a few facial features. How features interact, is clearly an important consideration; only a portion of the images may be useful—but which? Similarity judgements of CP effects along a happy-angry continuum are more likely to depend on image-based (congruent) features, whereas ABX discrimination will draw on different salient features. Second, the strength of the affect signal, or its ‘‘energy’’ (Benson, 1995; Katsikitis & Benson, submitted), provides a simple and clear indication that some (category) exemplars may not be as representative or as good members as others. This is particularly evident in Etcoff and Magee (1992) and may help to account for the failure to find categoricity of surprise (Benson, 1995). It should be clear that these two factors (feature salience, energy) are intrinsically related and therefore one should expect to observe modulations in the magnitude, or absence, of CP effects dependent on category verisimilitude (Benson, 1995). Prototypical or enhanced expressions may better qualify for category membership, the prediction being that optimal representations will yield optimal CP.

In the present study, as in earlier studies using the CP paradigm, subjects are good at discriminating between two stimuli that belong to the same category. This finding is compatible with CP conclusions and it has been argued (Harnad, 1987) that there are no good grounds for expecting discrimination ability within category pairs to be at chance level. Nevertheless, the issue of the basis of good discrimination ability is worth pursuing because of its importance for understanding facial expression perception, whether or not this is category-based. Examining the goodness ratings and their possible relation to this discrimination ability may be a first step.

We have already mentioned that a situation comparable to a first stage of physical information processing may not obtain, in the case of facial expression perception. Like sounds, faces are complex stimuli, but unlike the former, faces are likely to be perceived configurationally rather than componentially with perception being built-up from separate features. A two-stage model of facial expression perception in which better than expected discrimination would be due to continued post-categorical presence of low level differences is thus not very likely. Better than chance discrimination might result from the fact that subjects while assigning various stimuli to the same category, nevertheless note critical differences in the extent to which these different same-category stimuli represent the expression typical for that category. Two interesting hypotheses for future research are that goodness judgements are predominantly related to perceived quality of expression of the stimulus as a whole, and that subjects do not base their response on separate components of the face only.

As noted in the Introduction, very little is known about perceptual expression categories, and claims concerning a functional architecture and the emotional categories which comprise it will require converging evidence, at least in developmental neuropsychology. The present study presents such developmental data. These allow us to make a comparison with results obtained from a group of autistic subjects. In this population, deficits in the perception of facial expressions have been documented. Thus, the prediction is that an impairment would also be observed with the CP task. This was indeed the case (Teunisse & de Gelder, submitted). The autistic subjects showed poorer identification and discrimination. On further investigation of the subjects, interesting within-group differences emerge. Poor performance on the facial expression task is predicted by a low score on communication-skill tests, but not by general intelligence or verbal intelligence skills. This suggests that poor performance in facial expression tests is not a matter of general learning skills or of ease of manipulating verbal labels.

Some final caveats. The question whether facial emotions are perceived categorically is not so much interesting in itself, but rather for its potential relevance to discussions on the locus of emotion perception in the cognitive architecture of the organism. Neither our present data nor previously obtained results allow inferences that generalise beyond the perception of emotions in facial expressions. Issues of cognitive or functional primitives underlying the perception of facial expressions may be connected to the notion of natural categories or prototypes. Such inferences raise interesting questions on the domain-specificity of the facial expression perception mechanisms. Are those emotions presented in voices through speech prosody, equally perceived categorically? And if so, does one and the same mechanism sustain perception in the two modalities? Recent evidence (de Gelder, Vroomen, & Teunisse, 1995) suggests that this might be the case because there is a clear cross-modal effect from one modality to the other. There is increasing evidence that the perception of some emotional states is subserved by specialised cognitive and neurological circuitry (Adolphs, Tranel, Damasio, & Damasio, 1994). From this perspective, it is tempting to consider basic emotions, at least the cognitive ability underlying the perception of such, as consisting of a limited set of expressive primitives and to view CP results as providing crucial evidence. For example, Ekman (1994) argues that findings of categorical perception of facial expressions supports the notion of basic emotions that would be expressive universals, invariant across cultures. Current candidates for membership of the class of basic emotions must however be considered each on their own; and findings supporting categoricity of an emotion will need to be integrated with neurobiological, neuropsychological, and developmental data before inferences about expressive primitives can be drawn.

Manuscript received 26 May 1995
Revised manuscript received 13 December 1995

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.

Benson, P.J. (1992). Unpublished doctoral dissertation, University of St. Andrews, Scotland.

Benson, P.J. (1995). Perspectives on face perception: directing research by exploiting emergent prototypes. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition* (pp. 204–224). London: Routledge.

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

Carey, S. (1981). The development of face perception. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 9–38). London: Academic Press.

de Gelder, B., Vroomen, J., & Teunisse, J.P. (1995, November). *Hearing smiles and seeing cries: The bimodal perception of emotion*. Paper presented at the 36th Annual Meeting of The Psychonomic Society, Los Angeles.

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

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). Measuring facial movement. *Journal of Environmental Psychology and Nonverbal Behavior*, *1*, 56–75.

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

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

Finney, D.J. (1964). *Probit analyses*. Cambridge: Cambridge University Press.

Frijda, N. (1986). *The emotions*. New York: Cambridge University Press.

Harnad, S. (Ed.) (1987). *Categorical perception*. New York: Cambridge University Press.

Harnad, S., Hanson, S.J., & Lubin, J. (1994). Learned categorical perception in neural nets: implications for symbol grounding. In V. Honavar & L. Uhr (Eds.), *Symbol processors and connectionist network models in artificial intelligence and cognitive modelling: Steps toward principled integration* (pp. 191–206). London: Academic Press.

Katsikitis, M., & Benson, P.J. (submitted). *Classification of FACS-instructed facial expressions*.

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review*, *72*, 275–309.

Liberman, A.M. (1996). *Categorical perception*. Cambridge, MA: MIT Press.

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *53*, 385–368.

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

Nummenmaa, T. (1988). The recognition of pure and blended facial expressions of emotion from still photographs. *Scandinavian Journal of Psychology*, *29*, 33–47.

Repp, B.H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in Basic Research and Practise*, *10*, 234–335.

Teunisse, J.P., & de Gelder, B. (submitted). *Impaired categorical perception of facial expressions in high-functioning autistics*.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*, 471–491.

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

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

APPENDIX

Stimuli: Expression Continua for Angry-Sad, Happy-Sad, and Angry-Fear