A Label Superiority Effect in Children's Categorization of Facial Expressions

James A. Russell and Sherri C. Widen, *University of British Columbia*

Abstract

In three studies (N's = 360, 68, 160), children (2 to 7 years of age) were asked to categorize various facial expressions. The emotion category was specified to the child by its label (such as happy*), its facial expression (such as a smile), or both. From the youngest to the oldest children and for all 3 emotion categories examined (happiness, anger, and sadness), results showed a Label Superiority Effect: emotion labels resulted in more accurate categorization than did the corresponding facial expression. Errors conformed to a structural model emphasizing the dimension of pleasure-displeasure.*

By the 1980s, psychologists largely accepted as a fundamental axiom that human beings possess an emotion signaling system: certain facial expressions signal specific categories of discrete 'basic' emotions. Inspired by the writings of Charles Darwin (1872/1965), researchers such as Izard (1971), Ekman (1972), and Tomkins (1962) developed a set of related theories, methods, and evidence that together have been called the 'Facial Expression Program' (Russell & Fernandez Dols, 1997, p. 4).

This program included and stimulated much developmental research. For example, Harris (1989) offered an account of the development of children's knowledge of mind as growing out of their knowledge of emotion, and their knowledge of emotion as growing out of their ability to recognize the emotional meaning of facial signals. Harris cited evidence that the ability to recognize specific categories of emotion from facial and vocal expressions is universal and develops quite early. Some psychologists speculated that young infants recognize a half dozen or so specific emotions from facial expressions (Bowlby, 1969; Field, Woodson, Greenberg, & Cohen, 1983; Field et al., 1983; Izard, 1971; but see Kaitz & Meschulach-Sarfaty, 1988; Nelson & Dolgin, 1985). Such early recognition suggests that the infant might not have to acquire a system of discrete mental categories for emotions because that system is already available. The ability to recognize discrete emotions from facial expressions, in turn, was theorized to pave the way for infant-caregiver attachment (Bowlby, 1969, 1988; Magai & McFadden, 1995), social referencing (Klinnert, Emde, Butterfield, & Campos, 1988), empathy (Thompson, 1998), and further understanding of emotion and mind (Harris, 1989).

James A. Russell, Department of Psychology, Boston College, McGuinn Hall, Chestnut Hill, MA 02467, USA. E-mail: *james.russell@bc.edu*

However plausible, the assumptions of the Facial Expression Program must be subjected to empirical scrutiny. There are several reasons to take a second look at it. One reason is that the notion of universal signaling and recognition of specific emotions via facial expression is currently in dispute in the adult literature (Carroll & Russell, 1996, 1997; Mandler, 1992; Fridlund, 1994; von Brakel, 1994; Russell, 1994; Fernandez Dols & Ruiz Belda, 1995, 1997; Reisenzein, 2000; Wierzbicka, 1992). Proponents of the Facial Expression Program have replied, in part, by softening some of their claims (Ekman 1994; Izard, 1994). Russell and Fernandez Dols (1997) summarized fourteen assumptions of the Facial Expression Program, and so testing the entire Program is beyond the scope of this or any other single study, even if one focuses only on the Program's developmental aspects. The present article is a small part of the reexamination of that Program.

The topic of this article is the nature of the mental category evoked in a preschooler by a facial expression. We assume that the mental category evoked is the meaning that the child recognizes in the facial expression. Much evidence shows that children, including infants, respond to facial expressions in a systematic manner (Caron, Caron & MacLean, 1988; Walker-Andrews & Lennon, 1991). With infants, systematic responses to facial expressions have been found with imitation (Field et al., 1982), habituation tasks (Serrano, Iglesias, & Loeches, 1992; Barrera & Maurer, 1981; Maurer & Barrera, 1981), looking time (Serrano, Iglesias, & Loeches, 1995), and social referencing (Klinnert et al., 1988). Still, questions remain about the precise meaning that children find in facial expressions—that is, about the nature of children's mental categories for emotion.

Here we examine two specific and related propositions suggested by the Facial Expression Program: (a) that a child's mental category for a basic emotion is equivalent to the adult's category and (b) that the facial expression for that emotion is an especially powerful means to evoke the correct mental category in the child. We consider each topic in turn, beginning with the second.

If recognition of anger, sadness, and other specific categories of emotion from facial expressions is 'easy' (Ekman, 1972; Izard, 1971), is present early, and is the foundation for the acquisition of other information (Denham, 1998; Harris, 1989), then a facial expression should be an excellent way to evoke the appropriate mental category for the specific emotion signaled. Conversely, if the emotion's label, such as *angry* or *sad*, is arbitrary in the sense that it varies from one language to the next, must be acquired, is acquired later than is the ability to recognize the corresponding facial expression, and is acquired in part on the basis of that facial expression, then the label should not be as good a way to evoke the category. Thus, at least during a major portion of this period of development, the Facial Expression Program leads one to anticipate a Face Superiority Effect: accurate categorization of emotion is easier on the basis of faces than labels. 'Easier' might be made operational in response latency, the likelihood of errors, or other standard measures. A Label Superiority Effect would be just the opposite: accurate categorization is easier on the basis of emotion labels.

Evidence against a Face Superiority Effect was the unexpected result of a study by Camras and Allison (1985). Children ranging in age from preschool to second grade were told very brief stories about a fictitious teenager (e.g., 'her mother has died'), and were asked what emotion the teenager was feeling. The child's answers were gathered with one of two formats: a choice among emotion labels (e.g. *happy*, *angry*, and *sad*) or a choice among still photographs of facial expressions of the same emotions (in this example, photographs showing a teenager smiling, frowning, or crying). As

expected, the children did well in choosing the appropriate emotion for each story. What was not expected was the complete lack of evidence for a Face Superiority Effect. Indeed, for some emotions, the children did better when their options were presented to them as labels than when presented as facial expressions.

Russell (1990) and Russell and Widen (in press) pursued Camras and Allison's intriguing finding. They, too, looked at emotion stories, but turned the task around by providing preschoolers (3 to 5 years of age) with the emotion (presented as a face, a label, or both) and asking them to complete the story, telling what caused the emotion and what happened next. For no emotion did showing a facial expression produce more accurate results than did giving the corresponding label. Just the reverse—the superiority of label over face was statistically significant overall, and clearest for fear and disgust.

The evidence provided by these three story-based studies (Camras & Allison, 1985; Russell, 1990; Russell & Widen, in press) leaves the status of the Label Superiority Effect unclear on at least three counts. First, the superiority of labels over faces was clear with only two emotions (fear and disgust). Second, the youngest children studied so far were preschoolers (age unspecified) in Camras and Allison's study, 4-year-olds in Russell's study, and 3-year-olds in Russell and Widen's study; perhaps the superiority of label over face is reversed for younger or older children. Indeed, much evidence makes a reversal at younger ages highly plausible and the empirical question is the age at which it takes place. Finally, all three studies included stories; perhaps the superiority of label over face depends on this narrative aspect of the task given the children. Faces are visually processed, whereas both emotion labels and stories obviously involve language.

The present studies used a non-narrative task in which emotion categories were presented to the children as a box into which people who feel only a certain way could go. The target emotion was specified by an emotion label (*happy*), by the corresponding unlabeled facial expression (smile), or by the combination of label and facial expression. In other words, the child was either *told* who goes in the box, or *shown* who goes in the box (or both). The child's task was to decide whether each of a series of people with various facial expressions (test faces) should go into the box or be left out. This task minimized the verbal demands on the children and thus minimized any advantage the label mode of presentation may have enjoyed in previous story-based studies. Indeed, when the target emotion was specified by a face, the child could solve the task by perceptually matching this defining face with the test face; any methodological advantage would seem to favor a Face Superiority Effect. The first study focused on the emotion categories of happiness and anger. Participants ranged in age from 2 to 7 years in order to examine the possibility that any advantage of labels or faces will change or disappear with age. A second study focused on sex differences, added a box for sadness, and made other small changes, including the use of other test faces. The third study focused on toddlers (2 and 3 years of age) and again examined the categories of happiness and anger.

The second issue that these three studies allowed us to address was the nature and development of children's mental categories of emotion. The writings we cited as illustrating developmental aspects of the Facial Expression Program implicitly presupposed that basic emotion categories in children are identical to those in adults. That is, to assert that a child recognizes anger is to assume that the child recognizes the same thing that an adult recognizes when the adult recognizes anger. Although children may make more errors than do adults because of limitations in attention, memory,

vocabulary, factual information, and the like, the underlying mental categories are assumed to be the same.

The present studies are part of a larger project exploring an alternative possibility: that children's mental categories for emotion systematically change over the course of development. Bullock and Russell (1984, 1985, 1986; Russell & Bullock, 1986a, 1986b) found the traditional assumption insufficient in understanding how children make sense of the emotions they witness (and, by inference, experience). A fine-grained analysis of young children's 'errors' found that their actual emotion categories undergo large developmental changes. Therefore, on Bullock and Russell's view, children's 'errors' are not always errors at all, but simply differences from an adult standard. On this account, children's initial mental categories for emotion are broad. The youngest children they examined (24 to 36 months) divided all emotions into two broad categories, the pleasant and the unpleasant. This model derives from and predicts evidence that differentiation on the basis of valence (pleasant–unpleasant) is an easy task for the young child. Over the course of development, differentiation occurs within these two broad categories, initially based on the degree of arousal and later on antecedent situations and resultant behaviors, including specifics of facial movement. For example, the mental category that includes anger initially also includes disgust, fear, and sadness as well. Over time, these four are differentiated from each other. Anger is differentiated first from sadness, then from fear, and finally from disgust. Although not the same as an adult's, this system enables children to interpret facial expressions of emotion meaningfully. These findings were replicated with a sample of German (Bormann-Kischkel, Hildebrand-Pascher, & Stegbauer, 1990) and deaf children (Hosie, Gray, Russell, Scott, & Hunter, 1998).

Study 1

The child's task was to decide which facial expressions should be included in each emotion category—portrayed to the child as a box. The experimenter began, 'I have a box [pointing]. See? This is a special box. In this box goes . . .' The emotion category was specified in one of three ways: a label, a facial expression, or both. The child was then shown, one at a time, a series of test facial expressions (each posed by a different woman and different from the woman who posed for the photograph used to define the box/category). The Face Superiority Effect predicts that defining the box/category with a facial expression will result in fewer errors than defining the box with the corresponding label. The Label Superiority Effect predicts the opposite. Bullock and Russell's (1986) model predicts which 'errors' are more likely to occur.

Method

(i) Participants. Children (N = 360) were recruited from urban daycares, preschools, and public elementary schools. All were proficient in English. There were 60 children at each age level, 2 through 7 years, defined by age at last birthday (174 girls, 186 boys). The mean (and standard deviation) for 2-year-olds was 32 months (2.92), for 3-year-olds 43 months (3.21), for 4-year-olds 53 months (3.67), for 5-year-olds 65 months (3.59), for 6-year-olds 77 months (3.09), and for 7-year-olds 89 months (3.54). Sex was not considered when selecting participants or assigning them to a condition and was therefore not evenly divided across age level or experimental condition. For

example, of the 5-year-olds in the label-and-face condition, 14 happened to be girls, 6 boys. All other cells were more evenly divided between sexes.

(ii) Photographs of facial expression. The photographs were $5'' \times 7''$ black and white glossy prints of facial expressions of basic emotions taken from Ekman and Friesen's (1976) Pictures of Facial Affect. Each photograph was of a different adult woman. This set of photographs was designed to provide clear, unambiguous, prototypical examples of facial expressions each of a single emotion near maximum intensity. For example, the 'happy' faces are all beaming smiles.

(iii) Design and procedure. Each child participated in three trials. The first concerned the color green, the second the emotion of happiness, and the third the emotion of anger. The happy trial preceded the angry trial because we wanted to minimize the chances of the younger children believing they could not do the task. The experimenter spent the first visit to the daycare getting to know each child (the first visit was omitted for children in elementary school). On a second visit, the experimenter began the actual study by inviting an individual child to help with some 'work.' Children were randomly assigned to one of three modes of presentation (label-only, face-only, or label-and-face).

   (a) Label-only mode. In the label-only condition, the target box/category was defined by a label. That label was *green* on the first (practice) trial, *happy* on the second trial, and *angry* on the third trial. Every effort was made to say the labels *happy* and *angry* in a neutral tone of voice, with the same intonation as the label *green* was said in the first trial. On the first trial, the experimenter showed the child an empty box and explained that this box was for green cards and only green cards could go in the box. The child was then asked to help sort seven colored test cards (3 shades of green, and one each of fuchsia, red, orange, and yellow). As each test card was presented, the experimenter asked, 'Is this card green?' If the child indicated that the card was green, the card was placed in the box. If the child indicated that it was not green, it was placed outside the box. The child was encouraged to place the card in or out of the box him- or herself. Errors were corrected. Most answers were correct on the first presentation.
   On the second trial, a new box was introduced and the experimenter explained that the new box was for happy people, and only happy people could go in the box. The experimenter then showed the child six test photographs, one at a time in random order. There were two happy, two angry, one disgust, and one fear expression. (For half of the subjects, a sad expression was substituted for disgust.) For each test expression, the experimenter asked, 'Is this person happy?' Errors were not corrected; all answers were mildly praised.
   On the third trial, the procedure was the same as for the second trial, except that the label *angry* was substituted for *happy*. The same six test photographs were presented in a new random order.

   (b) Face-only mode. The face-only condition was identical to the label-only condition, except that, for all three trials, the box/category was defined with a visual example. Thus, on the practice trial, instead of using the label *green*, the experimenter showed the child a green card and explained that the box was for 'this color' [point-

ing] and only cards the 'same color as this one' [pointing] could go in the box. The card used to define the green category was close, but not identical, to the green test cards. The defining card was placed upright just above and behind the box so that it was fully visible to the child. The seven test cards were the same as those used in the label-only condition. For each test card, the experimenter asked, 'Is this card [pointing] the same color as that one [pointing]?'

The second trial was identical to its counterpart in label-only condition except that, instead of using the label *happy*, the experimenter showed the child a photograph of (a woman posing) a happy facial expression, and explained that this box was for people who 'feel like this [pointing] and only people who feel like this [pointing] can go into the box.' The defining photograph was again placed upright just above and behind the box. For each test face, the experimenter asked, 'Does this person [pointing] feel the same as that one [pointing]?'

The third trial was identical except that the photograph defining the box was an angry expression.

(c) Label-and-face mode. This third mode combined the two preceding conditions. Thus, the experimenter told the child that the box was for 'green cards, like this one [pointing to a green card] . . . feels happy, like this person feels [pointing to a happy photograph] . . . feels angry, like this person feels [pointing to an angry photograph].'

Table 1. Proportion of Faces Included in Happy and Angry Boxes in Study 1

	Mode of Presentation			
Test Face	Label-Only	Label-&-Face	Face-Only	Mean
Happy Trial				
Happy	$.98_a$	$.99_a$	$.85_h$	.94
Angry	$.04_b$	$.02_b$	$.03_b$	.03
Fear	$.08_b$	$.03_b$	$.05_b$	.05
Disgust	$.05_b$	$.02_b$	$.04_b$	.04
Sad	$.08_b$	$.03_b$	$.04_b$	.05
Angry Trial				
Angry	$.95_a$	$.86_h$	$.63_c$	.81
Disgust	$.95_a$	$.85_h$	$.58_f$	.79
Sad	$.33_e$	$.28_e$	$.15_g$	.25
Fear	$.54_d$	$.56_d$	$.35_e$	.48
Happy	$.02_b$	$.02_b$	$.03_b$	.02
Mean	.41	.38	.29	

Note: Maximum possible is 1. Least Significant Difference comparisons (alpha = .05) were calculated on the cell means. Means in the same row that do not share subscripts differ at $p < .01$. Means in the same column that do not share subscripts differ at $p < .02$.

Results and Discussion

Ten younger children were omitted from the study when they failed to perform adequately on the practice (color) trial. Thus, before beginning the happy trial, all children whose data were included in the study had demonstrated a good understanding of the procedure. Of the 360 children included, only 11 made any errors at all on the practice trial. For the emotion trials, 86.0% of all responses were 'correct.'

Label Superiority Effect

The first analysis focused on the question of which mode of presentation better evoked the correct emotion category. Responses to the disgust, fear, and sad test faces were omitted (a) to balance the numbers of inclusion and exclusion errors (see later), and (b) to use an unambiguous criterion of correctness (from our perspective, including disgust, fear, or sad in the angry box is less an error than including them in the happy box; analyses of these responses are reported later). Thus, for the happy trial, each child's proportion correct was the number of happy faces placed in the box plus the number of angry faces left out of the box, divided by 4. For the angry trial, proportion correct was the number of angry expressions placed in the box plus the number of happy expressions left out of it, divided by 4. In a repeated measures ANOVA (alpha = .05), mode of presentation (label-only, face-only, and label-and-face) and age (6 levels) were between-subjects factors; emotion box (happy, angry) was a within-subject factor. Proportion correct was the dependent variable.

The main effect of mode was significant, $F(2, 342) = 34.92$, $p < .001$. The proportion correct in the face-only mode (.86) was significantly lower ($p < .001$) than in the label-only (.97) and in the label-and-face (.95) mode.[1] The advantage of label over face replicated at every age (Figure 1). The age × mode interaction was not significant, $F(10, 342) = 1.06$, $p = .39$.
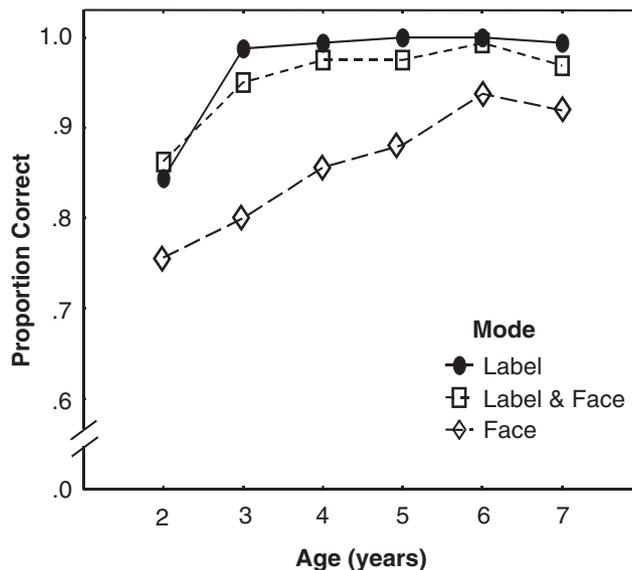

Figure 1. Proportion of correct responses with three modes of presentation in Study 1.

The advantage of label over face replicated for both happy and angry boxes, although the mode × emotion box interaction was significant, $F(2, 342) = 15.17$, $p < .001$. Least Significant Differences (LSD) comparisons indicated that, on the happy trial, the proportion correct in the face-only mode (.91) was significantly lower ($p < .001$) than in the label-only (.97) and in the label-and-face (.99) mode, which did not differ significantly from each other. On the angry trial, the proportion correct in the face-only mode (.80) was also significantly lower ($p < .001$) than in the label-only (.97) and in the label-and-face (.92) mode, but, on this trial, the latter two conditions differed significantly ($p = .001$). So, two sources of information in the label-and-face mode offered no advantage over the label-only mode in children's categorization in either box (and, indeed, might have been a disadvantage in the angry box).

Nature of Emotion Categories

The second analysis concerned Bullock and Russell's (1986) model, specifically the prediction that certain errors are more common than others. Both the Facial Expression Program and Bullock and Russell's models predict that the happy box/category includes only happy test faces and excludes all others (this decision relies on valence and should be easy); thus, errors are randomly spread across faces. Similarly, both models predict that the angry box/category includes angry test faces and excludes happy faces (again this decision relies on valence); errors are again random. The contrast between the two models occurs with disgust, fear, and sad faces in the angry box. For the angry box, these discriminations cannot be made on the basis of valence (these decisions are thus more difficult): On Bullock and Russell's models, more 'errors' are expected here than in other conditions. A second prediction of their model was that inclusion of disgust, fear, and sad faces in the angry box decreases with age, with sad faces being excluded first, followed by fear, then disgust. Bullock and Russell's model thus predicts a main effect for type of error and an age × error-type interaction.

Figures 2 and 3 show what was included in the happy and in the angry box (the data were collapsed across the three modes of presentation). These results were analyzed here, however, in terms of *errors*. A test of the main prediction examined five types of errors: (1) leaving a happy face out of the happy box; (2) leaving an angry face out of the angry box; (3) placing an angry face in the happy box or a happy face in the angry box; (4) placing a disgust, fear, or sad face in the happy box; and (5) placing a disgust, fear, or sad face in the angry box. To calculate the proportion of errors of type 1, for example, the number of times a child left a happy face out of the happy box was divided by the number of happy faces (two). The procedure was the same for the remaining four types of errors. In a repeated measures ANOVA (alpha = .05), age (6 levels) was the between-subjects factor and type-of-error (5 levels) was a within-subject factor. Proportion of errors was the dependent variable.

The main effect for type-of-error was significant, $F(4, 1416) = 199.60$, $p < .001$. As expected, Error-type 5—the proportion of disgust, fear, and sad faces put in the angry box (.50)—was significantly greater ($p < .001$, LSD) than each of the other four error-types (range: .02 to .18). Unexpectedly, error-type 2—the proportion of angry faces left out of the angry box (.18)—was significantly greater ($p < .001$) than each of the remaining three error-types, which did not differ significantly from each other.[2]

To investigate how children's category of anger changes with age, we examined each of the specific errors for the angry box separately. There were five possible types of error for the angry box: (1) leaving an angry face out; (2) including a disgust face; (3)
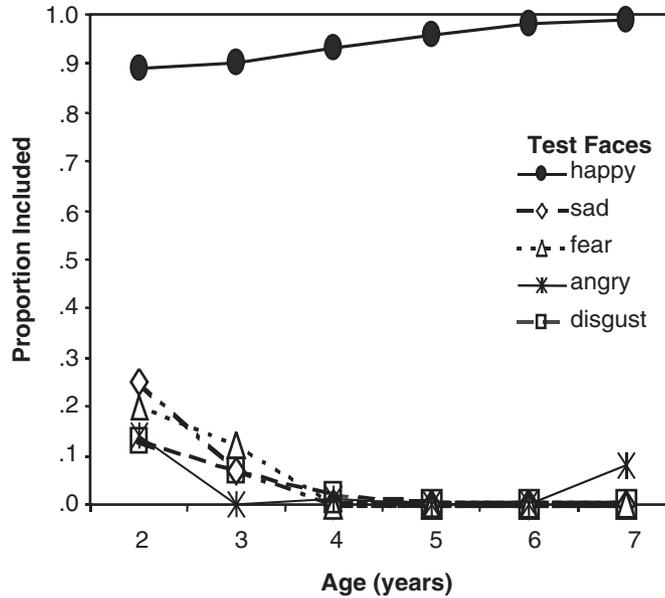
Figure 2. Proportion of each type of facial expression included in the happy box in Study 1.
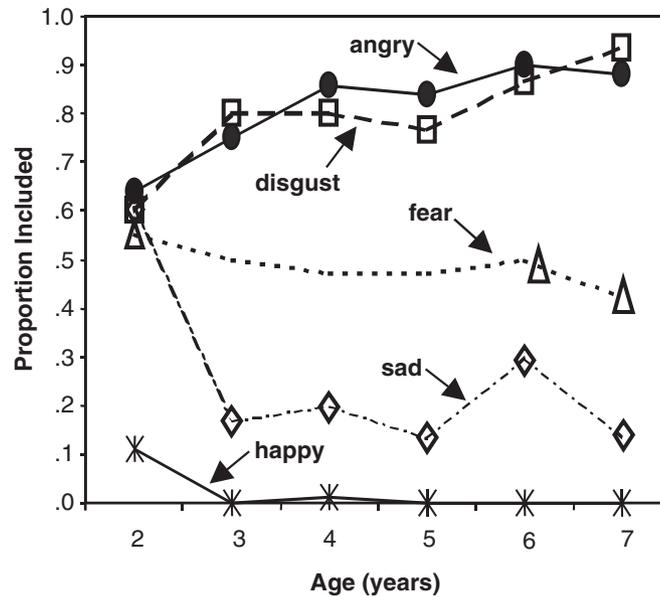

Figure 3. Proportion of each type of facial expression included in the angry box in Study 1.

including a fear face; (4) including a sad face; and (5) including a happy face. In a repeated measures ANOVA (alpha = .05), age (6 levels) was the between-subjects factor; type-of-error (5 levels) was the within-subject factor. Proportion of errors was the dependent variable. The main effect for type of error was significant, $F(4, 1044) = 47.24$, $p < .001$. Each error type was significantly different ($p < .03$) from each of the

others. Including the disgust face (.79) was most frequent, followed by including the fear face (.48), including the sad face (.26), excluding the angry face (.18), and finally, including the happy face (.02). With the exception of the difference between angry and happy faces, this pattern conformed to that predicted by Bullock and Russell (1986).

The predicted age × error-type interaction was also significant, $F(20, 1044) = 3.02$, $p < .001$, but the pattern of results was different from what we had anticipated (Figure 3). The 2-year-olds excluded happy faces from the angry box, but were equally likely to include anger, disgust, fear, and sad faces. This result was as expected. What was unexpected was what then happened with increasing age. As predicted, the proportion of sad faces included in the angry box decreased significantly ($p = .001$, in LSD comparisons) with age. However, the proportion of fear faces included did not change significantly with age, and the proportion of disgust faces actually increased significantly ($p = .02$) with age. Indeed, the proportion of disgust faces included did not differ significantly from the proportion of angry faces. Thus, the narrowing of children's anger category was far from complete even by 7 years of age. Put differently, whereas the 2-year-olds showed near-categorical perception for valence (including negative and excluding positive faces from the angry box), even the 7-year-olds failed to show categorical perception of a discrete category of anger.

Effect of Mode on Inclusion of Faces

The results also showed another unanticipated effect: mode of presentation affected not only the number of errors, but also the likelihood of inclusion of any face (Table 1). In a repeated measures ANOVA (alpha = .05), age (6 levels) and mode (3 levels) were between-subjects factors; and test facial expression (10 levels; i.e., 2 trials × 5 types of facial expression per trial, irrespective of whether it was correct or incorrect) was the within-subject factor. The likelihood of the facial expression being included in the box was the dependent variable. The main effect for mode was significant, $F(2, 324) = 26.82$, $p < .001$, indicating that defining the box by a label (label-only and label-and-face) resulted in the inclusion of more faces overall than defining the box by a face alone. Mode also interacted significantly with test face, $F(14, 2268) = 2.04$, $p = .01$, an effect due to the Label Superiority Effect. Mode did not interact significantly with age $F(10, 342) = 1.46$, $p = .15$, indicating that the label mode had a similar effect for all ages, even the 2-year-olds.

Thus, when the category/box was specified by a face alone, it was not merely the case that children made more errors, but they also included fewer exemplars overall. This effect cannot be interpreted solely in terms of errors or quality of performance, because including fewer exemplars results in more exclusion errors but fewer inclusion errors. Rather, this finding raises the possibility that the category evoked by a label might be broader than the 'same' category evoked by a face, even if both ostensibly concern the same emotion. It might thus be necessary to distinguish children's label-categories of emotion from their facial-expression-categories for the same emotion. Put differently, at the ages included in this study, children might not possess a single mental category for happiness or anger, but rather a category is formed on the spot to suit current circumstances.

Study 2

Study 2 served three primary purposes. The first was to examine the possibility of sex differences, which were not considered in Study 1 and were a potential but unlikely

source of variance. Studies with adult samples have found that women are generally better at recognizing emotions in others than are men (Toner & Gates, 1985). However, research done with children has been equivocal. Some studies found no sex differences in children's abilities to recognize emotions (Boyatzis & Satyaprasad, 1994), whereas others found small differences (Boyatzis & Chazan, 1993). Studies of the Label Superiority Effect found a similar mixed result: Russell (1990) found that girls were significantly better than boys overall; and Russell and Widen (in press) found that sex interacted with mode of presentation; but Camras and Allison (1985) found no sex differences.

The second purpose was to bolster the face-only mode of presentation by using two facial expressions of the same emotion to specify the target emotion. Our aim was to help the children in the face-only condition extract the emotion category common to the two facial expressions shown and thus define the box by the relevant emotion category.

The third purpose was to investigate further the Label Superiority Effect with a new emotion and across other minor changes in method. A sad box/category was added, resulting in a total of three emotion trials (happy, angry, sad). Prior to beginning the trials, the experimenter initiated a short conversation about emotions, in order to prime the child's emotion concepts and maximize the likelihood that the children's performance on the task would be the result of the independent variables rather than due to individual differences in children's abilities to access emotion concepts. In the pre-test that followed, children were taught the task by sorting animal and non-animal toys into a box rather than by sorting color cards. The purpose was to provide a practice trial in which the child would have to notice many features of the stimulus, thereby more closely approximating the emotion trials. This pre-test was identical in both modes of presentation. For the emotion trials, the label-and-face mode of presentation was omitted from Study 2 because it had not significantly improved performance over the label-only condition in Study 1. (Indeed, it curiously showed lower performance than the label-only mode in the anger box.) On each trial, fewer test faces (four) were presented to the child. The disgust faces were dropped altogether. Not all categories of emotions were shown in each trial: Happy and angry expressions were the test faces for happy and angry trials. Sad and fear expressions were the test faces in the sad trial.

Method

The method was identical to that of Study 1, except as noted.

(i) Participants. Sixty-eight preschoolers (34 boys and 34 girls) participated in the study. The majority of the participants (71%) were 4 years old, with a range from 3 years, 11 months to 5 years, 7 months; mean age was 54 months (standard deviation = 5.75). None had participated in Study 1.

(ii) Photographs of facial expressions. The sixteen photographs of women posing facial expressions (six happy, six angry; four sad, two fear) used in the current study were from Ekman and Friesen (1976). One of the happy faces, two of the angry faces and one of the sad faces had been used in Study 1; the twelve other photographs were new to this study.

(iii) Design and procedure. Each child participated in four trials: the first concerned animals, the second the category happy, the third the category angry, and the fourth the category sad.

Table 2. Proportion of Faces Included in the Happy, Angry, and Sad Boxes in Study 2

	Mode of Presentation		
Test Face	Label-Only	Face-Only	Mean
Happy Trial			
Happy	.98a	.90a	.94
Angry	.04b	.18d	.11
Angry Trial			
Angry	.96a	.81ef	.89
Happy	.04b	.18d	.11
Sad Trial			
Sad	.90a	.75f	.83
Fear	.35c	.44c	.40
Mean	.55	.54	

Note: Maximum possible is 1. Least Significant Difference comparisons (alpha = .05) were calculated on the means. Means in the same row that do not share subscripts differ at $p < .04$. Means in the same column that do not share subscripts differ at $p < .02$.

The experimenter spent the first visit to the daycare getting to know each child individually. On the second visit, the first trial (animals) was the same for both modes of presentation and was designed to teach the children the task. Each of six toys (three animals, three automobiles) was presented one at a time to the child, who was asked, 'Is this toy an animal? Does it go into the box?' If a toy was misidentified, the error was corrected and the toy was returned to the pile of as yet unsorted toys to be presented again later. The task continued until all the toys were correctly sorted.

Next, the experimenter introduced the topic of *feelings*, and spent a little while asking questions about when and how the child may have had different feelings (happy, angry, sad, and fear), in order to prime the child's emotion concepts. At this point children were randomly assigned to one of two modes of presentation (label-only, face-only) with the proviso of equal numbers of boys and girls in each mode. The face-only mode was identical to that used in Study 1, except that the box was defined with two visual examples.

Results and Discussion

Four children were screened from the study due to poor performance on the practice (animal) trial; these were replaced with same-sex children. Of the 68 children whose data were included in this study, only five made any errors at all on the practice trial. Overall, 84.2% of all responses on the three emotion trials were correct.

(iv) Label superiority effect. The first analysis focused on the question of which mode of presentation better evoked the correct emotion category. To balance the number of inclusion and exclusion errors, and to focus on unequivocal errors, only the data from the happy and angry trials were analyzed. In a repeated measures ANOVA, mode of presentation (label-only, face-only) and sex (two levels) were between-subjects factors; box (happy, angry) was the within-subject factor. Proportion correct was the dependent variable.

The main effect for mode was significant, $F(1, 64) = 9.72$, $p = .003$. The proportion correct in the face-only mode (.84) was significantly lower than in the label-only mode (.96).[3] This result replicated the advantage of label over face observed in Study 1. There were no significant effects involving sex and no other significant effects with alpha set at .05. The absence of any significant effects involving sex suggests that sex was not a factor in the results of Study 1.

We repeated this analysis with all three emotion boxes, and the main effect for mode was again significant, $F(1, 64) = 11.64$, $p = .001$. The proportion correct in the face-only mode (.78) was significantly lower than in the label-only mode (.90). The advantage of the label mode occurred for all three boxes, and the mode × emotion interaction was not significant, $p = .88$. There were no significant main or interaction effects involving sex.

(v) Nature of emotion categories. We next tested Bullock and Russell's (1986) prediction that certain errors are more likely than others. The happy and angry boxes/categories were predicted (other than random errors) to include only the target faces and to exclude the others. In contrast, the sad box was predicted to include not only the target (sad) faces, but, to a lesser degree, fear faces as well.

There were five types of errors: (1) leaving a happy face out of the happy box; (2) leaving an angry face out of the angry box; (3) leaving a sad face out of the sad box; (4) placing an angry face in the happy box or a happy face in the angry box; (5) placing a fear face in the sad box. In a repeated measures ANOVA (alpha = .05), sex (2 levels) was the between-subjects factor and type-of-error (5 levels) was a within-subject factor. Proportion of errors was the dependent variable.

The main effect for type-of-error was significant, $F(4, 264) = 16.54$, $p < .001$.[4] As expected, error-type 5—the proportion of fear faces in the sad box (.40)—was significantly greater ($p < .001$, LSD comparisons) than each of the other four types of error (range: .06 to .18). Unexpectedly, error-type 3—leaving the sad faces out of the sad box (.18)—was significantly higher ($p = .01$) than the proportion of error-type 1. There were no other significant differences between the error-types. There were no significant effects involving sex.

(vi) Effect of mode on inclusion of faces. Our final analysis investigated the effect of mode on the sheer inclusion of test faces (an effect observed in Study 1). In a repeated measures ANOVA (alpha = .05), mode of presentation (2 levels) and sex (2 levels) were between-subjects factors; and test facial expression (6 levels; i.e., 3 trials × 2 types of facial expression per trial) was the within-subject factor. The likelihood of a facial expression being included was the dependent measure. The main effect for mode was not significant ($p = .90$), thus failing to replicate the under-inclusion effect in the face-only mode from Study 1. The mode × face interaction was significant, $F(5, 320) = 4.47$, $p = .001$, and the pattern of results showed the Label Superiority Effect (Table 2).

—

Study 3

Our aim in Study 3 was to focus on toddlers (2- and 3-year-olds). Age groups were narrowed to six months in order to trace more closely the developmental changes in children's understanding and categorization of emotions between their second and fourth birthdays. We focused on these ages because in Study 1, 2-year-olds' results were significantly lower than any other age group, but 3-year-olds' results did not differ significantly from 4- and 5-year-olds'. Thus, the big transition in performance seems to occur between 2 and 4 years of age. In addition, the effect of mode had been weaker (albeit still significant) in 2-year-olds than in 3-year-olds.

Method

Study 3 was identical to Study 2, except as noted. Study 3 had two emotion trials (happy, angry) and focused on age differences. During the happy trial, only happy and angry test faces were shown. During the angry trial, only angry, fear, and sad test faces were shown.

(vii) Participants. One hundred and sixty children (80 girls, 80 boys) participated in this study. There were 40 children (20 girls and 20 boys) in each of four age groups: young 2s (24–29 months; mean = 28, s.d. = 2.16), older 2s (30–35 months; mean = 33, s.d. = 1.29), young 3s (36–41 months; mean = 40, s.d. = 2.01), older 3s (42–47 months; mean = 45, s.d. = 1.26). None had participated in either Study 1 or 2.

Table 3. Proportion of Faces Included in Happy and Angry Boxes in Study 3

	Mode of Presentation		
Test Face	Label-Only	Face-Only	Mean
Happy Trial			
Happy	.92$_a$	.91$_a$	.98
Angry	.24$_b$	.39$_c$	.32
Angry Trial			
Angry	.87$_a$	.88$_a$	.87
Fear	.64$_d$	.69$_d$	.67
Sad	.41$_c$	.59$_d$	.50
Mean	.62	.69	

Note: Maximum possible is 1. Least Significant Difference comparisons (alpha = .05) were calculated on the means. Means in the same row that do not share subscripts differ at $p < .002$. Means in the same column that do not share a subscript differ at $p < .03$.

(viii) Photographs of facial expressions. Fifteen photographs (5 Happy, 6 Angry, 2 Sad, 2 Fear) were used, each of a different adult woman. Seven of these photographs were introduced for this study; five had been used in both Studies 1 and 2; three had been used in Study 2 but not in Study 1.

(ix) Procedure. Each child participated in three trials. The first concerned animals, the second happiness, and the third anger. On the happy trial, the child was shown three happy[5] and two angry test faces. On the angry trial, the child was shown two angry (different from those used in the happy trial), two sad, and two fear test faces.

Results and Discussion

Five children who performed poorly on the first (practice) trial were omitted from the study and were replaced with same-sex age-mates. Of the 160 children whose data were included, only 16 made any errors on the practice trial. Overall, 69.6% of the responses on the emotion trials were correct.

(i) Label superiority effect. To examine which mode of presentation better evoked the correct emotion category, with unequivocal errors, data from the happy trial were analyzed. In a repeated measures ANOVA, mode of presentation (label-only, face-only), age (4 levels), and sex (two levels) were between-subjects factors. Proportion correct was the dependent variable. The main effect of mode was significant, $F(1, 144) = 5.65$, $p = .02$. The proportion correct was significantly lower in the face-only mode (.79) than in the label-only mode (.86). The advantage of label over face replicated at every age (Figure 4). The age × mode interaction was not significant, $F(3, 144) = .48$, ns.

The main effect of age showed steady improvement in children's categorizing of facial expressions between their second and fourth birthdays (Figure 4). LSD com-
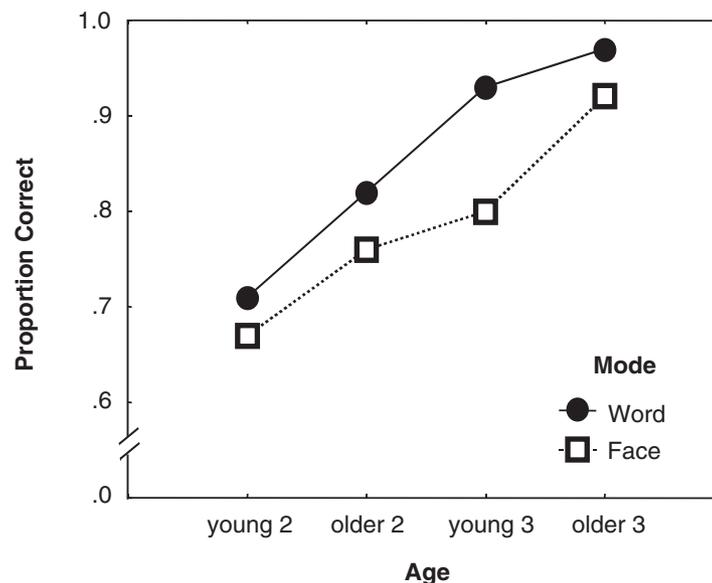

Figure 4. Proportion of correct responses with two modes of presentation in Study 3.

parisons showed that the young 2s' proportion correct (.69) was significantly lower (*p* < .02) than each of the other age group's; the older 2s (.79) did not differ significantly from the young 3s (.87), but were significantly lower (*p* < .001) than the older 3s (.95); the young 3s did not differ significantly from the older 3s.

We repeated this analysis with both emotion boxes, and the main effect for mode was again significant, $F(1, 144) = 10.16$, $p = .002$. The proportion correct in the face-only mode (.66) was significantly lower than in the label-only mode (.73). The advantage of the label mode occurred for both boxes, and the mode × emotion interaction was not significant, $p = .94$. There were no significant main or interaction effects involving sex.

(ii) Nature of emotion categories. We next tested Bullock and Russell's (1986) prediction that certain errors are more likely than others. Specifically, the most common type of error was predicted to be the inclusion of sad and fear faces in the angry box. There were five types of error: (1) leaving a happy face out of the happy box, (2) including an angry face in the happy box, (3) leaving an angry face out of the angry box, (4) including a fear face in the angry box, and (5) including a sad face in the angry box. In a repeated-measures ANOVA with alpha set at .05, mode of presentation (word, face), age (four levels), and sex of the child were between-subjects factors. The within-subject factor was the five types of error defined above.

The main effect of type of error was significant, $F(4, 576) = 77.37$, $p < .001$. As predicted, error-types 4 and 5—proportion of fear faces (.67) and sad faces (.50) included in the angry box—were significantly greater (*p* < .001, LSD comparisons) than the other types of errors. The proportion of angry faces included in the happy box (.32) was higher than expected. The two exclusion errors (leaving correct faces out of the box) were low (happy = .08; angry = .13) and not significantly different from each other.

Bullock and Russell's (1986) model predicts an interaction between age and type of error, and this interaction was significant, $F(12, 576) = 4.80$, $p < .001$. The pattern of results replicated that of Study 1. Exclusion errors (leaving happy faces out of the happy box or angry faces out of the angry box) did not differ with age: All four age groups, even the youngest, showed a similar proportion of exclusion errors (range: .08 to .15). In contrast, all other types of errors were more common and declined with age (e.g., the proportion of angry faces included in the happy box decreased from .65 to .06). The sharpest decline was in the happy box (Figure 5). There was a systematic, linear decrease with age in the proportion of angry faces included in the happy box; an asymptote was reached by the young 3s, and even the young 2s were significantly (*p* < .001) more likely to include happy than angry faces in the happy box. For the angry box (Figure 6), in contrast, young 2s included roughly equal (and high) proportions of anger, fear, and sad faces. The older 2s were significantly (*p* < .03) more likely to include an angry face than a fear or sad one in the angry box. Between their second and fourth birthdays, the anger category narrowed, with sad faces almost dropping out by children's fourth birthday, and fear faces declining, but not dropping out in this age range.

All the effects described so far must be re-examined in light of the three-way age × mode × type-of-error interaction, $F(12, 576) = 1.90$, $p = .03$. First, for the happy box, the three-way interaction was consistent with the pattern of results described above. Second, for the angry box, results with the angry and sad faces were consistent with the overall pattern. However, for fear faces there was an anomalous result:
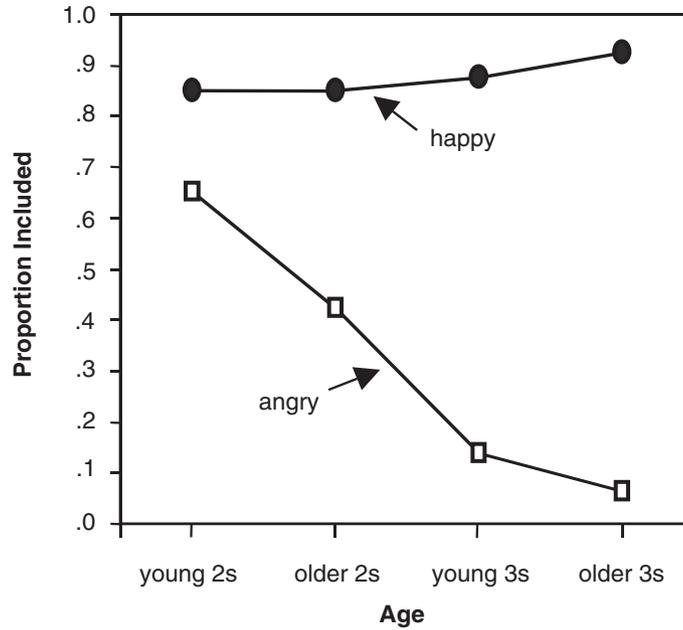
Figure 5. Proportion of each type of facial expression included in the happy box in Study 3.
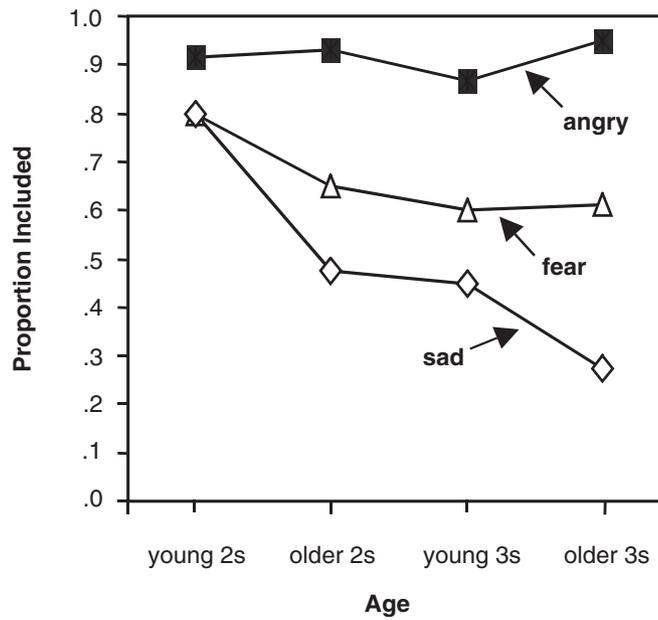

Figure 6. Proportion of each type of facial expression included in the angry box in Study 3.

In the label mode, instead of the usual decline of errors with age, both older and younger 3s made more errors than older 2s.[6]

(iii) Effect of mode on inclusion of faces. Our final analysis investigated the effect of mode on the sheer inclusion of test faces (an effect observed in Study 1). In a repeated measures ANOVA (alpha = .05), mode of presentation (2 levels), age (4 levels), and sex (2 levels) were between-subjects factors; test facial expressions (5 levels; each type of facial expression in each trial) was the within-subject factor. The likelihood of a facial expression being included was the dependent measure. The main effect for mode was significant, $F(1, 144) = 4.51$, $p = .003$, but, unlike the under-inclusion effect in the face-only mode from Study 1, the children here included *more* test faces in the face mode (.69) than in the word mode (.62). The mode × face interaction was also significant, $F(4, 576) = 1.38$, $p = .001$, with the pattern of results showing the Label Superiority Effect: The children included the same number of the target faces but fewer non-target faces in the label mode than in the face mode (Table 3).

General Discussion and Conclusion

In all three studies, even the youngest children were clearly able to sort stimuli into or out of a box—they understood the task. Our confidence in this conclusion stems from the use of a first (practice) trial (color cards or animal toys) as a training and screening device, which ensured that every child who participated in the emotion trials had already successfully done the task. In addition, performance on emotion trials overall was good, a result consistent with the general finding that even 2-year-olds have some understanding of facial expressions and emotion labels.

All three studies supported Bullock and Russell's (1986) model of children's emotion categories. Of course, children made random errors, but they also made more errors in some conditions than in others. That is, errors followed a predictable and systematic pattern. At all ages, children were more likely to include in the box non-target facial expressions of the same valence than non-target facial expressions of opposite valence (although this effect was tested here with negative valence categories only). Indeed, 2-year-olds' emotion categories were so broad, at least on the task we gave them, that it remains to be demonstrated that their categories are more than valence. Of course, no one study and no one task can be definitive, but the present results do add to mounting evidence on the nature of children's emotion categories: they do not coincide with adult categories but change over the course of development into adult categories.

A comparison of Figure 3 (from Study 1) with Figure 6 (from Study 3) yielded a surprising observation. Both figures show which test faces children included in the angry box/category. In both figures, 2-year-olds included anger, fear, and sad faces (but not happy ones) with similar probability. By their fourth birthday, the probability of including a fear face was cut in half and the probability of including a sad face was even lower. These trends are all consistent with Bullock and Russell's (1986) model. What was surprising was that, after 4 years, children showed little evidence of further development toward a discrete anger category up to the oldest age we included, 7 years. Although 4-, 5-, 6-, and 7-year-olds excluded happy expressions, they continued to include sad, fear, disgust, and angry faces, with unchanging (albeit different) probabilities. This apparent lack of development is quite intriguing.

There was also an unexpected result in Study 1: children included fewer faces in the face-only mode than in either the label-only or the label-and-face modes. This

result was not replicated in Study 2 and was reversed in Study 3. Thus, it may be a chance event. On the other hand, it remains possible that categories specified by a *single* face (as in Study 1) differ from those specified by two facial expressions of the same emotion (as in Studies 2 and 3) or by the corresponding emotion label. One possibility is that the children in Study 1 interpreted the single faces used to define the boxes as limited to only certain degrees or only certain types of the target emotion. Rather than 'happiness,' perhaps some children thought 'intense happiness' or 'contentment.' Several considerations speak against this possibility. Our results here and in previous studies indicate that young children tend to form broader rather than narrower categories. For example, as we showed, their category of anger tends to include disgust, fear, and sad faces. The second consideration is that all stimuli to be sorted into or out of the box were facial expressions prepared by Ekman and Friesen (1976). Although posed by different women, the facial expressions themselves consisted of specific muscle movements and were carefully developed to convey a single basic emotion near maximum intensity. The more interesting possibility is that children's emotion categories are not stable or unitary (i.e., not the same when evoked by a face and by a label). Perhaps children create emotion categories in response to the current situation. In any case, if this finding can be replicated, its interpretation would be worthy of further research.

Children's performance in the face-only mode was always above chance and our results are thus generally consistent with prior studies that demonstrated that even young children find meaning in faces. Nevertheless, our three studies found no evidence of a Face Superiority Effect: specifying an emotion category with one or two prototypical facial expressions of a discrete emotion reliably resulted in less accurate performance than did specifying the category with the emotion's label. These results invite further research on the assumption that facial expressions play a key role in the development of children's understanding of emotion. For instance, facial expressions might play an important role, but only in combination with situational, vocal or behavioral cues. Such ecological questions remain surprisingly neglected.

Instead, a Label Superiority Effect was found with all three emotions examined (happiness, anger, and sadness), thus expanding earlier findings of a similar effect for fear and disgust (Camras & Allison, 1985; Russell, 1990; Russell & Widen, in press). In Study 1, two sources of information (label-and-face mode) produced no advantage over the label-only mode (and perhaps a slight disadvantage for anger). The Label Superiority Effect was found with slightly different methods, across both sexes, and across a large range of ages. It remained in evidence even for 7-year-olds, and, perhaps more theoretically important, it was statistically significant for young 2-year-olds. Of course, we cannot extrapolate outside the age range actually covered. For older children, the effect of label and face might converge. For even younger children, the power of a label might well be less, but empirical evidence should replace assumption. Nevertheless, it is clear that by 24 to 29 months, emotion labels are an important part of these children's developing understanding of emotion.

This is the fourth report of a Label Superiority Effect. Three studies using stories as an independent (Camras & Allison, 1985) or a dependent measure (Russell, 1990; Russell & Widen, in press) had found that specifying fear and disgust with labels resulted in significantly better performance than specifying them with the corresponding faces. Of course, only further research can establish whether the Label Superiority Effect occurs for still other methods, or in more ecologically representative contexts. Thus, while it is not possible to say definitively that labels are better than

faces as hooks to children's mental emotion categories, the case for a Label Superiority Effect is mounting.

The finding of a Label Superiority Effect suggests that the role of emotion labels in children's acquisition of emotion knowledge might be greater than generally believed. Investigations of children's language acquisition and concepts has found a comparable result: labeling concrete objects increases children's later attention to the labeled object in 1-year-olds (Baldwin & Markman, 1989), helps 13-month-olds form categories (Waxman & Markow, 1995), and enables 4-year-olds' to group objects on the basis of category membership rather than thematic relations (Gelman & Markman, 1986). Thus, evidence of a Label Superiority Effect in children's understanding of emotion concepts complements evidence regarding the power of labels and labeling of concrete objects.

The current studies do not constitute counter evidence to the more traditional theories advocated by Izard (1971), Harris (1989), Denham (1998), and others based on the Facial Expression Program. Rather, our evidence for a Label Superiority Effect and for Bullock and Russell's (1986) model of children's emotion categories suggests, at least for children beyond their second birthday, that traditional assumptions on the nature of children's emotion categories and on the power of facial expressions to evoke the correct emotion category are incomplete and need to be refined. Ultimately, integration of the traditional accounts with the newer evidence on children's acquisition and understanding of emotion categories will clarify how exactly children of different ages interpret facial expressions and emotion labels and come to form emotion categories that match the adult standard.

References

Baldwin, D. A. & Markman, E. M. (1989). Establishing word-object relations: A first step. *Child Development*, *60*, 381–398.

Barrera, M. E. & Maurer, D. (1981). The perception of facial expressions by the three-month-old. *Child Development*, *52*, 203–206.

Bormann-Kischkel, C., Hildebrand-Pascher, S., & Stegbauer, G. (1990). The development of emotional concepts: A replication with a German sample. *International Journal of Behavioral Development*, *13*, 355–372.

Bowlby, J. (1969). *Attachment and loss: Vol. 1: Attachment* (2nd ed.). London: Hogarth Press.

Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development.* New York, NY: Basicbooks.

Boyatzis, C. J. & Chazan, E. (1993). Preschool children's decoding of facial emotions. *Journal of Genetic Psychology*, *154*, 375–382.

Boyatzis, C. J. & Satyaprasad, C. (1994). Children's facial and gestural decoding and encoding: Relations between skills and with popularity. *Journal of Nonverbal Behavior*, *18*, 37–55.

Bullock, M. & Russell, J. A. (1984). Preschool children's interpretation of facial expressions of emotion. *International Journal of Behavioral Development*, *7*, 193–214.

Bullock, M. & Russell, J. A. (1985). Further evidence on preschooler's interpretations of facial expressions of emotion. *International Journal of Behavioral Development*, *8*, 15–38.

Bullock, M. & Russell, J. A. (1986). Concepts of emotion in developmental psychology. In C. E. Izard & P. B. Read (Eds.), *Measuring emotions in infants and children, Vol. II*, (pp. 203–237). Cambridge: Cambridge University Press.

Camras, L. A. & Allison, K. (1985). Children's understanding of emotional facial expressions and verbal labels. *Journal of Nonverbal Behavior*, *9*, 84–94.

Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional expressions: The role of face and voice. *Child Development*, *59*, 604–616.

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

Carroll, J. M. & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology*, *72*, 164–176.

Darwin, C. (1965). *The expression of the emotions in man and animals.* Chicago: University of Chicago Press. Original work published in 1872.

Denham, S. A. (1998). *Emotional development in young children.* New York: Guilford Press.

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

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

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

Fernandez-Dols, J. M. & Ruiz-Belda, M. A. (1995). Are smiles a sign of happiness? Gold medal winners at the Olympic games. *Journal of Personality and Social Psychology*, *69*, 1113–1119.

Fernandez-Dols, J. M. & Ruiz-Belda, M. A. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression: Studies in emotion and social interaction, 2nd series* (pp. 255–274). New York, NY: Cambridge University Press.

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1983). Discrimination and imitation of facial expressions by neonates. *Annual Progress in Child Psychiatry and Child Development*, 119–125.

Field, T. M., Woodson, R., Cohen, D., Greenberg, R., Garcia, R., & Collins, K. (1983). Discrimination and imitation of facial expressions by term and preterm neonates. *Infant Behavior & Development*, *6*, 485–489.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view.* San Diego: Academic.

Gelman, S. A. & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.

Harris, P. L. (1989). *Children and emotion: The development of psychological understanding.* Oxford, UK: Basil Blackwell.

Hosie, J. A., Gray, C. D., Russell, P. A., Scott, C., & Hunter, N. (1998). The matching of facial expressions by deaf and hearing children and their production and comprehension of emotion labels. *Motivation and Emotion*, *22*, 293–313.

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

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*, 288–299.

Kaitz, M. & Meschulach-Sarfaty, O. (1988). A reexamination of newborns' ability to imitate facial expressions. *Developmental Psychology*, *24*, 3–7.

Klinnert, M. D., Emde, R. N., Butterfield, P., & Campos, J. J. (1988). Social referencing: The infant's use of emotional signals from a friendly adult with mother present. In S. Chess, A. Thomas, & M. Hertzig (Eds.), *Annual progress in child psychiatry and child development* (pp. 26–39). New York: Brunner/Mazel.

Magai, C. & McFadden, S. H. (1995). *The role of emotions in social and personality development: History, theory, and research.* New York: Plenum Press.

Mandler, G. (1992). Emotions, evolution and aggression: Myths and conjectures. In K. T. Strongman (Ed.), *International review of studies on emotion*, Vol. 2 (pp. 97–116). Chichester: Wiley.

Maurer, D. & Barrera, M. E. (1981). Infants' perception of natural and distorted arrangements of a schematic face. *Child Development*, *52*, 196–202.

Nelson, C. A. & Dolgin, K. G. (1985). The generalized discrimination of facial expressions by seven-month-old infants. *Child Development*, *56*, 58–61.

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion*, *14*, 1–38.

Russell, J. A. (1990). The preschooler's understanding of the causes and consequences of emotion. *Child Development*, *61*, 1872–1881.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *155*, 102–141.

Russell, J. A. & Bullock, M. (1986a). Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition*, *4*, 309–341.

Russell, J. A. & Bullock, M. (1986b). On the dimensions preschoolers use to interpret facial expressions of emotion. *Developmental Psychology*, *22*, 97–102.

Russell, J. A. & Fernandez-Dols, J. M. (1997). What does a facial expression mean? In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression: Studies in emotion and social interaction, 2nd series* (pp. 3–30). New York, NY: Cambridge University Press.

Russell, J. A. & Widen, S. C. (in press). Words versus faces in evoking children's knowledge of the causes of emotions. *International Journal of Behavioral Development*.

Serrano, J. M., Iglesias, J., & Loeches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear, and surprise in 4- to 6-month-old infants. *Developmental Psychobiology*, *25*, 411–425.

Serrano, J. M., Iglesias, J., & Loeches, A. (1995). Infants' responses to adult static facial expressions. *Infant Behavior and Development*, *18*, 477–482.

Tomkins, S. S. (1962). *Affect, imagery, consciousness: Vol. I. The positive affects*. New York: Springer Publishing.

Thompson, R. A. (1998). Empathy and its origins in early development. In S. Braten (Ed.), *Intersubjective communication and emotion in early ontogeny. Studies in emotion and social interaction, 2nd series* (pp. 144–157). New York, NY: Cambridge.

Toner, H. L. & Gates, G. R. (1985). Emotional traits and recognition of facial expression of emotion. *Journal of Nonverbal Behavior*, *9*, 48–66.

von Brakel, J. (1994). Emotions: A cross-cultural perspective on forms of life. In W. M. Wentworth & J. Ryan (Eds.), *Social perspectives on emotion, Vol. 2* (pp. 179–237). Greenwich, CT: Jai Press.

Walker-Andrews, A. S. & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior and Development*, *14*, 131–142.

Waxman, S. R. & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*, 257–302.

Wierzbicka, A. (1992). Defining emotion concepts. *Cognitive Science*, *16*, 539–581.

Notes

1. We reanalyzed the data by treating the happy trial as a screening device and excluding children who made any errors in it. This procedure allowed us to ensure that, in addition to being able to perform the task with color cards, the children could also perform the task with faces, and to eliminate the possibility that performance on the angry trial was influenced by performance on or feedback from the happy trial. In all, 57 children (27 two-, 16 three-, 8 four-, 2 five-, 1 six-, and 3 seven-year-olds) made at least one error on the happy trial and were excluded. In the reanalysis (N = 303), the main effect for mode was significant, $F(2, 285) = 13.40$, $p < .001$. The proportion correct in the face-only mode (.87) was significantly lower ($p < .001$) than the proportion correct for label-and-face mode (.93), which was in turn significantly lower ($p = .01$) than the proportion correct for the label-only (.97) mode.

2. In a reanalysis including only those children who made no errors on the happy trial, the main effect for type-of-error was significant, $F(2, 594) = 241.29$, $p < .001$.

3. In Study 2, 14 children (7 boys, 7 girls) made at least one error on the happy trial. In a re-analysis (N = 54) including only those children who made no errors on the happy trial, the main effect for mode was not significant ($p = .17$), but the proportion correct was lower in the face mode (.90) than in the label mode (.97).

4. In the reanalysis including only those children who made no errors on the happy trial, the main effect for type-of-error was significant, $F(3, 156) = 10.97$, $p < .001$.

5. Because many of our participants were so young for a task of this nature, an extra happy face was included in the happy trial in Study 3. All indicators are that this is the easiest face to categorize. The extra face was included for both the word and face modes of presentation, and we could think of no plausible way this extra stimulus would account for the results we obtained.

6. The proportion of times a fear face was included in the angry box was $.70_a$, $.48_b$, $.65_{ab}$, and $.73_a$ for the four age groups from youngest to oldest (means that do not share a subscript differ significantly in LSD comparisons, alpha = .05).

Acknowledgements