

A Closer Look at Preschoolers' Freely Produced Labels for Facial Expressions

Sherri C. Widen and James A. Russell
Boston College

Children's performance on free labeling of prototypical facial expressions of basic emotions is modest and improves only gradually. In 3 data sets ($N = 80$, ages 4 or 5 years; $N = 160$, ages 2 to 5 years; $N = 80$, ages 3 to 4 years), errors remained even when method factors (poor stimuli, unavailability of an appropriate label, or the difficulty of a production task) were controlled. Children's use of emotion labels increased with age in a systematic order: *Happy*, *angry*, and *sad* emerged early and in that order, were more accessible, and were applied broadly (overgeneralized) but systematically. *Scared*, *surprised*, and *disgusted* emerged later and often in that order, were less accessible, and were applied narrowly.

How exactly do children of different ages interpret the facial expressions of those they see around them? An answer to this question is needed to understand children's perceptions, cognitions, and actions in face-to-face encounters with other people (Baldwin & Moses, 1996; Denham, 1998; Harris, 1989, 1994; Izard, 1994; Magai & McFadden, 1995). An answer will come from a variety of methods used with children at different ages.

Research with the youngest children must rely on behavioral measures, including neurophysiological recordings, social referencing, habituation, looking time, and facial expression. A review of these methods is beyond the scope of this article—suffice it to say that different behavioral methods are useful for different purposes but that they generally leave open the question of the precise interpretation a child makes of another's face. For example, in a social referencing study, when a child alters his or her behavior in response to the caregiver's facial expression, has the child interpreted that expression at all (as opposed to emitting a simple behavioral response or undergoing a change in state; Baldwin & Moses, 1996)? And, if so, in what terms has the child interpreted the expression—Fear? Attention? Negative emotion? Or something else?

With older children and especially adults, the question of interpretation is typically approached through their use of verbal labels. Verbal labels cannot be the gold standard, but how children come to label emotions is an important complement to other measures. In

addition, it is an interesting question in its own right. Forced-choice labeling of still photographs of facial expressions has been the most commonly used method with verbal children, but the forced-choice method has its drawbacks and can yield artifactual results: Children (and adults) can be forced to choose a label they would not have thought of spontaneously—or would even have rejected (Russell, 1994). In this article, we reexamine an older and apparently more straightforward method known as *free labeling*.

In the free labeling procedure, the experimenter typically shows a child a still photograph of a facial expression and asks, "How is this person feeling?" Darwin (1872/1965) used this method with adults, and the earliest research with children used this method (Gates, 1923; Kellogg & Eagleson, 1931). Use of free labeling then declined, although it has been used in a few later studies with children (Harrigan, 1984; Izard, 1971; Markham & Adams, 1992; Wiggers & Van Lieshout, 1985). Free labeling is worth a second look, because it appears less subject to the artifacts of forced choice and thus complements other methods. It is the method that seems to come closest to tapping a child's spontaneous specification of the emotion seen in a face.

Empirical analyses of behavioral measures of emotion recognition in infants and theoretical analyses of the evolutionary advantages of an emotion signaling system in the production and recognition of facial expressions (Bowlby, 1969, 1988; Denham, 1998; Harris, 1989; Izard, 1971) have been taken to suggest that well before their 2nd birthdays, children recognize specific emotions from facial expressions. Against this background, children's behavior when asked to label facial expressions has been surprising: Performance is modest and improves only gradually with age. For example, in Izard's (1971) study, the 2-year-olds' proportion of correct free label responses was less than .10 (see Figure 1). Izard's 3-year-olds fared somewhat better (.18), but improvement with age was very gradual, and fewer than half of the responses made by even the oldest children he tested (9-year-olds) were labels that Izard considered correct. Izard included nine emotions: enjoyment, surprise, anger, disgust, shame, distress, fear, interest, and contempt. Some of these proved difficult even for adults to label. Subsequent research (Markham & Adams, 1992) using a smaller set (happiness, sadness, anger, fear, surprise, and disgust) reported higher, though still modest, free labeling performance by

Sherri C. Widen and James A. Russell, Department of Psychology, Boston College.

This study was funded by the Social Science and Humanities Research Council of Canada through a grant to James A. Russell.

Study 1 in this article includes a reanalysis of data reported by Widen and Russell (2002).

We thank D. Geoffrey Hall, Michael J. Chandler, Lisa Ferguson, Ann Chawla, Melanie Parkin, Amit Taneja, Sarah Knowlden, Sherisee Pham, and Anita Christy. We also thank Linda Camras for providing the photographs of facial expressions used in these studies.

Correspondence concerning this article should be addressed to James A. Russell, Department of Psychology, McGuinn Hall, 140 Commonwealth Avenue, Boston College, Chestnut Hill, Massachusetts 02467. E-mail: james.russell@bc.edu

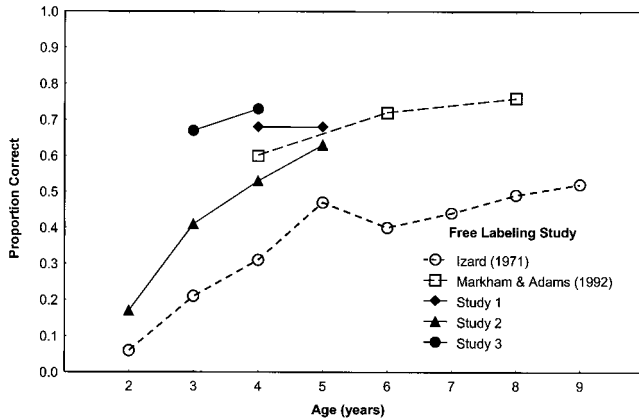


Figure 1. Mean proportion correct for free labeling of facial expressions in the three current studies and from two prior studies (Izard, 1971; Markham & Adams, 1992).

young children and only modest improvement with age (see Figure 1). Taken at face value, these results fall short of the level of performance one might anticipate if young children do indeed recognize specific basic emotions from faces.

A key question must be answered before such data from free labeling can be properly interpreted: Why do children make errors on this task? One obvious possibility is that children make errors when they do not interpret the facial expression correctly, at least in terms of what would be considered correct by the researcher. On this interpretation, children mislabel the sad face because they do not interpret a person displaying it as sad. If this is the case, children's "errors" reveal something of their actual understanding of a facial expression.

In a series of studies, Bullock and Russell (1984, 1985, 1986; Russell & Bullock, 1986a, 1986b) found complementary evidence that, indeed, preschoolers do not interpret facial expressions in terms of the same specific categories of emotion implied by researchers with the words *sad*, *scared*, *angry*, and the like. Bullock and Russell suggested that children interpret facial expressions but in a different way. Their earliest conceptual system for emotion is based on the dimensions of pleasure and arousal. With such methods as multidimensional scaling and asking children to select faces that corresponded to an emotion label, Bullock and Russell found that children's initial emotion categories are wide, encompassing any facial expression that is similar in levels of pleasure and arousal. For a 2-year-old, the mental category evoked by a facial expression of "anger," for instance, includes the unpleasant states of anger, disgust, fear, and sadness. Thus, a young child who labels the anger expression as *sad* is making an "error," but only from the researcher's point of view. The child's response is nonetheless an accurate reflection of the child's conceptual system for emotions. During the course of development, emotion categories narrow in a systematic manner until the adult system is achieved. These findings were replicated with a sample of German children (Bormann-Kischkel, Hildebrand-Pascher, & Stegbauer, 1990) and a sample of deaf children (Hosie, Gray, Russell, Scott, & Hunter, 1998).

Bullock and Russell's (1986) model has not been examined with free labeling data. One purpose of the present studies was to do so.

The "errors" produced in free labeling are predicted to conform to a structural model and to show initially broad categories that narrow with development. In addition, to the extent that free labeling reveals children's spontaneous categorization of facial expressions, the labels they choose may offer further insight into the development of emotion categories.

An alternative interpretation of free labeling errors is also possible and indeed was favored by past investigators and was responsible for the declining popularity of the free labeling method. According to this interpretation, errors are mainly or entirely artifacts that are due to the method itself. That is, preschool children really do understand facial expressions in the researcher's terms (they are correct as to specific emotion), but the method of free labeling *conceals* this ability. This interpretation has often been asserted but has not been empirically tested. We suggest three possible methodological reasons that free labeling might underestimate what a child knows: (a) The facial stimuli are inadequate or presented poorly. The earlier studies are especially subject to this criticism (for reviews, see Ekman, Friesen, & Ellsworth, 1972, and Izard, 1971). Even with adequate stimuli, if too many are shown, or if they are shown too briefly for a child to grasp, the child might be overwhelmed or not have enough time to respond. Or, if the expressions were posed by different and unknown adults, the child might be trying to identify the persons—their age, sex, ethnicity, and the like—rather than the emotion displayed. Seeing only the emotional facial expression of a stranger, the child has no neutral baseline against which to compare the expression. (b) The child must be able and willing to produce a label on demand, must understand the instructions, and must not be too shy or intimidated to perform properly. Thus, a production task might be too difficult for younger children. (c) The child might know the correct emotion category but not know its label. That is, if at a certain age the correct emotion word is simply unavailable (i.e., is not in the child's vocabulary), then the free labeling task would be inappropriate for that child.

OVERVIEW

Later in this article (Part 2), we use free labeling data to test and extend Bullock and Russell's (1986) model. But we begin (Part 1) with the prior question of the "method artifact" interpretation of free labeling data. We report three studies, in all of which the target emotions were restricted to the more successful six (happiness, sadness, anger, fear, surprise, and disgust). The problem of poor stimuli was controlled by using prototypical facial expressions of emotion presented in the simplest way we could conceive. The stimuli met Ekman and Friesen's (1978) criteria for the facial action units for each emotion. All of the expressions were posed by a single model to avoid the possibility that the child might focus on variations in age, sex, or ethnicity among different models. Before seeing the model's emotional facial expressions, the child saw the model with a neutral expression. Children were given as much time as they wanted to view each stimulus.

We begin with a reanalysis of data from another study (Widen & Russell, 2002) to explore label availability. Children (4 and 5 years of age) had been asked to name the emotions expressed by prototypical facial expressions and, separately, the emotions elicited by stereotypical emotion-evoking events. The reanalysis focused on those cases in which a child failed to label a specific

facial expression correctly. The question was whether the needed label was available in the child's vocabulary. We simply checked to see if that child had used the needed label elsewhere in the study (for another face or for an emotion-evoking event). We call this reanalysis Study 1.

Study 2 was a preliminary attempt to explore the idea that a production task is too difficult for younger children. Children (2 to 5 years of age) were asked to label six animals prior to labeling six facial expressions. If the children were willing and able to produce labels for photographs of animals, then the free labeling task in and of itself could not be blamed for children's poor performance with faces.

Encouraged by results from the first two studies, we designed Study 3 to eliminate all three methodological explanations for children's erroneous labeling of facial expressions. The stimuli were two sets of six prototypical facial expressions of emotion. Children (3 and 4 years of age) again labeled animals. To explore label availability, we developed an active priming procedure in which each child included in the sample was led to produce each of the target emotion labels spontaneously. Drawing on these three complementary data sets, we argue that for children 3 years of age or older, method factors play at best a minor role in free labeling performance.

PART 1: FREE LABELING PERFORMANCE AS A METHOD ARTIFACT

Study 1: Reanalysis of Data on Availability of Emotion Labels

This study is a reanalysis of data reported by Widen and Russell (2002). The method and procedure for Study 1 are fully described in that article.

Method

Participants

Participants were 80 children (40 girls and 40 boys) between the ages of 4 years 0 months and 5 years 11 months. Their mean age was 4 years 9 months ($SD = 6.06$ months). All of the children were fluent in English (as indicated on the permission form by the parent, by the day-care workers' opinions, and by the experimenter's opinion of the child's fluency in conversation) and were enrolled in day care in Vancouver, British Columbia, Canada.

Materials

Photographs of facial expressions. Two sets of 5 black-and-white 5×7 in. photographs (one set posed by a boy, one by a girl) of prototypical facial expressions of emotion (happiness, fear, anger, disgust, and sadness) were used. Except for hairstyle, the faces of the boy and the girl were in fact identical, having been "morphed" together with computer software (Morph 2.5 by Sierra On-Line, Inc., 1998). The original photographs were provided by Linda Camras. Camras, Grow, and Ribordy (1983) described the development of the photographs, their coding according to Ekman and Friesen's (1978) Facial Action Coding System, and their use in a study on recognition of emotional expressions.

Stories of emotional events. Five stories of stereotypical emotion-eliciting events were created that were based on prior work in our laboratory in which children generated causes and consequences of specific emotions (e.g., the story for happiness described the protagonist's birthday

party, and the story for sadness described the death of the protagonist's fish; for the complete stories, see Widen & Russell, 2002). The children were shown a drawing of the setting of each story while the story was being read.

Procedure

Each child participated in the three parts of the procedure in a single session. Each child's emotion concepts were first primed. Then each child was asked to label five facial expressions (of either the boy or the girl) and, separately, five stories of emotional events (order was counterbalanced).

The experimenter first spent time playing with a child until the child seemed comfortable with the experimenter. The experimenter asked the child for the names of two people at home with whom the child played games (call them X and Y). In order to prime the child's emotion concepts, the experimenter began a conversation in which six emotion words were inserted (*happy, sad, mad, scared, disgusted, and yucky*¹). The experimenter asked, for example, "Does Y ever feel *happy*?" "Do you sometimes feel *mad*?" "Does X ever get *scared*?" "Does Y ever feel *sad*?" and "Did you ever feel *disgusted*?" The experimenter did not discuss when or why these emotions might occur. If the child spontaneously offered an example of when someone had felt a particular emotion, the experimenter listened but did not comment on the child's story or encourage further explanation. Every effort was made throughout the experiment to use a neutral tone of voice when presenting the emotion words.

Faces. In the face mode of presentation, the experimenter introduced the faces by saying

I brought some pictures of Judd (Suzy). [In the face-first condition, the phrase was "a boy named Judd (a girl named Suzy)."] Would you like to look at them with me? Okay, here is a picture of Judd (Suzy) [showing the neutral expression]. Do you know what Judd (Suzy) is going to do? He (she) is going to show us how he (she) feels sometimes.

The experimenter then showed the child the five facial expressions, one at a time in a random order. For the first face, the experimenter said, "One day, Judd (Suzy) felt like this [pointing to the face]." For the other faces, the experimenter said, "One week later, Judd (Suzy) felt like this [pointing to the picture]." After each picture, the experimenter asked, "How do you think Judd (Suzy) feels in this picture?"

Responses were not corrected, and all were mildly praised (e.g., "Good answer"; "You are good at this game."). If no response was given, the experimenter used various prompts ("Have you ever made this face?" "What do you think happened to make Sally feel this way?"). If the child still did not respond, the experimenter went on to the next photograph and, after the other trials, returned to any to which the child had not responded. At no time did the experimenter use the word *emotion*, provide any other emotion label, or otherwise direct the child to try to use an emotion label beyond asking how Judd (Sally) was feeling.

Stories. The experimenter introduced the stories by saying

I'm going to tell you some stories about things that happened to Judd (Suzy). [In the story-first condition, the phrase was "a boy named Judd (a girl named Suzy)."] After each one, you get to tell me how you think Judd (Suzy) feels. How does that sound? Remember: Listen carefully, because you have to tell me how Judd (Suzy) feels.

¹ Two words, *disgusted* and *yucky*, were used in priming the disgust concept: "Another feeling is *disgusted*—like when something is really *yucky*." We used both labels to help ensure that the children's concept, however labeled, would be primed. Both *disgusted* and *yucky* were scored as correct in children's responses to the disgust face.

The experimenter then presented the stories, one at a time in a random order. The first story began with “Once upon a time,” and the other stories began with “One week later.” After each story, the experimenter asked, “How do you think Judd (Suzy) feels?”

Scoring

The child participants were allowed to use any label they chose. Collectively, the 80 children had 800 opportunities to label a face or story. These 800 opportunities yielded 39 different types of responses plus one category we labeled “nonresponse” that included only uninterpretable (silence) or nonsensical responses (e.g., “She’s just making a face”; “I don’t know”; “a mask”). For each of the responses, two raters made judgments regarding the specific emotion category into which each response fit. Responses from the current study rated correct for the happy category were “happy” and “excited”; for fear, “scared”; for disgust, “disgusted,” “yucky,” and “gross”; for anger, “angry,” “mad,” “cross,” “frustrated,” and “grumpy”; and for sad, “sad.” Responses varied from those just listed either in syntax or by embedment in a phrase (e.g., “very scared,” “totally grossed out”). These were all the labels children used that came close to specifying one of the specific target emotion categories. The two original raters agreed on category for 87% of the response types. In cases where the two original raters disagreed, a third rater rated the response, and the specific emotion category was determined by the majority (i.e., the category that two of the three judges chose). There were no cases in which all three judges chose a different category.

Results

Proportion correct for each emotion is given in Table 1. Overall, children were correct 68% of the time and incorrect 32% of the time. These results are comparable to past results (see Figure 1).

To examine the availability of emotion labels as an explanation for children’s responses, we focused on their errors (see Table 2). Of the 400 opportunities to label a face, children failed to use the correct label on 129 (32.2%). On 74 (57.4%) of these 129 trials, the children produced the needed label on a different trial: In 34 cases, the label was applied to another face; and in another 40 cases, it was applied to a story. Therefore, for over half of the cases of errors, unavailability can be ruled out as the explanation for the error.

Our test for label availability was extremely conservative. By “conservative,” we mean that the test is asymmetric and could underestimate the number of children for whom a given label was available. If a child produces a label, then the label is clearly available to that child. But the converse is not true. Perhaps a child

for whom the label was available nevertheless did not find it appropriate for the stimuli used in this study: These stimuli were a facial expression of other (and inappropriate) emotions or stories of various specific emotion-eliciting events (one of which was appropriate to the label).

Because the number of errors differed greatly for different emotions, it was difficult to compare the availability of the different emotion labels. For happiness, sadness, anger, and fear (see Table 2), the children who failed to label the face correctly used the needed label elsewhere 88.0% of the time (range = 75.0%–100.0%). In contrast, the percentage of times that children used the word *disgust* after failing to label the disgust face correctly was lower (26.7%) than the percentage for any other label. Thus, a reasonable hypothesis is that *disgust* is available to fewer children (or less accessible to them) than are the other emotion labels tested here, but this suggestion requires further investigation.

Study 2: Free Labeling Task With Animals and Faces

One purpose of Study 2 was to investigate the possibility that the free labeling task per se is too difficult for young children. Thus, we asked children to label animals as well as facial expressions. We used a procedure similar to that of Study 1, especially the free labeling response format, but one that differed in five ways: First, the sample included four age groups (2-, 3-, 4-, and 5-year-olds). Second, a comparison free labeling task was introduced (labeling animals). Third, surprise was included as one of the target emotions, resulting in a total of six emotions (happiness, surprise, fear, anger, disgust, and sadness). Fourth, no stories were presented. Fifth, in order to investigate children’s spontaneous interpretation of facial expressions, and to identify the ages at which children begin to use the target emotion terms, no priming procedure was used.

Method

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

Participants

Participants were 160 children (80 girls and 80 boys). There were 40 children (20 girls and 20 boys) in each of four age groups: 2-year-olds (age range = 28 to 35 months; *M* = 30 months, *SD* = 3.7), 3-year-olds (age range = 36 to 47 months; *M* = 42 months, *SD* = 3.6), 4-year-olds (age range = 48 to 59 months; *M* = 53 months, *SD* = 3.3), and

Table 1
Proportion Correct on Each Facial Expression in Each Study

Type of facial expression	Study 1		Study 2				Study 3			95% confidence interval
	Age 4	Age 5	Age 2	Age 3	Age 4	Age 5	Age 3	Age 4	<i>M</i>	
Happiness	.94	.96	.39	.90	.95	.98	1.00	1.00	1.00	—
Sadness	.93	.92	.28	.50	.78	.65	1.00	1.00	1.00	—
Anger	.78	.77	.33	.83	.88	.98	.98	1.00	.99	.96–1.00
Fear	.52	.58	.03	.10	.30	.45	.53	.68	.60	.46–.71
Surprise	—	—	.00	.10	.25	.65	.38	.50	.47	.33–.54
Disgust	.22	.15	.03	.03	.03	.10	.13	.20	.17	.08–.25

Note. Ages are in years.

Table 2
Availability of Emotion Labels in Study 1: When Children Did Not Correctly Label a Face, Did They Use That Label for Another Stimulus?

Face	No. of errors	Of children who did not correctly label the target face, number who used the label for			Availability ^a	(%)
		Another face	Story but not face	Never		
Happiness	4	2	1	1	3/4	(75)
Sadness	6	4	2	0	6/6	(100)
Anger	18	17	0	1	17/18	(94)
Fear	37	4	27	6	31/37	(83)
Disgust	64	7	10	47	17/64	(27)
Total	129	34	40	55	74/129	(57)

Note. Maximum possible number of errors for each row is 80.

^a Availability = number of children who failed to label the face with the predicted label but who used the needed label at least once, divided by total number of children who failed to label the face with the predicted label.

5-year-olds (age range = 60 to 71 months; $M = 65$ months, $SD = 3.1$). None of these children had participated in Study 1.

Materials

Photographs of animals. The animal pictures were six color photographs, one each of a cat, dog, horse, cow, turkey, and goose, that were included in the book *Who Are You? Country Friends* (1992).

Photographs of facial expressions. Each of seven photographs (one each for happiness, sadness, anger, fear, disgust, surprise, and neutral) was of the same 13-year-old girl. The photographs were provided by Linda Camras.

Procedure

The experimenter spent the first visit getting to know each child. On a subsequent visit, the experimenter invited an individual child to look at the special books the experimenter had brought with her. The pictures of animals were presented in their original book format, and the photographs of facial expressions were placed one per page in a photo-album. Each child participated in two sets of trials. The first set of trials concerned animals; the second set, facial expressions.

The animal trials served as a practice session and a comparison task. The experimenter said, "This is my special book about animals." The experimenter then opened the book and pointed to the first page, saying, "Do you know what kind of animal this is?" As in Study 1, responses were not corrected and were mildly praised equally (e.g., "Good answer"; "You are very good at this game"). Six animals were shown, always in the same order. The experimenter ended by saying, "That was really fun. Do you want to look at my second book?"

For the second set of trials, the order of the emotional facial expressions was varied randomly except that the first trial was always the smile because pilot testing indicated that even young children frequently labeled this expression as happy.

Scoring

Animals. Owing to mechanical difficulties, 5 participants' responses to the animal pictures were not recorded. Thus, collectively, the 155 children had 930 opportunities to label an animal. For these 930 opportunities, there were 46 different response types plus one category we labeled "nonresponse" that included only uninterpretable or nonsensical responses (e.g., "must be a horn"; "I don't know"; or the child saying nothing). For the

responses, two raters each made two judgments: (a) broad category and (b) specific category.

For the broad category ratings, the raters' task was to indicate whether the response was a mammal (e.g., cat, dog, donkey), a bird (e.g., turkey, goose, eagle), or neither (e.g., frog). The two raters agreed on the broad category for 98% of the response types. Disagreements were resolved by discussion.

For the specific category ratings, the raters' task was to indicate into which one of six specific animal categories each response fit (cat, dog, horse, cow, turkey, or goose) or if it did not fit into any of these categories. The labels included in the cat category were "cat" and "kitten"; in the dog category, "dog" and "puppy"; in the horse category, "horse"; in the cow category, "cow" and "calf"; in the turkey category, "turkey"; and in the goose category, "goose." The two raters agreed on the specific category for 93% of the response types. Disagreements were resolved by discussion.

Facial expressions. Collectively, the 160 children had 960 opportunities to label a facial expression. These 960 opportunities yielded 153 different types of responses plus one category we labeled "nonresponse" that included only uninterpretable or nonsensical responses (e.g., "I dunno"; "a harebrain"; "Mommy"; "turn the page"; or the child saying nothing). For each of the 153 types of responses, two raters made two judgments: (a) valence and (b) specific emotion category. Disagreements were resolved by a third rater who rated only those responses on which the two original raters disagreed.

For the valence ratings, the raters' task was to indicate whether the response was positive (e.g., "happy," "good"), negative (e.g., "angry," "sad," "bad," "shy"), or uninterpretable in regard to valence (e.g., "funny," "a little bit smile"). (Because surprise can be pleasant or unpleasant, "surprised" was not scored for valence.) The two original raters agreed on the valence for 85% of the response types. In cases where the two original judges disagreed, a third judge rated the response, and the valence was determined by the majority (i.e., the valence that two of the three judges chose). There were no cases in which each judge chose a different valence rating.

For the specific emotion category ratings, the raters' task was to indicate into which one of six emotion categories each response fit (happy, sad, angry, surprise, disgust, or scared) or if it was uninterpretable in regard to these six categories. Responses rated correct for the happy category were "happy," "excited," "fine," "going to play," "good," "she's thinking about someone she likes," "nice," "better," and "yeah!"; for the sad category, "sad," "upset," "hurt," and "she got an owie"; for the angry category, "angry," "mad," "cross," "frustrated," "grouchy," and "grumpy"; for the

scared category, “scared,” “frightened,” “nervous,” and “shy”; for the surprised category, “surprised” and “shocked”; and for the disgusted category, “yucky” and “gross.” Responses varied from those just listed in syntax or by embedment in a phrase (e.g., “very scared,” “when someone hurts you”). These were all the labels children used that came close to specifying one of the specific target emotion categories. The two original raters agreed on the category for 84% of the response types. In cases where the two original judges disagreed, the third judge rated the response and the specific emotion category was determined by the majority (i.e., the category that two of the three judges chose). There were no cases in which all three judges chose a different category.

Results

Animals

Figure 2 shows the results with animals. Overall, the children were correct on 73.8% of the 930 trials. For the four mammals, the children did very well: Even the 2-year-olds produced a high proportion of correct labels (85% to 97%). Every child named at least two mammals correctly. These results are clear evidence that even the youngest 2-year-old understood the instructions and was willing and able to produce a label.

The children did less well with the birds: 15% to 54% for the turkey and 5% to 38% for the goose. Analysis of errors was revealing. Of the 310 opportunities to label a bird, there were 235 errors. Of these, 26% were nonresponses, and 74% were incorrect as to specific category but correct as to broad category (i.e., they named some type of bird); there were no responses that were incorrect as to broad category. In other words, children either were silent or chose a label from the same general category, calling the turkey a chicken or a duck, for example. No one called it a fish or a tiger. These results show that an analysis of errors can reveal implicit broad categories, a result consistent with the general finding that children of this age overgeneralize a given word to members of a broader category (Gelman & Markman, 1986).

Facial Expressions

Figure 3 shows a comparable analysis of emotion labels. Overall, the children were correct (as to specific category) on 42.9% of

the 960 trials. Proportion correct varied both with face, $F(5, 760) = 131.66, p < .01$, and with age, $F(3, 152) = 50.55, p < .01$. Results are given in Table 1. These results are comparable to past free labeling results (see Figure 1).

The children’s performance on labeling animals was significantly better than the same children’s performance on labeling facial expressions of emotion: The mean number of correct animal labels (4.4 out of 6) was significantly greater than the mean number of correct emotion labels (2.6 out of 6), $t(154) = 15.85, p < .01$. We return to an analysis of errors with facial expressions in a later section.

Study 3: Main Study

To summarize our results so far, older preschoolers (4- and 5-year-olds) in Study 1 did make the expected errors, but in at least half the cases, the children’s spontaneous use of the needed label elsewhere in the study ruled out label unavailability as an explanation for most errors. In Study 2, even the youngest children (2-year-olds) were willing and able to produce labels for mammals, which indicates that they understood the instructions and the task. Thus, their poor performance when labeling facial expressions cannot be attributed to the productive nature of the free labeling task.

Encouraged by these preliminary results, we designed Study 3 to control all three method factors to which children’s poor free labeling performance has traditionally been ascribed. First, as in Studies 1 and 2, we limited our study to six emotions. We used prototypical facial expressions of emotion that met Ekman and Friesen’s (1978) specifications, and we presented them using a simple procedure that minimized distracting information. Second, each child began by labeling three mammals: This step served to introduce the free labeling procedure and to replicate our finding that the children are willing and able to produce labels. Third, we developed a technique we call active priming through which each child was led to spontaneously produce each of the six target emotion terms. A child who successfully completes active priming has the emotion labels available. This procedure also allowed us to

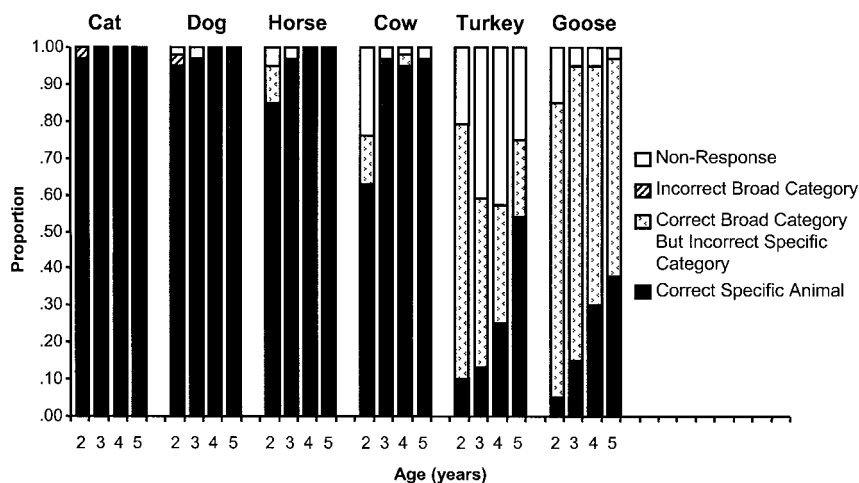


Figure 2. Proportion of responses to the animal pictures that fit into each of four mutually exclusive response categories in Study 2.

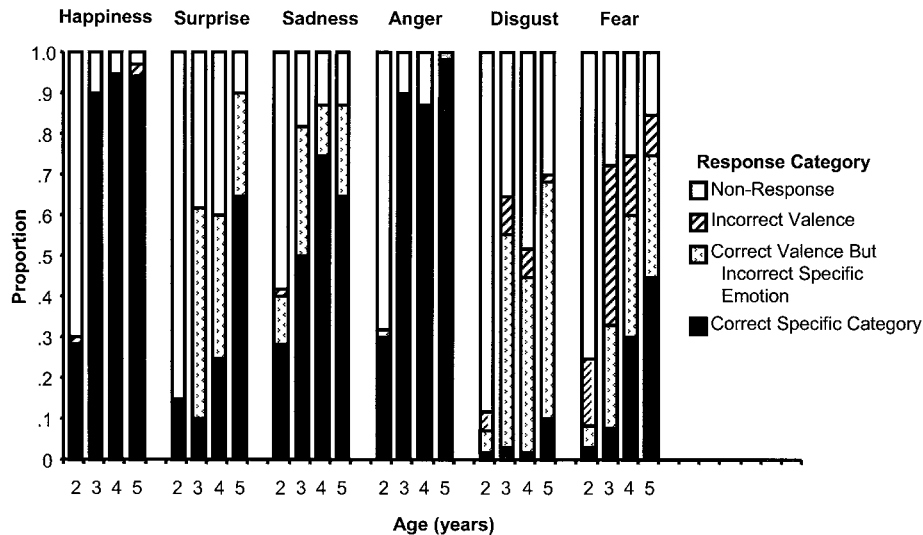


Figure 3. Proportion of responses to the facial expressions that fit into each of four mutually exclusive response categories in Study 2.

begin exploring the hint in Study 1 that among available labels, some are more accessible than others.

The procedure of Study 3 differed from that of Study 2 in four principal ways: First, the participants were 3 or 4 years of age. (The 2-year-olds' performance on the face labeling task in Study 2 was very low, and we doubted that active priming would be successful with a sufficient number of 2-year-olds for all labels. The 5-year-olds' performance in Study 2 did not differ significantly from that of the 4-year-olds.) Second, the children labeled three mammals (instead of six animals). Third, there were two sets of facial expressions (those posed by the 13-year-old girl used in Study 2 and a set posed by a 12-year-old boy) to allow a more generous criterion for a child's being able to label the emotion expressed by a face. Fourth, the active priming procedure was carried out prior to the labeling of the facial expressions.

Method

The method was identical to that in Study 2 except as noted.

Participants

Participants were 80 children (40 boys and 40 girls). There were 40 children in each of two age groups: 3-year-olds (age range = 36 to 47 months; $M = 43$ months, $SD = 3.4$) and 4-year-olds (age range = 48 to 59 months; $M = 54$ months, $SD = 3.5$).

Materials

Photographs of animals and facial expressions: The animal pictures were three color photographs, one each of a cat, horse, and rabbit. In addition to the set of facial expressions posed by the 13-year-old girl in Study 2, there was another set, also provided by Linda Camras, that was posed by a 12-year-old boy.

Procedure

The experiment was spread across three visits, each separated by anywhere from 1 to 7 days ($M = 2.5$ days). The experimenter spent the first

visit getting to know each child, the second visit on the active priming procedure, and the third visit on the three labeling tasks.

Active priming. On the second visit, the experimenter invited an individual child to play a game designed to elicit the target emotion terms from the child: *happy, angry, sad, scared, surprised, and disgusted* (or a synonym as described below). The game consisted of a three-step procedure. For the first step (free listing), the experimenter began by saying, "Today we are going to talk about feelings. Can you name some feelings for me?" The child was credited with any of the target emotion labels produced in response. When the child indicated that he or she could not think of any more feelings, the experimenter listed the remaining terms by saying, "Some other kinds of feelings are . . ."

The second step (experimenter's stories) concerned the emotion terms not produced by the child in the first step. The experimenter said, "Now I am going to tell you about some things that sometimes happen to people. And you can tell me how you would feel if they happened to you." For example, for a child who had not already produced the word *sad*, the experimenter asked, "How would you feel if you had an ice cream cone, and it fell on the ground?" If the child still did not produce *sad*, the experimenter offered a second story: "What if your favorite toy broke? How would you feel?"

The third step (own story) was introduced for those emotion terms the child had not produced in Steps 1 or 2. For example, for the child who still had not produced *sad*, the experimenter asked, "What would make *you* sad?" If the child produced the word *sad* within his or her response (e.g., "When X happens, I'm sad") and if the response was plausible, the child was credited with the word. If the child did not produce the word within his or her own example, the experimenter asked, "So, if X [child's own example] happened, how would you feel?" For each emotion term, the experimenter judged whether the child had used the term meaningfully (i.e., that the child was not simply echoing the experimenter). If it seemed that the child was merely echoing the experimenter, the experimenter rejected that response as a production of the term.

Animal labeling task. On the final visit, the child was invited to play a new game in which he or she was asked to label animals and facial expressions. The animal labeling task was presented first and served as a practice session and a comparison free labeling task. The experimenter asked, "Do you know what I brought with me today? I brought some pictures of animals. Would you like to see them?" The experimenter then

showed one of the animal pictures and asked, "What kind of animal is this one?" The order of the animals was varied randomly.

Facial expressions. Next, the experimenter introduced the pictures of the girl's (Sally's) facial expressions. The presentation of the facial expressions was identical to that in Study 2 except that the happiness expression was not always presented first; the order of the emotional facial expressions was varied randomly. After the child had labeled each of Sally's facial expressions, the experimenter introduced the boy's (Jesse's) facial expressions. Upon completion of the last trial, the experimenter praised the child for his or her participation (regardless of how the child had responded) and thanked him or her for playing the game.

Scoring

Active priming. Data from the active priming procedure were scored in two ways. First, a pass/fail criterion (availability) was based on whether the child produced a given term on any one of the three steps. The second way of scoring the data was, for the available terms, to give points for the ease of eliciting the term (accessibility): 3 points if the term was elicited in Step 1 (free listing); 2 points if the term was elicited in Step 2 (after one of the experimenter's stories); and 1 point if the term was elicited in Step 3 (after the child's own story).

Animals. The labels scored as correct in the cat category were "cat" and "kitten"; in the horse category, "horse" and "horsey"; and in the rabbit category, "rabbit" and "bunny."

Facial expressions. The 480 opportunities to label a facial expression yielded 19 different types of responses plus "nonresponse."

Results

The active priming procedure was used as a screening device: Any child who did not produce all six target emotion terms was replaced by a same-sex age-mate. Ten children had to be replaced.² Thus, all of the children included in this sample had produced each of the target emotion terms in the active priming trials, thereby indicating that the target emotion terms were available in their vocabularies.

Ease of Eliciting Emotion Terms (Active Priming)

The ease of eliciting each different emotion term (a possible index of its accessibility) varied with emotion, $F(5, 380) = 54.40$, $p < .01$. On the basis of the second scoring method, in which a low score (minimum = 1) indicates difficult elicitation and a high score (maximum = 3) indicates easy elicitation, the means for ease of production for the emotion terms were as follows: *sad*, 2.53; *happy*, 2.50; *angry*, 2.23; *scared*, 1.81; *disgusted*, 1.76; and *surprised*, 1.30. Pairwise comparisons (t tests, $df = 79$) of these means indicated that the means for *sad*, *happy*, and *angry* were each significantly higher ($p < .01$) than the means for each of the other three labels (*scared*, *disgusted*, and *surprised*). For 3-year-olds, the median was 3 for *happy*, 2 for *sad*, *angry*, *scared*, and *disgusted*, and 1 for *surprised*. For 4-year-olds, the median was 3 for *happy*, *sad*, and *angry*, 2 for *scared* and *disgusted*, and 1 for *surprised*. Thus, even terms that are available may vary in accessibility.

Labeling Mammals

Collectively, the 80 children had 240 opportunities to label a mammal. Overall, the children were correct on 98.8% of these. For

the rabbit and cat, there were no errors: All children were correct. For the horse, 3 children (two 3-year-olds and one 4-year-old) labeled the horse incorrectly (1 called it a "cow," 1 a "camel," and 1 a "dog"); Performance was 96.3%. Thus, all 80 children correctly labeled at least two of three mammals, and 77 of the 80 children correctly labeled all three. (The three errors were consistent with the results of Study 2: Children chose a label from an implicit broader category of mammal. Thus, by the "broad" criterion of choosing some name applicable to a mammal, 100% were correct for every mammal.) These results indicate that every child understood the instructions and was willing and able to produce a label.

Facial Expressions

To be credited with correctly labeling a facial expression, a child had to accurately label at least one exemplar (Sally, Jesse, or both) of that expression; any child who labeled neither exemplar correctly was scored as failing.

Collectively, the 80 children had 480 opportunities to label a facial expression (with each opportunity involving two trials). Overall, the children correctly labeled at least one of the two exemplars of a facial expression on 70.3% of the 480 opportunities. These results are comparable to previous results (see Figure 1), although slightly higher (perhaps because children had two opportunities to label each type of facial expression rather than one). The proportion of children who produced correct emotion terms (.70) was significantly lower than the proportion producing correct mammal labels (.99), $t(79) = 16.57$, $p < .01$. This result replicated the finding of Study 2.

Proportion correct (shown in Table 1) varied with type of facial expression, $F(5, 380) = 93.59$, $p < .01$. For the happy and sad faces, performance was perfect; for the remaining faces, it was less than perfect. A 95% confidence interval showed that performance for anger also included the possibility of ceiling level performance. But for the remaining three emotions (fear, surprise, disgust), the upper limit on the confidence interval was less than 1.0. Between 40% and 84% of the children did *not* link these three faces with the target emotion label.

Performance also varied with age, $F(1, 76) = 4.35$, $p = .04$. As have all prior researchers using this task, we found only gradual

² Of the 10 children excluded, 5 were 3-year-old girls, 2 were 4-year-old girls, and 3 were 4-year-old boys. One (4-year-old girl) failed to produce *angry*; 1 (3-year-old girl) failed to produce *scared*; 6 (three 3-year-old girls, one 4-year-old girl, and two 4-year-old boys) failed to produce *surprised*; and 3 (one 3-year-old girl and two 4-year-old boys) failed to produce *disgusted*. However, for three of the four emotion terms, at least some of the children who failed to produce the term in active priming used it in free labeling: For *surprised*, 2 of the 6 children used it in free labeling; for *angry*, the 1 child used it in free labeling; for *scared*, the 1 child used it in free labeling; and for *disgusted*, none of the 3 children used it in free labeling. So we cannot conclude that the 10 children excluded because of failure to produce a term during active priming did not have that term accessible. Instead, it appears that our active priming procedure could be improved so that an even greater percentage of children would succeed. The proportions of these 10 children who labeled at least one example (Sally, Jesse, or both) of each facial expression correct were 1.0 for happiness and sadness, .90 for anger, .40 for fear, and .20 for surprise and disgust.

improvement with age (the overall mean for 3-year-olds was .67, and for 4-year-olds it was .73; see Figure 1). The smallness of change with age was due in part to a ceiling effect for three emotions (happiness, sadness, and anger). Still, for fear, surprise, and disgust, the mean improvement was gradual (for 3-year-olds, .36; for 4-year-olds, .46). Similar improvement with age occurred for each emotion (the Age \times Face interaction was not significant, $p = .57$).

Discussion of Part 1: Is Poor Performance on Free Labeling a Method Artifact?

The proportion correct when labeling faces that was found here replicated that found in prior studies (Harrigan, 1984; Izard, 1971; Markham & Adams, 1992; Wiggers & Van Lieshout, 1985): The children's performance was modest and improved only gradually (see Figure 1). The first question is whether the errors children made are primarily an artifact of method.

The first possible method artifact concerned the quality of the facial stimuli and the difficulty of presentation. The facial expressions used here conformed to the criteria for prototypical expressions of basic emotions defined by Ekman and Friesen (1978). Our method of presenting them was as simple and helpful as we could think to make it. Of course, our stimuli and method can be questioned with regard to ecological validity. (Indeed, we believe that actual emotional displays are rarely so prototypical [Carroll & Russell, 1996] and rarely occur with such optimal presentation. More naturally occurring, ecologically valid facial stimuli are worthy of more research.)³ Nevertheless, this manner of presentation seemed to us near optimal for allowing children to make their best judgment for the single cue of a facial expression.

The second possible methodological explanation is that the label production task per se is too difficult for young children. We found that every child in Study 2 labeled at least two mammals correctly. Even the 2-year-olds correctly labeled the four mammals on a large proportion of trials (85% to 97%), and the 3-, 4-, and 5-year-olds were 100% correct with dogs and horses. Similarly, in Study 3, the 3-year-olds labeled the three mammals correctly on 98.8% of the trials; all of the children were correct with rabbits and cats. We acknowledge that naming animals is not a parallel task to labeling emotions from facial expressions.⁴ Nonetheless, the fact that every child produced at least two correct labels shows that the instructions and label production method per se were well within every child's grasp.

The third factor—unavailability of emotion labels—was partially addressed in Study 1 and ruled out in Study 3 for children at least 3 years of age. In Study 1, in over half (57%) of the cases in which a child failed to correctly label a face, that child spontaneously used the needed label for another face or for a story. Given that the children had responded to only one set of facial expressions and one set of stories, the test of availability used in Study 1 was highly conservative, thereby rendering the results even more encouraging. In Study 3, a more rigorous method of ruling out label unavailability was developed: active priming. This procedure established the availability of all target emotion labels for all the children in the sample. Study 3 therefore showed that for children 3 years of age and older, even with all labels available,

performance was low for fear (.53), surprise (.38), and disgust (.13).

Altogether, our results speak against an otherwise plausible and theoretically motivated account of children's free labeling of facial expressions, namely, that from a very early age (well before their second birthdays) children understand the specific basic emotion conveyed by its facial expression but that this understanding is concealed by the method of free labeling. This account holds that as children gradually acquire the words for the basic emotion categories they already know, and as they come to understand what they are to do in a production task, they come to label faces correctly. If this account is inconsistent with the data, what alternative might be considered? Clearly, there are many possibilities; here we pursue one that we find promising.

PART 2: TESTING AND EXTENDING BULLOCK AND RUSSELL'S (1986) ACCOUNT

The theme of Bullock and Russell's (1986) account is that children's responses can be analyzed for what they reveal about children's manner of interpreting facial expressions rather than simply categorized as correct or incorrect by a researcher-imposed theoretical criterion. "Errors" are not random and are not simply products of method. One specific prediction is that free labeling errors conform to a structural model previously found in children's similarity judgments and forced-choice labeling (Bormann-Kischkel et al., 1990; Bullock & Russell, 1986; Hosie et al., 1998). A second prediction is that children's emotion categories are initially few in number but that the number increases with age and experience as the original categories narrow and new ones are acquired. Thus, in Part 2, we first analyzed the data of the three studies from Part 1 to test whether children's errors conformed to

³ As have other researchers, we chose still photographs, and another important ecological question from this line of research is whether the results can be generalized to dynamically moving stimuli. Outside the laboratory, of course, children see faces embedded in a context of voice, other behavior, antecedent situation, and the like. The emotion they attribute to another given this full set of complementary cues is a topic of major importance. Nevertheless, one theoretically interesting and legitimate avenue has been to ask about the child's response to the face alone.

⁴ One might be tempted to attribute superior performance with animals to deliberate training, and indeed preschoolers are trained how to label animals. Books of labeled animal pictures are readily available. On the other hand, at the day-care centers we visited, there was often a great deal of discussion of emotions as well. The child-care workers deliberately modeled the use of emotion words and encouraged the children to notice how others feel and to "use their words" instead of acting out. These day-care centers had a variety of books and puzzles about emotions and the facial expressions associated with each. Amazon.com lists 1,565 children's books about feelings and emotions, including *The Feelings Book* (Parr, 2000) and *All My Feelings at Preschool: Nathan's Day* (Conlin, 1991), as well as a variety of books about particular feelings, such as *I'm Mad* (Crary, 1992) and *Some Things are Scary* (Heide, 2000). In addition, Amazon.com lists a variety of books about facial expressions, including *Feelings and Faces: Feelings Activity Book* (Boulden, 1993), *Toby's Silly Faces* (Szekeres, 2000), *Mrs. Mustard's Baby Faces* (Wattenberg, 1998), and *Baby Faces* (Suter & Levin, 1998). These books make some emotion words more prevalent in the child's environment than others and make explicit the face that goes with some emotion words.

the structural model and whether the number of emotion categories children used (as indicated by labels used) increased with age.

Bullock and Russell's (1986) account is preliminary. For example, they did not specify which emotion labels children apply to their initial categories or the order in which later-acquired emotion categories and labels are added. The data from the current studies were thus also analyzed to explore these questions. Specifically, we examined what labels children applied to faces at different ages and whether the emergence of specific labels was systematic. In other words, we attempted to extend Bullock and Russell's (1986) account in a data-driven manner.

Testing Bullock and Russell's (1986) Structural Model

To investigate whether children's free labeling responses in each study conformed to Bullock and Russell's (1986) structural model, we counted the number of times a label from each of the six target emotion categories was produced for each face (see Figure 4). In Figure 4, the emotion categories are rank ordered in accordance with the structural (circumplex) model, with emotions that are most alike adjacent to one another (e.g., anger and disgust) and those that are least alike furthest apart (e.g., happiness and sadness). With this arrangement, all the "correct" responses to the faces are on the center (white) diagonal, and all the "errors" are on the other diagonals. Those responses that were incorrect by what in Bullock and Russell's account is one step (e.g., calling the disgust face *angry* or *sad*) are on the two closest diagonals (lightest gray), one on either side of the center diagonal. Those responses that were incorrect by two steps are on the next pair of adjacent diagonals (slightly darker gray), and so on, to the maximum number of five steps (darkest gray diagonal) away from the target emotion category (e.g., calling the happy face *sad*, or the sad face *happy*). The prediction was that the likelihood of choosing a label is proportional to the number of steps.

To test this prediction, we summed the cells of each diagonal and divided it by the number of cells in that diagonal; for example, in Study 1 (see Figure 4A), the relative frequency of Step 1 = $(0 + 2 + 4 + 8 + 9 + 4 + 43 + 1)/8 = 71/8 = 8.88$. In all three studies, the prediction was generally confirmed, although in Study 1 (see Figure 4A), there was a reversal between Steps 2 and 3, and in Study 2 (see Figure 4B), there was a reversal between Steps 3 and 4. Errors were not random but were more likely to be from emotion categories adjacent to the target category than from categories further removed. As was found earlier with labeling animals, "errors" suggest broader implicit categories. That "errors" conform to a structural model complements our earlier conclusion that errors are not simply artifacts of method. This structural model, although evident in the data, was not expected to tell the whole story; that is, its predictions are only approximate, for at least three reasons. First, only ordinal values for the relations among the emotions are used. Second, we collapsed data across ages. Third, the marginal probability of using a label is ignored. We thus turn next to frequency of label use.

Frequency of Label Use

Free labeling allows one to examine the children's spontaneous use of the labels *happy*, *sad*, *angry*, *scared*, *surprised*, and *dis-*

gusted. A simple but important point is made in Figure 5, which shows data from Study 2 (chosen because no priming procedure was involved) arranged to show the number of these labels used over the span from 2 to 5 years of age. (By "label," we mean any word earlier scored as correct for a given emotion; thus, synonyms count as the same label.) The modal number of different emotion labels used by the 2-year-olds was 0, although 57.5% used 1, 2, or 3 (different) labels in response to the six faces. The modal number of labels used by the 3-year-olds was 2. The modal number used by the 4-year-olds was 4 (although their mean number of labels used was lower: 3.35). The modal number used by 5-year-olds was also 4 ($M = 4.03$). Thus, the number of different emotion labels children used increased with age.

Frequency of Use of Specific Labels

More interesting than growth in sheer number of different labels used was the unequal frequency in the use of specific labels. In all three studies, children responded to an equal number of emotions (as represented by faces). Nevertheless, Figure 4 (column totals) shows that in all three studies, some labels were used more frequently than others. *Angry* was used most, followed by *happy*, *sad*, *scared*, *surprised*, and finally *disgusted*. Similar results have been reported before (Harrigan, 1984; Markham & Adams, 1992; Wiggers & Van Lieshout, 1985). One interpretation of this pattern has been that some facial expressions (anger, happiness, sadness) are identified earlier or more easily than others (fear, surprise, disgust; e.g., Gosselin & Simard, 1999; Izard, 1971, 1994). However, this explanation cannot be the whole story because, as is also shown in Figure 4, the frequency of *incorrect* uses (errors) of these labels showed the same rank order. The pattern is therefore not a result of the faces per se but rather a result of the children who are interpreting those faces. This analysis could be tested by examining children's labeling of other cues to emotion, such as voice, context, or action. (Our result with frequency of errors raises the possibility that the repeated finding that happiness, sadness, and anger expressions are easier to interpret might be, in part, an artifact of label use.)

Systematic Emergence of Labels

The two previous observations—that, with age, children use a greater number of labels and that children favor some labels over others—suggest that labels might emerge systematically during development. Because Study 2 lacked any priming procedure, data from Study 2 were used to explore this possibility. We first sorted all children, irrespective of age, by the number of labels they used. We found that if a child used only one label, that label was most likely to be *happy*. If he or she used two labels, they were most likely to be *happy* and *angry*. If three labels, they were most likely to be *happy*, *angry*, and *sad*. If four labels, *surprised* was added. If five labels, *scared* was added. And if six labels, *disgusted* was added. Of the 160 children, 67.5% fit this pattern—a proportion significantly greater ($p < .001$) than the percentage expected by chance. (If, as the number of labels that children used increased, any label was as likely to be added as any other, then 20.3% would

A

Face	Response Category					Total	Number of Steps from Target	Relative Frequency
	Happiness	Fear	Anger	Disgust	Sadness			
Happiness	76	0	1	0	0	77	Correct	271/5= 54.2
Fear	9	43	2	4	11	69	Step 1	71/8= 8.9
Anger	0	4	62	4	4	74	Step 2	13/6= 2.2
Disgust	0	1	43	16	8	68	Step 3	13/4= 3.3
Sadness	0	2	3	1	74	80	Step 4	0/2= 0.0
Total	85	50	111	25	97	368		
Errors	9	7	49	9	23	97		
(%)	(10.6)	(14.0)	(44.1)	(36.0)	(23.7)	(26.4)		

B

Face	Response Category					Total	Number of Steps from Target	Relative Frequency	
	Happiness	Surprise	Fear	Anger	Disgust				Sadness
Happiness	124	0	0	0	1	0	125	Correct	411/6= 68.5
Surprise	28	40	13	4	1	5	91	Step 1	126/10= 12.6
Fear	33	8	34	11	1	16	103	Step 2	70/8= 8.8
Anger	0	0	1	119	1	2	123	Step 3	19/6= 3.2
Disgust	10	0	0	56	7	27	80	Step 4	16/4= 4.0
Sadness	1	0	2	30	1	87	121	Step 5	1/2= .5
Total	196	48	50	220	12	117	643		
Errors	72	8	16	101	5	30	233		
(%)	(36.7)	(16.7)	(32.0)	(45.9)	(41.6)	(25.6)	(36.3)		

C

Face	Response Category					Total	Number of Steps from Target	Relative Frequency	
	Happiness	Surprise	Fear	Anger	Disgust				Sadness
Happiness	160	0	0	0	0	0	160	Correct	610/6= 101.7
Surprise	43	63	23	6	1	11	147	Step 1	218/10= 21.8
Fear	30	7	68	21	5	28	159	Step 2	60/8= 7.5
Anger	3	0	1	150	2	4	160	Step 3	32/6= 5.3
Disgust	3	0	4	98	21	23	149	Step 4	14/4= 3.5
Sadness	0	0	0	11	0	148	159	Step 5	0/2= 0.0
Total	239	70	96	286	29	214	934		
Errors	79	7	28	136	8	66	324		
(%)	(33.1)	(10.0)	(29.2)	(47.6)	(27.6)	(30.8)	(34.7)		

Figure 4. The number of times each face was labeled with a response from each emotion response category in (A) Study 1, (B) Study 2, and (C) Study 3. The emotion categories and facial expressions are ordered in accordance with a structural model based on similarity of pleasantness and arousal. All the “correct” responses are on the center diagonal, and the “errors” are on the other diagonals.

fit this pattern.⁵) Thus, a good first approximation to the order in which emotion labels emerge developmentally is this simple ranking.

Nevertheless, 32.5% of the children did not fit this simple pattern. In Figure 6, we suggest a slightly more complex pattern. Again, *happy* emerges first. For two labels, there were two possible paths. Some children added *angry*, others added *sad*. For those who used three labels, the two paths merged and all the children used *happy*, *angry*, and *sad*. The next step again allows two paths: Some children added *surprised* and some added *scared*. For five labels, the two paths again merged. In the last step, *disgust* was added. A model that allows for these two alternative routes increases the number of children who fit the pattern to 81.3%—a proportion significantly greater ($p < .01$) than the percentage expected by chance. (If, as the number of labels that children used increased, any label was as likely to be added as any other, then 23.3% would fit this more complex pattern.) In addition, a comparison of percentages indicated that 81.3% for the more complex model was significantly greater ($p = .008$) than 67.5% for the simpler model.

This hypothesized developmental order in the use of emotion labels derives from a single task: labeling facial expressions. Other free labeling studies (Harrigan, 1984; Izard, 1971; Markham & Adams, 1992; Wiggers & Van Lieshout, 1985) have analyzed children's responses for accuracy rather than investigating their use of different labels, and therefore it is not clear if their results mirror ours. (Their obtained mean levels of accuracy are consistent with our results.) Observational studies of children's spontaneous use of emotion labels in everyday life offers some support for our observations. For example, Wellman, Harris, Banerjee, and Sinclair (1995) reported, similar to our own findings, that *happy*, *sad*, and *angry* (and, in their study, *scared*) emerged in spontaneous production before *surprised* and *disgusted*. New data are needed to examine this hypothesis in other contexts, such as with children's labels for vocal, behavioral, or contextual cues to emotion.

Nature of Emotion Categories

Our results on children's use of emotion labels raise questions about children's emotion categories. For example, if these labels

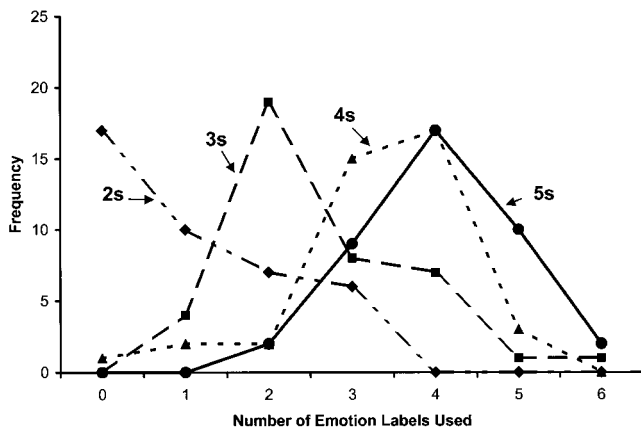


Figure 5. Number of children at each age (2s, 3s, 4s, and 5s = 2-, 3-, 4-, and 5-year-olds) who used a given number of emotion labels in Study 2.

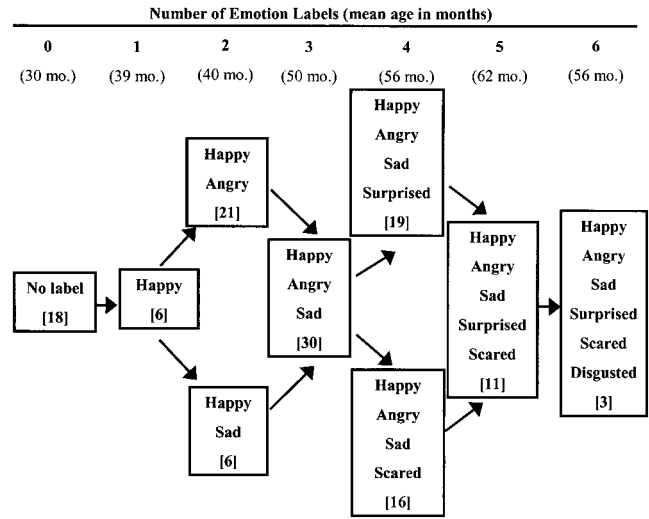


Figure 6. Systematic emergence of emotion labels (Study 2). The number of children who used the specified set of labels is given in brackets [n].

reflect categories, the question arises as to how a child is able to make sense of the emotional world with such a limited number of available categories. One possible answer to this question is Bullock and Russell's (1986) suggestion that the emotion categories of younger children are not equivalent to adults' categories even when children and adults use the same labels. Instead, children's categories are broader. That is, they admit as members (labels are applied to) a broader range of events. At the beginning stages of development, they admit any event of similar valence and level of arousal. At the end, a category is restricted to a specific type of emotion.

The developmental scheme of Figure 6 suggests that the broadest categories would be labeled *happy*, *sad*, and *angry*. To explore this possibility, we again turned to data from Study 2, again because it involved no priming. In that study, children's uses of *happy*, *sad*, and *angry* each followed a similar pattern, illustrated for *happy* in Figure 7A. (The emotions/faces in Figure 7 are arranged along the abscissa in accordance with Bullock and Russell's [1986] structural model, with emotions that are most alike adjacent to one another.)

The 2-year-olds used *happy* (Figure 7A) much less often than older children. But when they did use *happy*, they did so not only for the happy smile but for other expressions as well (indeed, the only exception was the anger face). The 3-year-olds used *happy* much more frequently (.90) for the happy smile. But they also used *happy* to label the surprise and fear faces and a few others. The 4-

⁵ The number of children who would have fit the model by chance alone was calculated by, first, counting the total number of combinations possible for each number of labels (e.g., for two labels used out of the six target labels, there were 15 possible combinations). Next, we divided the number of children who produced that number of labels (e.g., 30 children used two labels) by the total number of possible combinations (i.e., 30/15 = 2). This number was the number expected by chance for that number of labels used. This process was repeated for each number of labels used (from 0 to 6), and the numbers expected by chance for each were added together (32.52) and divided by the number of children in the sample (160).

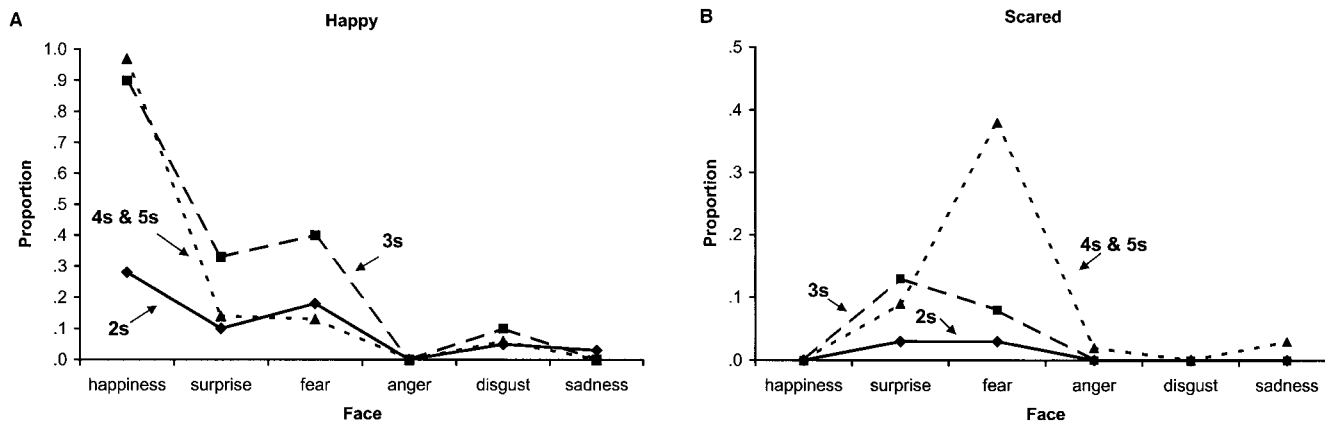


Figure 7. Proportion of children in each age group (2s, 3s, 4s, and 5s = 2-, 3-, 4-, and 5-year-olds) who used (A) *happy* and (B) *scared* for each facial expression in Study 2. Because the 4- and 5-year-olds did not differ, the two groups were combined and the resulting means are presented here.

and 5-year-olds (whose data were so similar that we combined them) used *happy* (.97) for the smile even more frequently, but they used it to label the surprise and fear faces less frequently. Thus, two processes are evident. With age, the probability of using the label *happy* increases. And with age, use narrows so that it is used for the smile more but for other expressions less. Results for *angry* and *sad* showed the same two processes.

Scared, *surprised*, and *disgusted* showed a pattern (illustrated in Figure 7B for *scared*) that is in some ways similar and in some ways different. First, the overall frequency of use of these labels was much lower. The 2- and 3-year-olds rarely used *scared*. When they did, use was restricted to the fear and surprise faces. The 4- and 5-year-olds used *scared* both more frequently and more narrowly than did the 3-year-olds. Again, we see the two processes of increasing probability of use and narrowing of use. The difference is that, even initially, the label *scared* was not used in the very broad way that *happy*, *sad*, and *angry* were. (Because the sheer number of uses of *scared*, *surprised*, and *disgusted* was lower than the number of uses of *happy*, *sad*, and *angry*, our hypothesized pattern for the former must be taken as more tentative.)

CONCLUSION

Asking children to examine a good photograph of a prototypical facial expression and to say what the person shown is feeling turns out to be a more revealing task than commonly thought. It is not, we believe, dominated by simple method artifacts. When children's responses are simply categorized as right or wrong (Harrigan, 1984; Izard, 1971; Markham & Adams, 1992; Wiggers & Van Lieshout, 1985), the main conclusion from this research has been that the task is too difficult for children. But taking a more descriptive approach proved quite interesting. Let us conclude with some thoughts about the development of children's responses to this task.

The lowest level of performance we found (with children past their second birthdays) was a failure to label facial expressions: 18 children in Study 2, 17 of whom were 2-year-olds (mean age = 2 years 6 months), used no labels. Lack of a response is difficult to interpret, but recall that all of these same children had just labeled

four mammals (and were correct on at least two of them). At 2 years of age, emotion label availability is questionable (in Study 3, availability was established for 3- and 4-year-olds, but 2-year-olds were not included in that sample), although other aspects of the data caution against assuming that unavailability necessarily accounts for the 2-year-olds' lack of labeling. Other "method artifact" explanations (poor stimuli or presentation, difficulty of production task) play a minor role at best. Investigation of this performance level, mainly in 2-year-olds, is a high priority for future research.

We believe from other evidence (e.g., Repacholi & Gopnik, 1997) that children at this performance level do interpret faces. That is, they possess mental categories that allow them to find meaning in facial expressions. If they have also acquired verbal labels, those labels have not yet been linked to the mental categories. The nature of those categories and the meaning of those labels, however, remain obscure. For example, Repacholi and Gopnik interpreted their results as showing that 18-month-old children interpret faces in terms of the category of disgust. We believe that their results are also consistent with the hypothesis that their children interpreted faces in terms of broad dimensions such as pleasant–unpleasant.

The next level of performance is the use of one label, most likely *happy* (mean age = 3 years 3 months). Bullock and Russell's (1986) model does not predict this level in children's understanding of emotion. For children at this level, the meaning they associate with *happy* is difficult to characterize. Their responses to animals suggest that they might use broad categories. Perhaps, *happy* at this stage means something very broad, such as "excited" or even "emotional." (Figure 7A shows that 2-year-olds applied the label *happy* to all but one facial expression.) Of course, different children might attach somewhat different meanings to the word. These possibilities, too, merit further research.

The next level is the use of two labels, one positive (*happy*) and one negative (either *sad* or *angry*). This third level (mean age = 3 years 4 months) quickly follows the second. We suggest that this level corresponds to what Bullock and Russell (1986) thought of as a two-category interpretation of emotion. Thus, *happy* means

positive, and *angry* (or *sad*) means negative. If so, the use of two labels stems from children's linking of these labels to their preexisting broad categories based on the pleasant–unpleasant dimension. With a larger sample of children at this level, their use of the word *angry* (or *sad*) could be examined to see if it covers the full spectrum of negative emotions.

The next level is the use of three labels, with negative emotions divided into *angry* and *sad*. This level is achieved on average about 10 months after Level 3 (mean age = 4 years 2 months). Each of these three labels was applied broadly by both younger and older children. This is a level not described in Bullock and Russell's (1986) account, although one interpretation of this level consistent with that account is that *happy* continues to mean positive, whereas negative emotions are divided into those with high arousal (*angry*) and those with low arousal (*sad*). Our results, although from a small sample, suggest a different interpretation. *Angry* was used for both anger and disgust faces, and *sad* for sadness and fear faces.

After, on average, another 6 months, the words *surprised*, *scared*, and *disgusted* are added to the child's lexicon for interpreting facial expressions. Unlike *happy*, *sad*, and *angry*, however, these last three labels are applied much more narrowly from the beginning. Even when in the child's vocabulary, these labels remain relatively low in accessibility. The initial three labels (*happy*, *sad*, and *angry*) continue to be used for the expressions of fear, surprise, and disgust even after these new labels are available.

This result hints at an explanation for the seemingly slow pace at which children improve in producing the "correct" label: They already have a way of interpreting faces that presumably serves reasonably well. Rather than replacing no knowledge with knowledge, children are fine-tuning their way of interpreting faces. This point is illustrated by the results from Study 3 for the disgust face. All of the children (100%) used the label *disgusted* during active priming. Most (93%) also labeled the disgust face. And yet, only 13% labeled the disgust face as *disgusted*. That the label *disgust* enters the children's lexicon before they apply it to the disgust face is a most intriguing finding. When preschool-aged children were shown an array of facial expressions and asked to find the person who felt disgusted, they chose a range of faces (Bullock & Russell, 1984) centered on anger, disgust, and sadness faces. It appears that initially, *angry* and *disgusted* are taken as synonyms, at least with regard to faces, with *angry* the more accessible term. Only later does disgust come to be differentiated from anger. The question therefore arises of just what propels this change.

One possibility is that disgust is not an emotion that is emphasized in the North American culture of our participants—at least not to the degree that other emotions are emphasized. Levy (1973) suggested that different cultures emphasize different concepts and supply a greater variety of words for those that are emphasized more. As a result, these "hypercognized" concepts are easy to talk about and more likely to be transmitted to children at an earlier age. For example, anger has a variety of labels in English: *angry*, *mad*, *frustrated*, *grumpy*, *enraged*, *outraged*, *cross*, *irritated*, *annoyed*, *irked*, and so forth. Shame has a variety of labels in Chinese (Russell & Yik, 1996). Cultures deemphasize other concepts (Levy, 1973), making them less easy to talk about and reducing the likelihood that children will learn about them at an early age. Thus, in English disgust has only a few labels: *disgusted*, *yucky*, and, the more slang label, *gross*. There is evidence (Izard, 1971) that

French children used the French equivalent for disgust (*dégoût*) earlier than American children. The French word *dégoût* is more closely associated with bad tastes and food and may also be easier to acquire than *disgusted*, which also refers to morality and persons as well as foods. Thus, the children we tested may have performed poorly in treating disgust as a distinct emotion because they had not been taught about it as thoroughly or from as early an age as they had been for other emotions. The possibility that children's performance in the current studies was due in part to cultural emphasis invites cross-cultural comparisons.

References

- Baldwin, D. A., & Moses, L. J. (1996). The ontogeny of social information gathering. *Child Development, 67*, 1915–1939.
- 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.
- Boulden, J. (1993). *Feelings and faces: Feelings activity book*. Weaverlyville, CA: Boulden.
- 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: Basic Books.
- 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 preschoolers' 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. 2, pp. 203–237). Cambridge, England: Cambridge University Press.
- Camras, L. A., Grow, J., & Ribordy, S. (1983). Recognition of emotional expression by abused children. *Journal of Clinical Child Psychology, 12*, 325–328.
- Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70*, 205–218.
- Conlin, S. (1991). *All my feelings at preschool: Nathan's day*. Seattle, WA: Parenting Press.
- Crary, E. (1992). *I'm mad*. Seattle, WA: Parenting Press.
- Darwin, C. (1965). *The expression of the emotions in man and animals*. Chicago: University of Chicago Press. (Original work published 1872)
- Denham, S. A. (1998). *Emotional development in young children*. New York: Guilford Press.
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding systems*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. New York: Pergamon Press.
- Gates, G. S. (1923). An experimental study of the growth of social perception. *Journal of Educational Psychology, 14*, 449–461.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.
- Gosselin, P., & Simard, J. (1999). Children's knowledge of facial expressions of emotions: Distinguishing fear and surprise. *Journal of Genetic Psychology, 160*, 181–193.
- Harrigan, J. A. (1984). The effect of task order on children's identification of facial expressions. *Motivation and Emotion, 8*, 157–169.
- Harris, P. L. (1989). *Children and emotion: The development of psychological understanding*. Oxford, England: Basil Blackwell.

- Harris, P. L. (1994). The child's understanding of emotion: Developmental change and the family environment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 35, 3–28.
- Heide, F. P. (2000). *Some things are scary*. London: Walker Books.
- 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*, 2, 288–299.
- Kellogg, W. N., & Eagleson, B. M. (1931). The growth of social perception in different racial groups. *Journal of Educational Psychology*, 22, 367–375.
- Levy, R. I. (1973). *Tahitians: Mind and experience in the Society Islands*. Chicago: University of Chicago Press.
- Magai, C., & McFadden, S. H. (1995). *The role of emotions in social and personality development: History, theory, and research*. New York: Plenum Press.
- Markham, R., & Adams, K. (1992). The effect of type of task on children's identification of facial expressions. *Journal of Nonverbal Behavior*, 16, 21–39.
- Parr, T. (2000). *The feelings book*. Boston: Little, Brown.
- Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, 33, 12–21.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 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., & Yik, M. S. M. (1996). Emotion among the Chinese. In M. H. Bond (Ed.), *Handbook of Chinese psychology* (pp. 166–188). Hong Kong, China: Oxford University Press.
- Sierra On-Line, Inc. (1998). Morph Version 2.5 [Computer software]. Bellevue, WA: Author.
- Suter, K., & Levin, J. (1998). *Baby faces*. New York: Golden Books.
- Szekeress, C. (2000). *Toby's silly faces*. New York: Little Simon.
- Wattenberg, J. (1998). *Mrs. Mustard's baby faces*. New York: Chronicle Books.
- Wellman, H. M., Harris, P. L., Banerjee, M., & Sinclair, A. (1995). Early understanding of emotion: Evidence from natural language. *Cognition and Emotion*, 9, 117–149.
- Who are you? Country friends*. (1992). Newmarket, England: Brimax Books.
- Widen, S. C., & Russell, J. A. (2002). Gender and preschoolers' perception of emotion. *Merrill-Palmer Quarterly*, 48, 248–262.
- Wiggers, M., & Van Lieshout, C. F. M. (1985). Development of recognition of emotions: Children's reliance on situational and facial expressive cues. *Developmental Psychology*, 21, 338–349.

Received October 29, 2001

Revision received August 9, 2002

Accepted August 22, 2002 ■

Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at journals@apa.org and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands.