THE DEVELOPMENTAL COURSE OF CHILDREN'S FREE-LABELING RESPONSES TO FACIAL EXPRESSIONS

by

SHERRILEA E. WIDEN

B.A., The University of British Columbia

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

FACULTY OF GRADUATE STUDIES
(Department of Psychology)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLBUMBIA

September 1999

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology.

The University of British Columbia
Vancouver, Canada

Date 24/09/99

Abstract

The current study investigated the developmental course of how young children label various facial expressions of emotion. 160 children (2 to 5 years) freely produced labels for six prototypical facial expressions of emotion and six animals. Even 2-year-olds were able to correctly label 5 of 6 animals, but the proportion of correct specific emotion category responses for this age group was ≤ .30 for each of the six facial expressions. The 5-year-olds' proportion of correct specific emotion category labels was at ceiling for the happy and angry faces, but significantly lower for each of the other four facial expressions, and at floor level for the disgust face. The type of errors in labeling facial expressions changed with age: when incorrect, the youngest children produced any emotion label; older children produced labels of the correct valence; and the majority of the 5-year-olds' responses were of the correct specific emotion category. These results indicate that the free-labeling task per se is not too difficult even for 2-year-olds, but that children's use of emotion terms is not initially linked to facial expressions. Thus, the children's production of emotion terms far exceeded their proportion of correct specific emotion category labels. With age, children's implicit definition of emotion terms develops to include the associated facial expression, though this process is not complete for all expressions before the age of 6 years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I'd like to thank Dr. Jim Russell for his help, guidance and patience. Thanks also go to Lisa Ferguson for teaching me how to work with children, and to Sarah Knowlden, Melanie Parkin, Amit Taneja, and Paula Castellon for their help with data collection, data entry, proof-reading, and moral support.

The Developmental Course of Children's Free-Labeling Responses to Facial Expressions

This study focuses on how young children (2 to 5 years) label facial expressions. As adults, we associate facial expressions with particular emotions. When, for example, we see a person smiling, we suspect that person is happy; when we see a person frowning, we suspect that person is angry. Do young children think that a smiling face is associated with being happy, or that a frowning face is associated with anger? The question of when and how children label faces with emotion terms is interesting for at least two reasons. First, it provides one possible avenue toward understanding children's acquisition and use of emotion terms. Second, how children label facial expressions provides a possible avenue to understanding their knowledge of emotions in general.

In previous research, children have been shown facial expressions, and asked, "How is this person feeling?" The children's poor performance in free-labeling facial expressions surprised many researchers. Gates (1923), for example, found that the increase in children's proportion of correct responses was very gradual, and that even 14-year-olds accurately labeled facial expressions on fewer than 80% of the trials. Izard (1971) observed that "... not much of an emotion labeling response could be obtained below age six..." (p. 332). Other free-labeling studies (Harrigan, 1984; Wiggers & van Lieshout, 1985; Markham & Adams, 1992) have also found that children's labeling of facial expressions was poor and improved only gradually with age. Commonly, this poorness and gradualness of performance has been attributed to lack of vocabulary and to the free-labeling task itself, but these explanations themselves need to be examined.

The account that follows begins with a discussion of the available evidence on children's spontaneous use of emotion terms; goes on to consider available theory and evidence for children's association of facial expressions with discrete emotions; and ends with a review previous facial expressions free-labeling studies.

Children's Acquisition and Use of Emotion Terms

One common line of thinking holds that children's acquisition and use of emotion terms occurs relatively late in development, whereas children's knowledge and understanding of emotion has its roots in an innate and early-occurring understanding of facial expressions. For example, in his textbook, Plutchik (1994) reported that, "by the age of 3 or 4 years, young children can recognize a number of basic emotions in others, and ... the complex vocabulary of emotion is gradually added much later" (p. 213). When children make errors in associating a facial expression with the standard label (e.g., failing to label a smiling face as *happy*), the error, as seen from their perspective, is seen to be due to young children's lack of vocabulary rather than to their lack of ability in understanding the meaning of the face (Harris, 1983; Field & Woodson, 1982).

This same account further assumes that acquiring knowledge of emotions (including emotion terms) is built on a still earlier comprehension of facial expressions. Giving voice to this view, Denham (1998), for example, wrote, "Comprehension of [vocal and facial] emotional expression can be seen as the perceptual bedrock for further understanding of emotions. As such, it stands preschoolers in good stead, giving them an initial ability to think and talk about emotional issues..." (p. 61).

Despite such claims, there is mounting evidence that children do acquire and use

some emotion terms much earlier than once anticipated – by their second birthday. Brown and Dunn (1991) found that at 24 months of age, children were already using emotion terms, particularly in reference to their own feelings, and that, with age, their use of such terms increased much more rapidly than indicated by the results of free-labeling studies, and that their use of emotion terms expanded to include the emotions of others as well. Similarly, Dunn, Bretherton, & Munn (1987) observed young children's (18 to 32 months) uses of emotion terms in conversations at home, recording instances of the children[1] producing a variety of emotion terms, including *happy*, *sad*, *angry*, *mad*, *surprise*, *afraid*, and *fear*.

Bretherton and Beeghly (1982) found that by the age of 28 months, at least 40% of the children in their study spontaneously produced tokens of *sad, scared*, and *happy*; a third of them produced tokens of *mad*, and a few of the children produced *yucky* (Table 1). On this evidence it would appear that, shortly after their second birthday, many children have acquired and are able to produce a variety of emotion words in conversation. In addition, parental reports of emotion terms these children had acquired by 28 months of age indicated that the proportion of children who understand and could

---

[1] Unfortunately, the proportion of children who produced, for example, one emotion term versus those who produce many emotion terms was not specified in the observational studies reported here (Dunn et al., 1987; Bretherton & Beeghly, 1982; Ridgeway et al., 1985). Therefore, it was not possible for me to be specific about how many of the children were producing emotion terms or even if there was a portion of children who produced very few or many emotion terms.

Table 1

Percentage of Children who Spontaneously Produced Emotion Terms in Observational Studies

	Observational Study				
	Bretherton & Beeghly (1986)	Ridgeway et al. (1985)			
		Age (years)			
Emotion	2	2	3	4	5
Happy	.40	.80	.95	.95	.97
Sad	.50	.62	.89	.92	.92
Angry/mad	.33	.54	.84	.88	.87
Surprised	.03	.28	.60	.75	.83
Scared	.43	.69	.87	.90	.94
Disgusted/ yucky	.13	.03	.19	.25	.34

Note. Ridgeway et al. (1985) reported data for age groups spanning six months. The data reported here were averaged across the two six-month age groups that composed each year.

produce these terms was higher than the proportion of children who spontaneously produced these terms during observation, ranging from .13 for *surprised* to .73 for *scared.*

Ridgeway, Waters, & Kuczaj's (1985) longitudinal study of thirty children from 24 to 59 months of age similarly demonstrated that children's use of emotion terms increased with age more rapidly than indicated by free-labeling studies, and that, for five emotion terms (happy, sad, angry, mad, and scared), production was at ceiling level or near it (> .85) by 4 years of age (Table 1). (Ridgeway et al. apparently did not record instances of *yucky*.[2]) In short, their study indicated that, at least for terms denoting

---

[2] Bretherton and Beeghly's (1982) inclusion of *yucky* as an emotion term, and Ridgeway et al.'s (1985) exclusion of it, draws attention to the problem of synonyms. Does *yucky* really mean the same thing as *disgusting*? What about *angry* and *mad*? This question has important implications for the scoring of responses in a free-labeling task: Which terms should be included in each specific emotion category, and which ones should be excluded? On the one hand, it is possible to be so narrow as to only accept the one target term (e.g., *disgust* but not *yucky*). On the other hand, it is possible to be so broad as to include virtually any term of the correct valence (e.g., *disgust*, *yucky*, *bad*, *unhappy*, etc.). Izard (1971) approached this problem (with his adult sample) by having fourteen people rate the responses produced by at least two participants, and included terms in the specific emotion category indicated by at least eight of the judges (e.g., *smug*, *aversion*, *sarcasm*, *disgust*, *skepticism*, and *repulsion* were considered correct for the "contempt" expression). Gates (1923) liberally added terms to her emotion categories if they were frequently produced by the participants (e.g., *don't like anybody*, *disagreeable proud*, *disgusted*, *jealous*, and *shows off* were all considered correct for the "contempt" expression). The problem is that many of the terms in both Izard's (1971) and Gates'

happiness, sadness, anger, and fear, more than half of children were producing emotion terms in appropriate contexts before their third birthday. Parental reports of the terms these children knew were again higher for each age group than indicated in the observational results.

Two shortcomings plague each of these observational studies (Brown & Dunn, 1991; Dunn et al., 1987; Bretherton & Beeghly, 1982; Ridgeway et al., 1985), at least in terms of their capacity to inform the current investigation. First, due to the observational nature of the studies, it is probable that the reports of the proportion of children who knew and could use the various emotion terms was conservative. That is, it is likely that some of the children knew, for example, the word *mad*, but that no situation arose during observation for them to use it. Thus, although obviously plagued by problems of their own, parental reports of children's emotion vocabulary might provide a more realistic representation of children's emotion vocabulary.

Second, there is no guarantee, when children produce an emotion term, that they hold the same implicit definition of that word as do adults. In an effort to ensure that the children understood the words they were using, the researchers (Brown & Dunn, 1991; Dunn et al., 1987; Bretherton & Beeghly, 1982; Ridgeway et al., 1985) included only those instances of emotion words which were judged to have been produced in appropriate contexts. Nonetheless, there remains the possibility that children's implicit definitions of emotion terms differ from the adult definitions of the same words. There is some evidence, for example, that children's understanding of emotion terms is broader

---

(1923) emotion categories were not synonyms at all.

than adults', and that the meanings of some emotion words is measurably different for children than for adults (e.g., 3- and 4-year-olds' definition of *surprised* is closer to the adult definition of "excited" or "happy" than "surprised") (Bullock & Russell, 1986). Furthermore, it is also possible that children's implicit definition of emotion terms is accurate but incomplete (e.g., they could know the types of situations that cause a particular emotion, but not what facial expression is associated with it – or vice versa).

In the current study, children's use of different emotion words was explored through their application of them to different facial expressions. Clearly this approach does not provide a measure of children's full understanding of different emotion terms. However, a full definition of, for example, *angry* or *sad* might reasonably be expected to include the facial expression commonly associated with that word. Thus, how children label facial expressions is one avenue to assessing at least this aspect of their understanding of the emotion terms they use.

Are Children Able to Associate Facial Expressions with Particular Emotions?

According to an evolutionary line of thought (e.g., Ekman, 1972; Izard, 1971), those members of the species who could both express emotions via facial expressions and who could also understand the communicative intent of such expressions (e.g., understanding that a relaxed stance and a smile indicated happiness and welcome, whereas a tense posture with clenched fists and a frowning face indicated threat and danger) were more likely to survive and pass on their genes to the next generation. Thus, facial expressions are assumed to have evolved as "relatively universal communication systems that promote the survival of humans" (Anderson & Guerrero, 1998, p. 50).

Ekman's (1972) and Izard's (1971) primary assumption was that facial expressions served the evolutionary purpose of emotional communication. They further assumed that recognition and understanding of facial expressions is innate, universal, and central to our concepts of emotion. Apparent support for this theory was found in infants' ability to "adapt their behavior" (Klinnert, Emde, Butterfield, & Campos, 1988, p. 36) to caretakers' facial expressions (Maurer & Barrera, 1981; Barrera & Maurer, 1981; Walker-Andrews, 1988). For example, studies of social referencing have found that, when one-year-olds are faced with an ambiguous situation, they will look to their parent and if the parent looks happy, children proceed, but if the parent looks scared, children do not (Klinnert, Emde, Butterfield, & Campos, 1988).

The assumption that infants understand facial expressions is widespread and current. Harris (1989), for example, cited evidence in support of the possibility that the ability to recognize emotions from facial and vocal expressions is universal and develops quite early: He reported that, "From about 10 weeks, babies appear to react differently and appropriately depending on the emotion that the mother expresses... [and] recognize and respond to the meaning of a caretaker's expression" (p. 19). Magai and McFadden (1995) similarly wrote, "Babies recognize and respond to angry, sad, fearful, and happy maternal expressions with similar expressions. Infants thus acquire considerable information about the world via their ability to read and respond to emotion signals" (p. 204).

The assumption made by Harris and by Magai and McFadden is that, at 10 weeks of age, infants already possess concepts for discrete emotions. That is, infants have a set of mental categories, one for each basic emotion, innately tied to facial expressions. This

assumption is implicit in psychologists' speculation that very young infants and perhaps even new-borns "recognize" a half dozen or so specific emotions from facial expressions (Bowlby, 1969/1982; Izard, 1971; Field & Woodson, 1982; Field, et al., 1983; but see also Kaitz & Meschulach-Sarfaty 1988; Nelson & Dolgin, 1985). Thus, Nelson and de Haan (1997) wrote,

> "[I]t is the face that conveys... affective state.... An infant's recognition of an angry expression displayed by a stranger could facilitate a crying response, which will in turn bring the caregiver to protect the infant. Similarly, an infant's recognition of a happy expression displayed by the caregiver could facilitate the expression of happiness in the infant, which could contribute to the development of the attachment relationship" (p.176).

Both Ekman (1972) and Izard (1971) based much of their research on the premise that each of the basic, fundamental emotions (e.g., happy, sad, angry, surprise, disgust, fear) had corresponding discrete facial expressions. Thus, having isolated prototypical expressions of the basic emotions, Ekman and Izard set out to test the hypothesis that these facial expressions were identified with particular discrete emotions both cross-culturally (Ekman, 1972; Izard 1971) and by young children (Izard, 1971).

There is a large body of evidence, gathered with a variety of tasks, that is purported to show that children do understand facial expressions in terms of discrete emotions. Forced choice has been one of the most commonly employed response formats used to investigate how children identify facial expressions. For example, Gnepp (1983) found that when 3½- to 5-year-olds were asked to identify whether line drawings of facial expressions were happy, sad, or afraid, 92% of the children chose the correct emotion.

Reichenbach and Masters (1983) found that when 4-year-olds were asked to identify whether photographs of facial expressions, posed by four different children, were happy, sad, mad, or neutral, 50% chose the correct emotion. Izard (1971) showed children (2 to 9 years) arrays of three facial expressions, each posed by a different adult, and asked them to choose the one who felt, for example, angry. He found 46% of the 2- and 3-year-olds, 57% of the 4-year-olds, and 66% of the 5-year-olds could choose the correct face. The children's performance increased steadily with age, with performance rising to 84% for the 9-year-olds.[3]

A variety of other methods have been employed to test children's understanding of emotion and facial expressions (Camras & Allison, 1985; Bullock & Russell, 1986, 1985, 1984; Russell & Bullock, 1986a, 1986b; Russell, 1989, 1990; Markham & Adams,

---

[3] The difference in the children's performance in these three forced choice studies (Gnepp, 1983; Reichenbach & Masters, 1983; Izard, 1971) may have been due, in part to the stimuli used in each study. Gnepp's stimuli were line drawings of a child with a smiling, crying, or scared expression. Both Reichenbach and Master's and Izard's stimuli were photographs of people posing facial expressions. Reichenbach and Master's models were four different children (4- to 5-years-old). Izard's models were adults; in addition, his task was reversed with the children identifying the face that was expressing the target emotion. Although, to my knowledge, there has been no systematic investigation of how these different types of facial expression stimuli affect children's responses, it seems probable that the complexity and realism of the stimuli would affect children's performance.

1992). In many cases, due to the young ages of the children or concerns about the difficulty of production tasks, the response format was nonverbal or minimally verbal (e.g., yes/no responses). Examples of these tasks include sorting tasks (Russell & Bullock, 1986a, 1986b), matching tasks (Russell & Widen, 1999), and tasks on which children are asked to indicate their responses using special scales (Stifter & Fox, 1987).

Markham and Adams (1992) investigated four tasks frequently employed in facial expressions studies with children: free labeling, forced choice, situation discrimination (choose the facial expression that matches situation and emotion term), and matching discrimination (match target face with one from response set). The stimuli were eighteen photographs of six facial expressions (happy, sad, surprise, anger, fear, disgust). Children's (4, 6, and 8 years) performance on each task improved with age. Children in each age group performed well on the situation discrimination, matching discrimination, and forced choice tasks, though performance in each task varied with facial expression. The children's performance on the free-labeling task was lower than their performance on any of the other three tasks. (Markham and Adams' free-labeling results are discussed further in the next section.)

In short, children perform well on most tasks involving facial expressions (the major exception being free labeling) and their performance has generally been interpreted as indicating that children understand facial expressions as communicating discrete emotions. Does the evidence warrant this interpretation? Not entirely. First, the theory underlying this research – namely that what facial expressions communicate are emotions -- has been questioned. Fridlund (1994) suggested that, rather than communicating emotions, facial expressions predict behaviors. For example, a person who is frowning

and whose lips are compressed or whose teeth are exposed (i.e., Ekman's [1976] hypothesized facial expression of anger), may be about to attack; a person with a down-turned mouth and eyebrows pulled together with the inner corners raised (i.e., Ekman's [1976] hypothesized facial expression of sadness), may be about to ask for help. Thus, while such expressions may also be correlated with particular emotions, Fridlund (1994) points out that facial expressions are not necessarily communicating that emotion, but that these expressions allow those around us to know how we are likely to behave. Thus, a task in which children choose among three emotions presupposes, rather than tests, the theory that children understand faces in terms of emotion.

Second, the available evidence is weak. For example, while social referencing studies have shown that 1-year-olds respond differently to expressions of happiness than to expressions of fear, there is no evidence, for example, that they respond differently to fear expressions than to angry expressions. That is, while it is evident that these children react differently to expressions of opposite valence, it is not clear that they react differently to expressions of the same valence. (This possibility is consistent with Bullock and Russell's [1986] evidence from sorting and other minimally verbal tasks that young preschoolers [2 and 3 years of age] interpret facial expression primarily in terms of valence and arousal, rather than in terms of discrete emotions.)

The evidence that children recognize facial expressions provided by other tasks is also weak and open to alternative interpretations. Take, for example, a forced choice task (e.g., Markham & Adams, 1992; Gnepp, 1983; Reichenbach & Masters, 1983), in which children have to choose the correct label from the three the experimenter offers for the facial expression. First, the forced choice method has been criticized as overestimating

children's degree of recognition of facial expressions in terms of discrete emotions (Russell, 1994). For example, given three labels (e.g., *happy*, *sad*, and *angry*) and the ability to discriminate between positive and negative expressions, with no further knowledge of facial expressions a child could achieve 50% recognition by chance; and 50% was not an unusual finding for preschoolers (Reichenbach & Masters, 1983; Izard, 1971). Second, the children's forced choice performance increases very gradually with age. This is a curious finding because if children innately recognize facial expressions in terms of discrete emotions then their performance should be near excellent and asymptotic from a very early age. Finally, forced choice is usually embedded in a paradigm that might teach children the correct answer. Thus, if on one trial they select *happy* for a smiling face, then the other expressions are implicitly "not-happy". On the next trial, if the face is not smiling, they can eliminate the *happy* from the list the experimenter offers. Thus, as they move through the trials, the children may remember whether they saw an expression earlier; if so, they can apply the same label to it; if not, they can eliminate from the choices the experimenter offers any that they have applied to other expressions.

Similar criticisms are possible for the situation discrimination task (e.g., Markham and Adams, 1992), in which the experimenter describes a short scenario and labels the target emotion, and the child has to choose the correct facial expression from an array of three photographs. In addition, a child might associate a particular scenario (e.g., a gift) with a facial expression (e.g., a smile) without associating either with a discrete emotion.

In addition to such criticisms, there is evidence that children's emotion concepts and their judgements of facial expressions are not discrete, but that they are initially

based on broad dimensions, and that only with age do they approach the adult standard. Bullock & Russell's (1984, 1985, 1986; Russell & Bullock, 1986a, 1986b) fine-grained analysis of children's "errors" found that children do have a conceptual system, based on the dimensions of pleasure and arousal, for understanding facial expressions from a very young age.

On Bullock and Russell's (1984) view, even when "accurate" in standard tasks, 2- and 3-year-olds did not interpret facial expressions in terms of the same categories implied by the adult words *anger*, *fear*, and the like. Instead, children's categories begin quite wide, encompassing any expression that has similar levels of pleasure and arousal. Thus, when two-year-olds were asked to identify facial expressions that were angry, they included angry, disgust, scared, and sad faces (i.e., any unpleasant facial expression). Older children's categories were narrower until, among the oldest children tested their ability to discriminate different expressions in terms of particular emotion was adult-like and discrete (Bullock & Russell, 1986). These findings have recently been replicated with a sample of deaf children (Hosie, Gray, Russell, Scott, & Hunter, 1998)

Similarly, Bormann-Kischkel, Hildebrand-Pascher, and Stegbauer (1990) replicated Bullock and Russell's (1984, 1985, 1986) studies with a sample of German children. They also found that young children's understanding of facial expressions was based on the dimensions of pleasure and arousal. In addition, they found evidence for a third dimension based on the positions of the facial features themselves. Thus, children's errors tended to occur more frequently for those expressions that shared, for example, faces with open mouths and exposed teeth or faces with closed mouths.

To summarize, the standard account of children's understanding of emotion and facial expressions has much evidence to cite in its favor, and therefore has been taken for granted. Nevertheless, the assumption has been that if children find any meaning in such facial expressions, it must be in terms of discrete emotions. A workable alternative explanation for findings that appear to support the conclusion that infants and young children understand facial expressions in terms of discrete emotions is that they understand the valence of facial expressions before they understand the discrete categories associated with them. Given that the results of social referencing, forced choice, and situation discrimination tasks are equally well explained by both alternatives, it is important not to take it for granted that children, especially preschoolers, understand facial expressions as discrete categories.

Children's Performance on Free-Labeling Facial Expressions

In the standard free-labeling procedure, the experimenter shows a child the to-be-named stimulus (e.g., a photograph of a posed prototypical facial expression of emotion) and asks a child to label it (e.g., "How is this person feeling?"). The standard free-labeling facial expressions task brings together children's use of emotion terms and their understanding of facial expressions. Ideally, the child is not given additional cues as to the target ("correct") response. The child's task is simply to label the stimulus. Results based on this task has sometimes been dismissed (perhaps because results with it are inconsistent with the taken-for-granted theory and seemingly with other evidence). Nevertheless, free labeling has advantages over other tasks. In particular, because the experimenter does not force the children to choose among preselected options, such a

procedure allows us a glimpse of the ways children apply emotion terms spontaneously and to investigate whether current theories account for these applications.

The use of the free-labeling procedure to investigate how children label facial expressions dates back to the 1920s (Gates, 1923), and in the intervening decades, children's performance on this task (Izard, 1971; Harrigan, 1984; Wiggers & van Lieshout, 1985; Markham & Adams, 1992) has been the same: relatively low, with gradual improvement with age (Figure 1). For example, the 2-year-olds' proportion of correct responses was less than .10 in Izard's (1971) study. Izard's 3-year-olds faired somewhat better (.18), but improvement with age was only gradual and less than half of the responses made by even the oldest children he tested (9-year-olds) were correct. While some of the other researcher's (Markham & Adams, 1992; Wiggers & van Lieshout, 1985) free-labeling results were somewhat higher, the proportion of correct responses given by preschoolers (< 6 years) has never exceeded .60.

The increase in children's proportion of correct responses in the free-labeling task remained gradual whether children performed the free-labeling task first (Gates, 1923; Harrigan, 1984; Wiggers & van Lieshout, 1985) or after another facial expression task (Izard, 1971; Harrigan, 1984; Markham & Adams, 1992). However, the priming provided by another task may have tended to inflate estimates of children's understanding of facial expressions in terms of discrete emotions. That is, another task – forced choice, for example – alerts children to the type of responses the experimenter is looking for (i.e., emotion terms), and it also provides them with the opportunity to make explicit associations between the facial expressions and the emotion labels. In both Harrigan's and Markham and Adams' studies, children's free-labeling of facial expressions

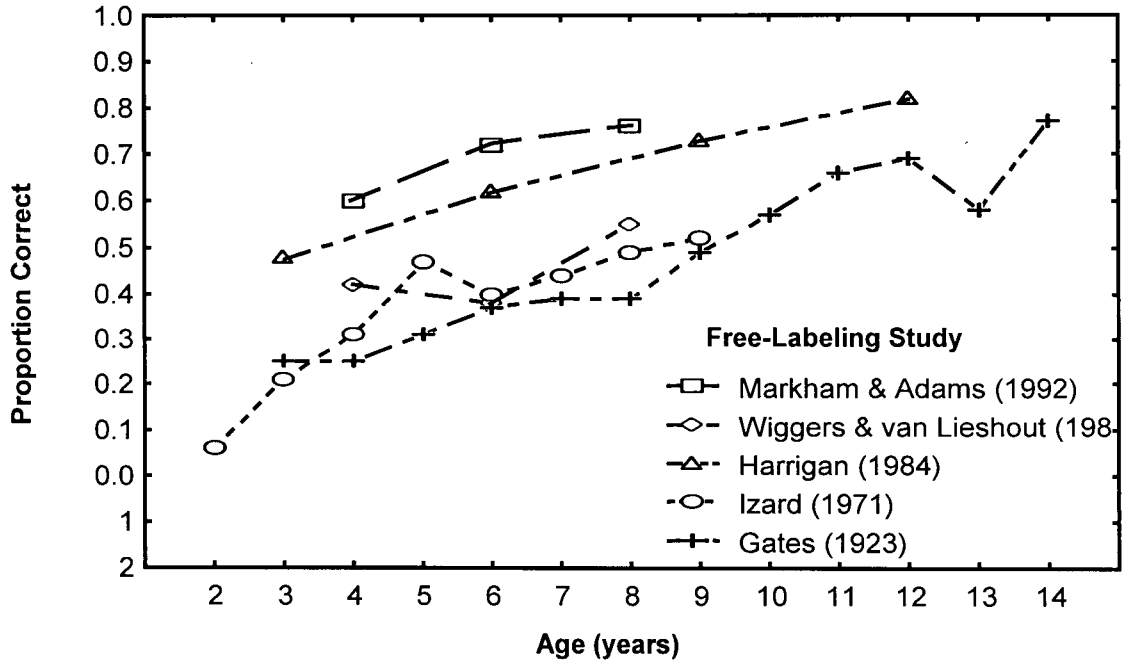


Figure 1. Mean proportion correct for each age group in Markham and Adams' (1992), Wiggers and van Lieshout's (1985), Harrigan's (1984), Izard's (1971), and Gates' (1923) free-labeling studies.

increased significantly when the free-labeling task was preceded by another facial expressions task over than when the free-labeling task was administered first. Thus, it is possible that children's free-labeling responses to facial expressions following another task are as much a result of performing that other task first as they are a result of children's actual associations between the expressions and discrete emotions.

Children's poor performance in free-labeling facial expressions was not uniform across all expressions. Instead, their accuracy varied with facial expression (Table 2). While the preschoolers' proportion of correct labels for the individual expressions varied in magnitude between the three studies that offered this information (Gates, 1923; Izard,

1971, Markham & Adams, 1992), the overall trend is the same. The preschoolers labeled the happy face well, and some of them also labeled the sad and angry expressions accurately, but the proportion of children who labeled the other three expressions was lower. All-in-all, while results of the free-labeling studies varied in absolute proportion correct, they each found that children's proportion of correct responses increased gradually with age and the same order of difficulty: happy was accurately labeled earliest, followed by sad and angry; then scared; surprised and disgusted were accurately labeled least often (Gates, 1923; Izard, 1971; Harrigan, 1984; Wiggers & van Lieshout, 1985; Markham & Adams, 1992).

Izard (1971) found children's poor free-labeling performance surprising, but ascribed the children's difficulty to the free-labeling task itself. Other researchers have also offered this explanation for poor performance. Markham and Adams' (1992) purpose was to compare children's performance on four tasks (free labeling, forced choice, situation discrimination, matching discrimination) in order to establish the relative difficulty of each. Task difficulty is one possible explanation of children's performance in free labeling, but it is an explanation that bears further investigation. In addition, no free-labeling study to date has taken advantage of the children's freedom of response by examining their "incorrect" responses.

Table 2

Proportion of Correct Responses for Young Children in Markham and Adams' (1992), Izard's (1971), and Gates' (1923) Free-Labeling Studies

	Free-Labeling Study							
	Gates (1923)			Izard (1971)				Markham & Adams (1992)
	Age (years)							
Face	3	4	5	2	3	4	5	4
Happy	.70	.70	.84	N/A	N/A	N/A	N/A	.99
Sad/ Distress/ Pain	.40	.40	.44	.00	.20	.21	.41	.77
Angry	.30	.40	.29	.12	.46	.62	.88	.77
Surprise	.10	.00	.00	N/A	N/A	N/A	N/A	.67
Disgust/ Contempt	.00	.00	.00	N/A	N/A	N/A	N/A	.47
Scared	.00	.05	.13	N/A	N/A	N/A	N/A	.63
Mean	.25	.26	.28	.06	.21	.31	.49	.60

Note. Proportions for Markham and Adams (1992) and Izard (1971) were estimated from graphs. Izard (1971) provided proportion of correct responses by age for only the distress and angry faces; the data for the other faces were not available (N/A) but the means for each age are for all six facial expressions.

The Study

The current study was designed to investigate the developmental course of how children label facial expressions and to replicate the results of the previous free labeling studies. The objective of the current study was to offer a more definitive exploration of how children free label. Changes in the procedure were introduced to increase the possibility that the children would perform optimally. I also included a comparison free-labeling task, in which the children were asked to label animals. In addition, rather than rating the children's responses as simply accurate or inaccurate, they were also rated in terms of the valence of their responses.

In the current study, in order to take a baseline measure of how children label facial expressions, the children's emotion concepts were not primed. Thus, without prior priming of their emotion concepts, children between the ages of 2 and 6 years were asked, "How is Sally feeling in this picture?" The term *feeling* was used in this question because the children clearly had to be asked something (i.e., simply showing them a picture was unlikely to elicit any kind of response) and "feeling" was the broadest relevant term available. In addition, *feeling* was the term that Gates (1923), Izard (1971), and Markham and Adams (1992) used in their free-labeling procedures.

In order to optimize the chances that the children would produce the correct label, I used new stimuli and introduced two additional steps in the procedure. The stimuli were photographs of facial expressions that were designed in accordance with Ekman & Friesen's (1976)[4] specifications (Camras, Ribordy, & Grow, 1983). Ekman-type facial

---

[4] Ekman and Friesen (1976) identified the facial expressions for each "basic" emotion

expressions have been tested extensively with adults and children, and with people from different cultures, and have been shown to be associated with their target emotion in a variety of testing procedures. Markham and Adams' (1992) used photographs of facial expressions developed by Ekman and Friesen (1976) and their free-labeling results were higher than in the other free-labeling studies. Given the extensive testing of these expressions, these stimuli do seem to offer the highest likelihood of being associated with the target emotion.

The procedural changes I introduced to my study were intended to make the children feel more comfortable and thus more willing to respond. The first change required the experimenter to spend time getting to know each child before testing. Thus, the experimenter did not attempt to do the procedure until the child seemed relaxed and comfortable in the experimenter's company. The other procedural change I introduced was an animal labeling trial, which preceded the facial expressions trial. The purpose of this trial was to introduce the children to the free-labeling procedure so that they would be more comfortable responding to the experimenter's questions in the facial expressions

---

(happy, sad, angry, surprised, disgusted, scared) that received the highest consensus in forced choice studies, and then identified the positions of the various facial muscles in each using their Facial Action Unit coding system. Using this system, models could create the precise desired expressions by learning to voluntarily control their facial muscles. Ekman's photographs are head-only shots, and the gaze direction is generally toward the camera; thus, there may be less extraneous information (e.g., head position, gaze direction, etc.) in these photographs.

trial.

The animal labeling trial was introduced for another important reason: to investigate whether children as young as 2 years of age were able to perform the free labeling task itself. Evidence from studies of how young children label other stimuli (e.g., Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; Johnson 1992; Johnson & Clark, 1988) indicates that children are quite capable of producing labels. While these studies have found that the number of labels an object has and how common the object is affects children's reaction time and accuracy, their general finding has been that children are able to produce labels of the correct category (e.g., Cycowicz et al., 1997, report that some children called the turkey a *chicken*, or a *rooster*), if not the correct instance label. Thus, by the age of 4 years, children are able to name objects in pictures from such diverse categories as animals, human body parts, clothing, fruit, and furniture (Johnson & Clark, 1992).

If the free-labeling task per se is too difficult for a child, then his or her performance should be uniformly poor for animals and facial expressions. Having the same children label animals and facial expressions should also provide clues about the kinds of limits on the task per se and the nature of the types of responses children make when they are incorrect.

In order to identify the developmental course of how children label facial expressions it is necessary to investigate the nature of the errors they make. Previous free-labeling studies (Gates, 1923; Izard, 1971; Harrigan, 1984; Wiggers & van Lieshout, 1985; Markham & Adams, 1992) concerned themselves only with how accurately children labeled the facial expressions, and thus only rated responses in terms of emotion

category. Although this method enables an investigation of how adult-like children's free-labeling responses are, it is difficult to make any other inferences about their performance. In the current study, the children's responses to the facial expressions were rated for valence (positive, negative) and for emotion category (happy, sad, angry, surprised, disgust, scared). The valence rating allowed the evaluation of the types of errors children made at different ages. The category ratings allowed the more traditional analysis for accuracy.

Finally, each age group in the current study was evenly divided by sex in order to investigate sex differences in how children label facial expressions. Some studies find no sex differences in children's recognition of emotions (Boyatzis & Satyaprasad, 1994; Camras & Allison, 1985), whereas others do find small sex differences (Boyatzis & Chazan, 1993; Russell, 1990). No prediction was made regarding sex differences.

Method

Participants

One hundred and sixty children (80 girls, 80 boys) from twenty-six daycare centers and preschools in or near Vancouver, British Columbia participated in the study. All children were proficient in English (as indicated on the permission form by the parent, by the daycare workers' opinion, and by the experimenter's opinion of the child's fluency based on conversations with the child).

There were 40 children (20 girls and 20 boys) in each of four age groups: 2s (2;0-2;11), 3s (3;0-3;11), 4s (4;0-4;11), and 5s (5;0-5;11).

Photographs of Facial Expressions

The photographs were six 5" x 7" black and white glossy prints of prototypical facial expressions. Each photograph (one each: happy, sad, angry, fear, disgust, surprise) was of the same 12-year-old girl (Figure 2). Each photograph was placed on a separate page of a photo-album. The photographs were provided by Dr. Linda Camras. Camras et al. (1983) describe 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.

Design and Procedure

Each child participated in two trials. The first concerned labeling pictures of animals; the second concerned labeling facial expressions.

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 she brought with her.

The first trial, in which the child was asked to identify six different animals (cat, dog, horse, cow, turkey, goose), was designed as a practice trial as well as as a comparison task. The experimenter said, "This is my special book about animals. [Opening the book and pointing to the first page] Do you know what kind of animal this is?" All responses were treated as correct and were praised. The purpose of this trial was both to introduce children to the task of producing labels, and to investigate how they freely label non-facial-expression stimuli.

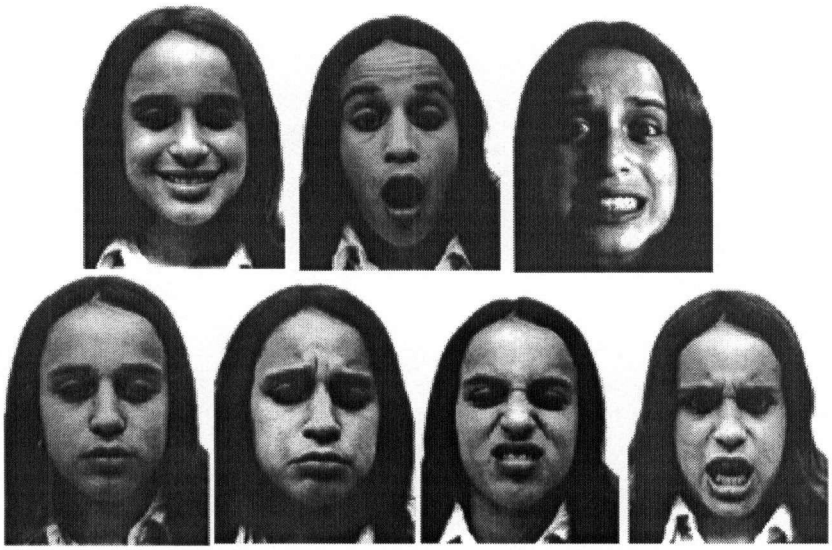


Figure 2. The facial expressions used in the current study, provided by Camras et al (1983).

For the second trial, the experimenter introduced the photo-album with the pictures of Sally. The order of the photographs was varied, though the first expression was always the smiling photograph, because piloting revealed that even young children frequently labeled this expression accurately. The experimenter began by saying, "That was really fun. Do you want to look at my second book? This is my book about a girl named Sally. This is a picture of Sally [neutral expression]. Do you know what Sally is going to do? She is going to show us how she feels sometimes. Are you ready? On the first day, Sally felt like this [pointing to the first picture]. How do you think Sally is feeling in this picture?" All responses were treated as correct and praised. 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 picture. At no time did the experimenter use the

word *emotion* or otherwise direct the child to try to use an emotion label beyond asking how Sally was feeling.

Scoring of Responses

Animals

Due to clerical error, 5 participants' responses to the six animal pictures were not recorded. Thus, collectively, the 155 children had 930 opportunities to label an animal. Of these 930 opportunities to respond, there were 46 different response types, excluding only uninterpretable- and non-responses (e.g., *must be a horn*; *I don't know*; the child said nothing). Two raters made two judgements: (a) broad category, and (b) specific category.

Broad 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*, *not a bird*). The two raters agreed as to broad category for 98% of the response types. Disagreements were resolved by discussion.

Specific category. For the specific category ratings, the raters' task was to indicate into which one of seven animal categories each response fit (cat, dog, horse, cow, turkey, goose), or if it did not fit into any of these categories. The labels included in each specific category are listed in Appendix A, Table A1. The two raters agreed as to 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 yielded 153 different types of responses, excluding only uninterpretable- and non-responses (e.g., *I dunno*; *I don't want to*; *I don't know these pictures*; the child said nothing). Two raters made two judgements: (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.

Valence. 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*), or uninterpretable in regard to valence (e.g., *fine*, uninterpretable responses such as *a harebrain*, *Mommy*, and *turn the page*). Responses to the surprise face where not rated in terms of valence, because it is a state of high arousal and neutral pleasure and thus has no valence. The two original raters agreed as to the valence for 85% of the response types. In cases where the two original judges disagreed, the 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.

Specific emotion category. 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, scared), or if it was uninterpretable in regard to these categories. The labels that were included in each category are listed in Appendix B, Table B1. The two original raters agreed as to 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 specific emotion category.

Results

In order to better understand children's skills with, and approach to, the free-labeling task, I first analyzed the children's performance with the animal pictures. I then analyzed the children's performance with the pictures of faces. First, two parallel repeated measures ANOVAs were calculated on the comparison animal-labeling task (specific category, broad category) and on the face-labeling task (specific emotion category, valence). Second, the types of responses the children made were analyzed. Finally, the underlying dimensions of the children's responses to the photographs of facial expressions were investigated with multidimensional scaling.

Animal Labeling

Two parallel repeated measures ANOVAs were calculated (alpha = .05) on the children's freely produced labels for the animal pictures. Age (four levels) and sex (two levels) were between-subjects factors, and animal (cat, dog, horse, cow, goose, turkey) was the within-subject factor. The dependent variable was whether the label was correct or not, scored 1 or 0.

Specific Animal Categories

In the first analysis, the labels were scored as correct if they matched the target specific animal category. There were significant main effects for age $F(3, 147) = 5.85$, $p < .001$, and animal, $F(5, 735) = 192.90$, $p < .001$. The age x animal interaction was also significant, $F(15, 735) = 2.95$, $p < .001$, as was the sex x age x animal interaction, $F(15, 735) = 2.14$, $p = .01$. There were no other significant effects with alpha set at .05.

The significant main effect for age (Table 3, column means) indicated that as age

Table 3

Effect of Age and Stimulus on Children's Production of Labels for Animals

Animal	Age (years) 2	3	4	5	Mean
	Proportion of Correct Specific Category Labels				
Cat	.97 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	.99
Dog	.97 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	.99
Horse	.97 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	.99
Cow	.85 $_{a}$	1.00 $_{a}$	.95 $_{a}$	.97 $_{a}$	.94
Turkey	.21 $_{b}$	.18 $_{b}$	.25 $_{b}$	.54 $_{d}$	.29
Goose	.77 $_{c}$	.90 $_{ac}$	.80 $_{c}$	.89 $_{ac}$	.85
Mean	.79	.85	.84	.90	
	Proportion of Correct Broad Category Labels				
Cat	.97 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	.99
Dog	.97 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	.99
Horse	.97 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	1.00 $_{a}$	.99
Cow	.98 $_{a}$	1.00 $_{a}$	.98 $_{a}$	1.00 $_{a}$	.99
Turkey	.85 $_{b}$	.70 $_{c}$	.58 $_{d}$	.76 $_{bc}$	.72
Goose	.92 $_{ab}$	.95 $_{a}$	.95 $_{a}$	.97 $_{a}$	.95
Mean	.94	.94	.92	.95	

Note. Alpha = .05 in all Tukey's Least Significant Difference (LSD) comparisons. For the specific category analysis, means in the same row that do not share a subscript differ at $p < .001$. Means in the same column that do not share a subscript differ at $p < .03$.

LSD comparisons on the main effect for age (column means) indicated that the 2s' proportion correct was significantly lower ($p \leq .05$) than every other age group; the 3s' proportion correct did not differ significantly from the 4s', but was significantly lower ($p = .04$) than the 5s'; the 4s' proportion correct was significantly lower ($p = .02$) than the 5s'. LSD comparisons on the main effect for animal (row means) indicated that the proportion correct for the turkey picture was significantly lower ($p < .001$) than for any of the other animal pictures; the proportion correct for the goose picture was significantly lower ($p < .01$) than any of the mammal pictures, which did not differ significantly from each other. For the broad category analysis, means in the same row that do not share a subscript differ at $p < .03$. Means in the same column that do not share a subscript differ at $p < .02$. LSD comparisons on the main effect for animal (row means) indicated that the proportion correct for the turkey picture was significantly lower ($p < .001$) than for any of the other animal pictures, which did not differ significantly from each other.

increased the children correctly labeled the six animal pictures more frequently.

The main effect for animal (Table 2, row means) was due to the children's difficulty in correctly labeling the turkey, and, to a lesser extent, the goose. For the other four animals (cat, dog, horse, cow), the proportion of correct labels produced was $\geq .94$.

Both main effects must be integrated in light of the age x animal interaction (Table 3, cell means). The effect of age was different for the two types of animals: even the 2s' performance was at or near ceiling for the four mammals (cat, dog, horse, cow), but there was improvement with age for the two bird pictures (Figure 3, correct specific category). For the goose picture, the proportion of correct specific category labels

produced by each age group was lower than for the mammal pictures, but higher than for the turkey. For the turkey picture, the 2s', 3s', and 4s' proportion of correct labels was very low (≤ .25), and even the 5s' performance for the turkey picture was significantly lower ($p < .001$) than every cell for each of the other five animals (Tukey's Honestly Significant Difference comparisons, alpha = .05).

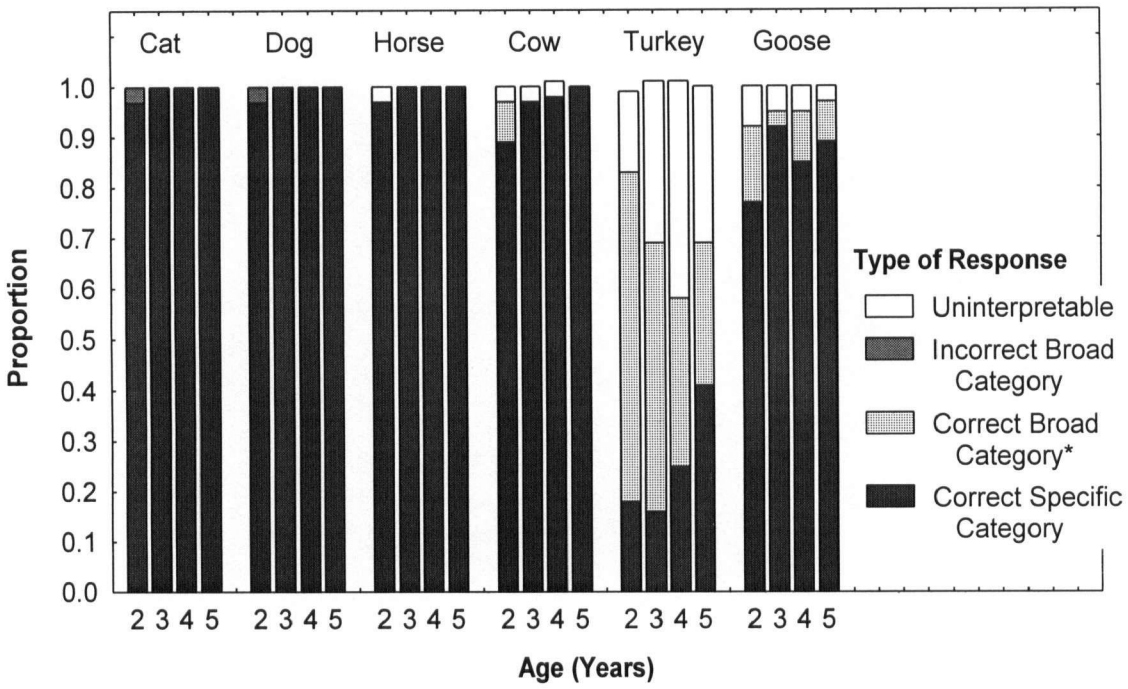


Figure 3. Proportion of responses to the animal pictures that fit into each of four mutually exclusive categories (correct specific category; *correct broad category but incorrect specific category; incorrect broad category; uninterpretable- or non-response).

For the sex x age x animal interaction, both the boys and the girls in each age group performed at ceiling level for the cat, dog, and horse pictures. For three of the forty-eight cells there was a significant difference between sexes (Tukey's Least Significant Difference, alpha = .05), but these differences did not favor either sex

consistently.

Overall, the analysis of specific category ratings indicated that children as young as two years of age freely produced labels in response to animal pictures. Furthermore, even the 2s' responses were correct almost 100% of the time for the four mammal pictures, and they were correct more than three-quarters of the time for the goose picture. The turkey picture was labeled correctly least frequently by each age group.

Broad Category

In this analysis, the labels were scored as correct if they matched the target broad category (mammal, bird). There were significant main effects for sex, $F(1, 147) = 4.56$, $p = .03$, and animal, $F(5, 735) = 44.28$, $p < .001$. The age x animal interaction was also significant, $F(15, 735) = 2.35$, $p = .003$, as was the sex x animal interaction, $F(5, 735) = 2.18$, $p = .05$. There were no other significant effects with alpha set at .05.

For the main effect for sex, the boys produced more labels of the correct broad category (.96) than did the girls (.92). Nonetheless, the vast majority of the boys and the girls produced animal labels of the correct broad category.

The main effect for animal (Table 3, column means) was due to the children's very high proportion of correct broad category labels (.99) produced for the cat, dog, horse, and cow pictures, as well as for the goose picture (.95), while their proportion correct for the turkey picture was lower (.72).

The age x animal interaction (Table 3, cell means) indicated that even the 2s produced correct broad category labels for each of the six of the animal pictures ($\geq .84$) (Figure 3). The older children's performance was also excellent ($\geq .95$) for five of the six animals (cat, dog, horse, cow, goose). Interestingly, the proportion of bird labels

produced for the turkey picture did not increase with age (Figure 3, correct broad category). Instead, the proportion of bird labels produced by the 2s was quite high, while the 3s' proportion correct was lower, the 4s' proportion correct was lower yet again, and the 5s proportion correct was on a par with the 3s' proportion correct. This seemingly anomalous finding is explored below in Types of Responses.

The sex x animal interaction was due to the boys (.78) producing a significantly higher ($p < .001$) proportion of correct broad category labels for the turkey picture than did the girls (.65).

Overall, even the 2s' responses were of the correct broad category on more than .90 of the trials for five of the six animals (cat, dog, horse, cow, goose), as were the older children's. For the turkey, for which the children produced the fewest correct specific category labels, a large proportion of the labels they produced were of the correct broad category. Thus, the children identified the correct broad category for each animal and produced correct specific or broad category labels for each.

Types of Responses (Animals)

Another way of investigating children's free-labeling responses is to consider all the types of responses they made, whether correct or not. Thus, the children's responses to the animal pictures were re-coded into four mutually exclusive categories: correct specific category; correct broad category but incorrect specific category; incorrect broad category; and uninterpretable- or non-response. The analysis of the types of responses enabled the investigation of all the responses the children made, rather than of only their correct specific or broad category responses as was the case for the ANOVAs. Based on

the re-coded data, the difference between the specific category ratings and the broad category ratings is evident. In addition, the proportion of incorrect broad category responses and of uninterpretable- or non-responses can be considered separately. See Appendix A, Table A2 for the proportion of responses to each animal picture that fit into each animal category for each age group.

For the four mammals, even the 2s produced a very high proportion of correct specific category labels (≥ .89), and their proportion of correct goose labels was quite high, with an additional portion producing correct broad but incorrect specific category labels (Figure 3). Thus, for these five animals, the 2s' proportion of correct broad or specific category labels was very high.

For the turkey, although the proportion of correct specific category labels increased with age, the proportion of correct-broad-category-but-incorrect-specific-category labels did not, as noted in the age x animal interaction for the broad category ratings. Specifically, the 2s produced a high proportion of correct broad category but incorrect specific category labels for the turkey picture, while the 3s, 4s, and 5s each produced fewer correct broad category labels for this picture (Figure 3). This result suggests that even the 2s recognized the turkey as a member of the bird category and were willing to offer labels from that category in response to the picture. The three older age groups also recognized that the turkey was a bird, as indicated by the fact that they did not offer any incorrect broad category labels (e.g., none of them called the turkey a "fish" or a "tiger"). However, their production of fewer other-bird labels for the turkey picture suggests that, with increasing age, more of the children recognized that the turkey was not one of the birds for which they knew a label and thus offered no label at all.

Multidimensional Scaling of Children's Animal Labels

The multidimensional scaling analysis was calculated on the similarity of the animals. In essence, the question I asked in creating the similarity matrix for multidimensional scaling was, "When the children labeled this animal, what animal labels did they produce?" Similarity for any pair of animals was taken as how often they were confused for each other. Specifically, for any pair of animals, similarity was measured by counting the number of times the "correct" label for one of the pair of animals was applied to the other, and vice versa. For example, the similarity of the cat and the dog was determined by counting the number of times the dog was labeled "cat" plus how often the cat was labeled "dog". Because there were very few incorrect responses for five of the six animals (cat, dog, horse, cow, goose), the analysis was calculated on the responses of all the children as a single group.

The measure of interest is "stress," a badness-of-fit statistic which indicates how poorly the relationship of objects is described by that number of dimensions. Thus, when stress is low, the fit is good. For the one-dimensional solution, stress was very low (.09), indicating that there was one dimension underlying the children's responses to the animals. I interpreted this dimension as type-of-animal (mammal, bird). This interpretation was based on children's ceiling level of correct specific category responding to the mammal pictures, with the few errors they made falling into the mammal category, and from their responses to the bird pictures, for which the errors fell into the bird category.

Facial Expression Labeling

After a few brief descriptive comments regarding the children's proportion of labels in the facial expressions labeling trial, the results of two parallel ANOVAs calculated on children's proportion of correct free-labeling responses to the facial expressions are reported. Then, the children's responses to the facial expressions were re-categorized into four mutually exclusive categories (correct-specific-emotion-category, correct-valence-but-incorrect-specific-emotion-category, incorrect valence, uninterpretable or non-response), and the types of responses that the children made to each face were analyzed. Finally, the underlying dimensions of children's responses to facial expressions were explored in a multidimensional scaling analysis.

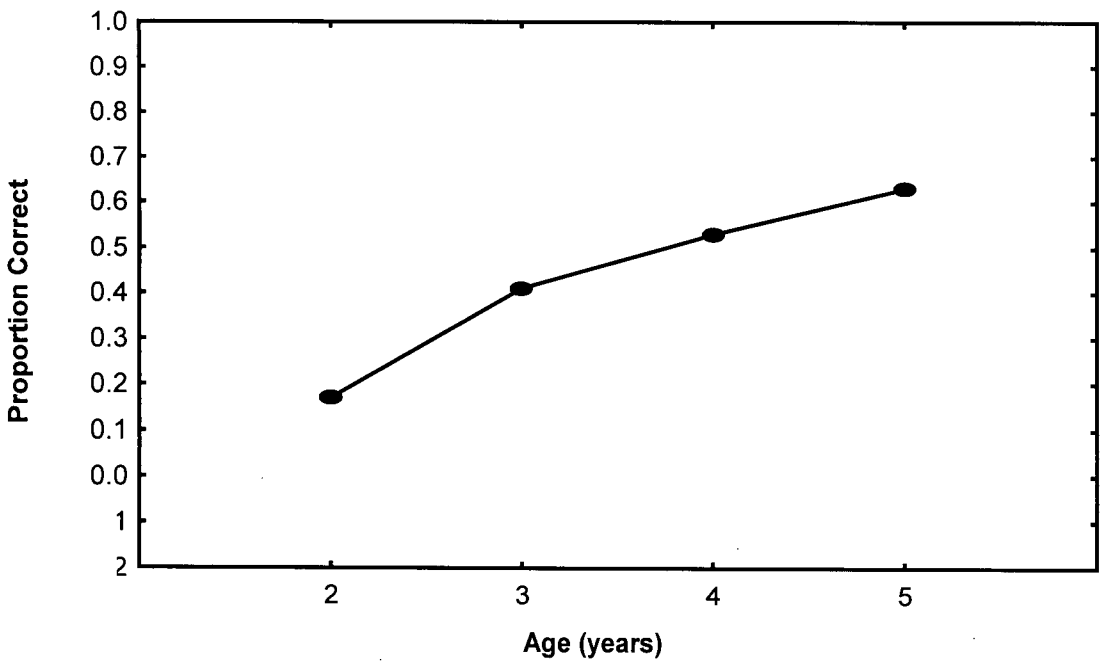


Figure 4. The proportion of correct responses to the facial expressions made by the children in each age group.

As in previous research, the children's proportion of correct responses in the facial expressions labeling trial was poor, and increased only gradually with age (Figure 4). In addition, their level of responding to the faces as well as their proportion of correct responses was low compared to their performance on the animal labeling trial. For example, the proportion of 2s' responses to the faces that were interpretable in terms of specific emotion category (whether correct or incorrect) was .26, whereas their level of responding to the animal pictures was .95.

Two parallel repeated measures ANOVAs were calculated (alpha = .05) on children's free-labeling responses to the faces.[5] Age (four levels) and sex (two levels) were between-subjects factors, and face (happy, sad, angry, surprised, disgusted, and scared in the analysis of specific emotion category; happy, sad, angry, disgusted, and

---

[5] The analyses reported in this article were based on emotion categories in which descriptive terms for the faces (e.g., *smile* for the happy face, *cry* for the sad face, *frowned* for the angry face, etc.) were scored as incorrect. These descriptive responses were re-coded as correct and the analyses repeated. Overall, there was very little effect. The 2s produced ten of the twelve responses that were re-coded. Thus, their proportion of correct responses increased slightly. In the ANOVA calculated on the specific emotion category ratings of the children's responses, the sex x face interaction was not significant ($p = .09$), but all the other significant effects remained the same. There were no differences in the ANOVA calculated on the valence ratings of the children's responses.

scared in the analysis of valence – surprise was excluded because, according to Bullock and Russell's model it is a state of high arousal with neutral valence) was the within-subject factor. The dependent variable was whether the label was correct or not, scored 1 or 0.

Specific Emotion Category

The labels were scored as correct if they matched the target specific emotion category. There were significant main effects for age, $F(3, 152) = 50.55$, $p < .001$, and face, $F(5, 760) = 131.66$, $p < .001$. The age x face interaction was also significant, $F(15, 760) = 7.75$, $p < .001$, as was the sex x face interaction, $F(5, 760) = 2.30$, $p = .04$. There were no other significant effects with alpha set at .05.

The significant main effect for age (Table 4, column means) indicated that as age increased a larger proportion of the children correctly labeled the six facial expressions. Each older age group's proportion correct was significantly greater ($p < .01$) than the younger groups'. The most dramatic increase in proportion correct was between the 2s and 3s.

The significant main effect for face, indicated that children labeled some facial expressions more accurately than others (Table 4, row means). A majority of the children labeled the happy, angry, and sad faces correctly, but not the disgust, surprise, and scared faces.

The significant age x face interaction (Table 4, cell means) was due to different rates of increase in proportion correct with age for each face. This increase with age was significant for every face. For the happy face, the improvement with age was not significantly different for the 3s, 4s, and 5s, but the 2s' proportion correct for this face

was significantly lower than each of the other age groups'.

For the sex x face interaction, the only significant difference between sexes was for the scared face ($p = .001$): the boys' proportion correct (.30) was greater than the girls' (.13). This difference was to due to the girls higher proportion of uninterpretable- or non-responses for the scared face: .41 of their responses were uninterpretable in terms of emotion, while only .30 of the boys' responses were uninterpretable. Both the boys (.23) and the girls (.19) offered *happy* for the scared face. However, the modal response for the boys was *scared*, while the modal response for the girls was *happy*. There was no general trend for one sex to demonstrate an advantage over the other for the other faces.

Overall, the analysis of specific emotion category ratings indicated that the children improved gradually with age in their facial expression labeling accuracy. The children's accuracy in labeling the faces varied by expression, with the happy face being correctly labeled most frequently, followed by the angry, sad, surprised, scared, and disgust faces (in that order). This order was the same for each of the three older age groups. The task of free-labeling facial expressions was very difficult for the 2s, whose proportion of correct responses never exceeded .40, and was less than .05 for the surprised, scared, and disgusted faces.

Valence

In this next analysis, the labels were scored as correct if they matched the valence (positive, negative) of the target emotion. There were significant main effects for age, $F(3, 152) = 66.00$, $p < .001$, and face, $F(5, 760) = 52.77$, $p < .001$. The sex x face interaction was also significant, $F(4, 608) = 3.32$, $p = .01$. There were no other significant effects with alpha set at .05.

Table 4

Effect of Age and Stimulus on Children's Production of Labels for Faces

Face	Age (years) 2	3	4	5	Mean
	Proportion of Labels with the Correct Specific emotion category				
Happy	.38 $_{a}$	.90 $_{cf}$	.95 $_{c}$	.98 $_{c}$	.80
Sad	.28 $_{a}$	.50 $_{d}$	.78 $_{f}$	.65 $_{ef}$	.55
Angry	.33 $_{a}$	.83 $_{f}$	.88 $_{cf}$	.98 $_{c}$	.75
Surprised	.00 $_{b}$	.10 $_{b}$	.25 $_{a}$	.65 $_{e}$	.25
Disgusted	.03 $_{b}$	.03 $_{b}$	.03 $_{b}$	.10 $_{a}$	.04
Scared	.03 $_{b}$	.10 $_{b}$	.30 $_{a}$	.45 $_{d}$	.22
Mean	.17	.41	.53	.63	

Note. Alpha = .05 for all LSD comparisons. Means in the same rows that do not share a subscript differ at $p < .05$. Means in the same column that do not share a subscript differ at $p < .02$. LSD comparisons on the main effect for emotion in the analysis of category ratings (row means) indicated that the proportions correct for happy, and angry did not differ significantly from each other and were each significantly greater ($p < .001$) than the proportion correct for sad, surprised, disgusted, and scared. The proportion correct for sad was significantly higher ($p < .001$) than surprised, disgusted, and scared. The proportion correct for surprised and scared did not differ significantly, and both were significantly higher ($p < .001$) than disgusted. LSD comparisons on the main effect for age in the analysis of the valence ratings (column means) indicated that the proportion correct for each age group was significantly different from the other age groups ($p < .02$).

The significant main effect for age indicated that as age increased the children produced labels of the correct valence for the six facial expressions more frequently. Each older age group's proportion correct was greater than the preceding groups', though the 3s (.75) and 4s (.81) did not differ significantly, nor did the 4s and 5s (.88). The most dramatic increase in proportion correct was between the 2s (.25) and 3s (.75).

The significant main effect for face, indicated that children produced more labels of the correct valence for some facial expressions than for others. Most of the children produced labels of the correct valence for the happy (.83), sad (.83) and angry (.81) faces. They had some difficulty producing labels of the correct valence for the disgust (.51) and scared (.43) faces, for which the proportions correct were significantly lower than each other, and each significantly different from the other three faces. The types of responses that the children made for each face are explored below in Types of Responses.

For the sex x face interaction, the only significant difference between sexes was for the scared face ($p = .003$); the proportion correct for boys (.50) was greater than that for girls (.35). This difference was to due to the girls' higher proportion of uninterpretable- or non-responses for the scared face: .44 of their responses to the scared face were uninterpretable in terms of valence, whereas only .23 of the boys' responses were uninterpretable. There was no general trend for one sex to demonstrate an advantage over the other for the other faces.

Overall, the analysis of valence ratings indicated that the children gradually improved with age in their production of labels of the correct valence for facial expressions. Children's proportion of correctly valenced responses exceeded their proportion of correct specific emotion category responses, indicating that they frequently

understood the valence of facial expressions even when they were unable to produce the correct specific emotion category label for that face. Across age groups, the children produced the largest proportion of correctly valenced responses for the happy and sad faces, followed by the angry, disgust, and scared faces (in that order). This order was generally the same for each age group.

Types of Responses (Facial Expressions)

The children's responses to the faces were re-coded into four mutually exclusive categories: correct specific emotion category; correct valence but incorrect specific emotion category; incorrect valence; and uninterpretable- or non-response. The analysis of the types of responses enabled the investigation of all the responses the children made to the facial expressions, rather than of only their correct specific or broad category responses as was the case for the ANOVAs. (Note: The responses to the surprise face were not considered in this analysis, as surprise has no valence.) Based on the re-coded data, the difference between the specific emotion category ratings and the valence ratings is evident. In addition, the proportion of incorrect-valence responses and of uninterpretable- or non-responses can be considered separately. See Appendix B, Table B2 for the proportion of responses to each face that fit into each category for each age group.

The 2s' proportion of interpretable responses, whether correct or incorrect, was very low, even for the happy, sad, and angry faces for which the other three age groups did very well (Figure 5). Of the interpretable responses the 2s did offer for the happy, sad, angry, and disgust faces, the majority were of the correct specific emotion category

or valence. For the scared face, the 2s' proportion of incorrect valence responses was greater than their proportion of correct specific emotion category and correct-valence responses combined (i.e., more of their responses were rated as having positive valence than negative valence). This result is explored below.

The 3s' production of correct specific emotion category or correct-valence responses was much higher than the 2s' for the happy, angry, sad, and disgust faces, and on a par with the 4s' and 5s' (Figure 5). This result indicates that by 3 years of age, the children freely produced labels from correct specific emotion category or valence for some faces, though their proportion of correct responses continued to increase with age.

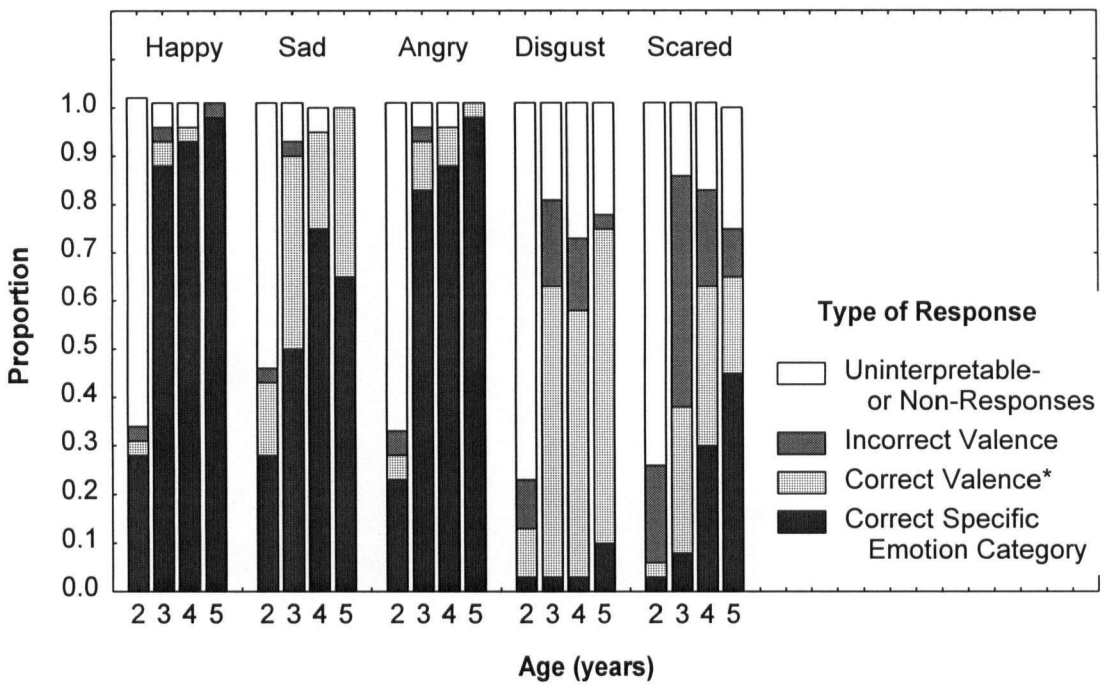


Figure 5. Proportion of responses to the facial expressions that fit into each of four mutually exclusive categories (correct specific emotion category; *correct valence but incorrect specific emotion category; incorrect valence; uninterpretable- or non-response).

Although the 3s', 4s', and 5s' proportions of correct specific emotion category or valence responses for the happy, angry, sad, and disgust faces varied from very high to very low, for each of these faces the proportion of incorrect-valence responses was relatively low, ranging from .00 to .18 (Figure 5). Thus, when the children did not know the correct label for one of these faces, they tended to produce either a label of the correct valence or offered no label at all.

For the scared face, the trend was quite different, especially for the 2s and 3s. The two youngest age groups each produced a greater proportion of incorrect-valence responses than their total correct emotion category and correct-valence responses for this face. This trend occurred only for the scared face, suggesting that there was something about this expression that allowed these children to interpret it as having either positive or negative valence, rather than predominantly interpreting it as one or the other. The 4s and 5s also produced incorrect-valence responses for the scared face, but their proportion of correct specific emotion category and valence responses were each greater than their proportion of incorrect-valence responses, indicating that with increased age the children correctly identified this expression (or at least its valence) more frequently. An explanation of this result is explored in Multidimensional Scaling.

Summary

The results of the facial expressions trial were very different from the results of the animal labeling trial on which even the youngest children performed very well. In the facial expressions trial, the 2s produced a very low proportion of responses (correct or incorrect) overall, and even the 5s produced a low proportion of correct specific emotion

category responses for the disgust and scared faces. However, more than .60 of the responses children as young as 3 years of age produced for the happy, sad, angry, and disgust face were of the correct valence. These results indicate that, although children could freely label animal pictures by 2 years of age, their association of facial expressions with emotion labels develops more slowly.

The proportion of correct specific emotion category labels increased with age for each face. For each face, and for each age group, there was an additional proportion of correct-valence-but-incorrect-specific-emotion-category labels, which indicated that there were children who could identify the valence of the expressions even when they couldn't produce the correct specific emotion label. While there were at least a few incorrect-valence responses produced for each face, for the happy, sad, and angry faces responses of this type were minimal, indicating that the children's understanding of the valence of these faces was strong. For the disgust face, there were somewhat more incorrect-valence responses, but the proportion of correct-valence labels was greater, indicating that the majority of children understood the valence of this face. These results offer support for Bullock and Russell's (1986) finding that children's early understanding of emotion, and their perception of facial expressions, is based on their initially broad valence categories, rather than on discrete emotion categories.

For the scared face, there was a much larger proportion of incorrect-valence labels produced, especially by the 2s and 3s, than for the other faces. This result suggests that the scared face was different from the other faces with negative valence. Specifically, for the other negatively valenced faces the children produced few positively valenced labels, while for the scared face the children produced more positively valenced responses than

negatively valenced ones. This finding is explored further in Multidimensional Scaling.

Multidimensional Scaling of Children's Emotion Labels

Multidimensional scaling was employed by Bullock and Russell (1986) and Bormann-Kischkel et al. (1990) to investigate the nature of children's understanding of facial expressions. The two primary dimensions identified in these studies were pleasure/displeasure and level of arousal. Bormann-Kischkel et al. also identified a third dimension based on the position of the facial features.

The multidimensional scaling of children's labeling of the facial expressions was identical the multidimensional scaling analysis of their animal labels.

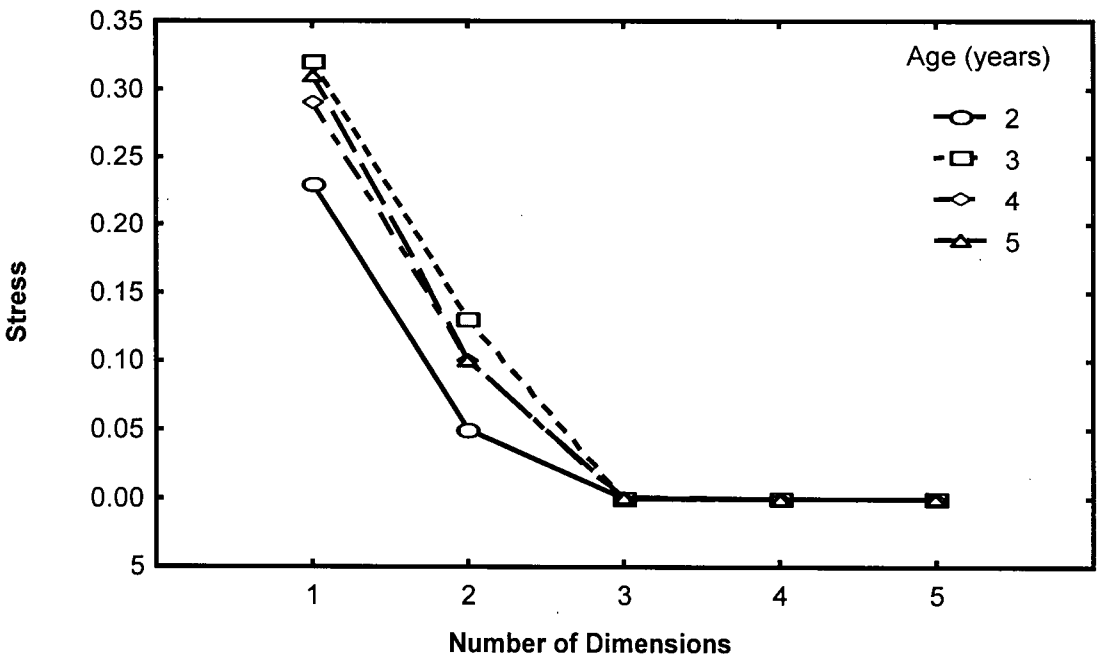


Figure 6. Stress as a function of the number of dimensions in multidimensional scaling solutions.

There was a sharp decrease in stress from a one-dimensional to a two-dimensional solution for each age group (Figure 6) indicating that there were at least two dimensions underlying the children's labeling of facial expressions. The decrease in stress from a two-dimensional solution to a three-dimensional solution for all age groups suggested the possibility of a third dimension (Figure 6).

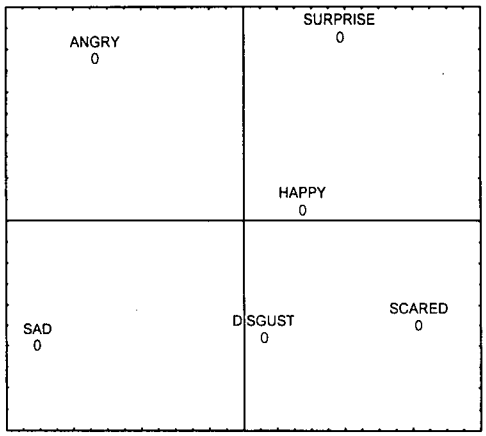
It was not possible to interpret the third dimension meaningfully and so only two-dimensional solutions were considered. In addition, the 2s' two dimensional solution (Figure 7) was uninterpretable, likely as a result of their low overall rate of responding. Because the array of the six faces in the two dimensional space varies from one age group to the next (Figure 7) interpreting the dimensions was difficult at best. For the 3s, 4s, and 5s, the two dimensions discussed below were the best candidates to explain the results of the multidimensional scaling, but other interpretations are also possible.

In the 3s', 4s', and 5s' two dimensional solutions (Figure 7), *happy* was located far from *angry*, *disgust*, and *sad*, and *happy* was also far from *scared* for the 4s and 5s. Based on this observation, I interpreted the first dimension as pleasure. This finding offers partial support for Bullock and Russell's (1986) finding that young children's understanding of emotions and facial expressions is based on the dimension of pleasure.

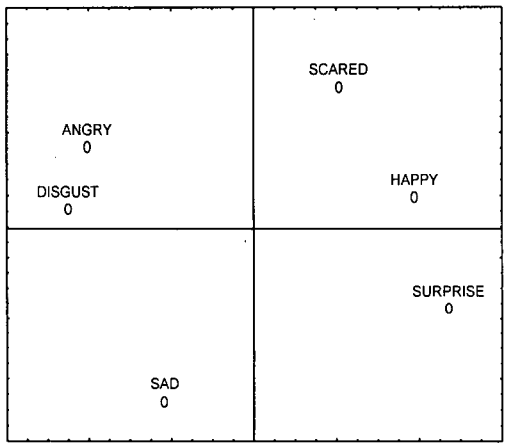
As noted in Types of Responses, some children in each age group (especially the 3s) labeled the scared expression *happy* (see Figure 5, incorrect valence). In the 3s' two-dimensional solution, *scared* and *happy* were much closer together than would be expected if the children's responses were based only on levels of pleasure in the expressions. An explanation for the mislabelings of the scared face as *happy* may be that the happy and scared faces have similar mouth positions: stretched wide with teeth

exposed (Figure 2). In addition, in the 3s', 4s', and 5s' two-dimensional solutions, the angry and disgust faces were clustered together (Figure 7). The mouth positions of these two faces were somewhat similar: the lips are pulled back and either the upper or lower teeth are exposed. Thus, the second dimension was interpreted as relating to the position of the facial features themselves, specifically, the position of the mouth and exposure of teeth. This interpretation of the second dimension offers support to Bormann-Kischkel et al.'s (1990) finding of a dimension based on the position of facial features in children's understanding of facial expressions.

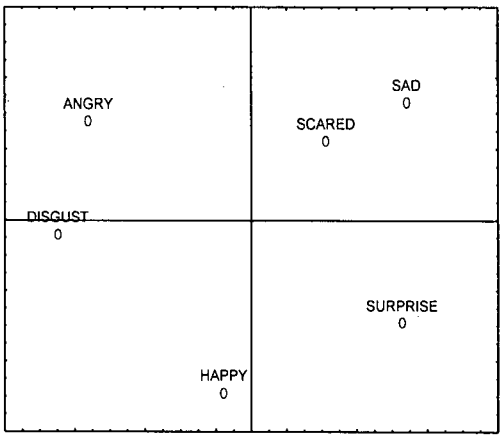
The presence of these two dimensions (pleasure, facial feature position) in the multidimensional scaling analysis confirms my analysis of the types of errors that the children made. The pleasure dimension indicates that it was more likely that the children would confuse faces of the same valence, which they did (e.g., calling the sad and disgust faces *angry*), and unlikely that they would confuse faces of opposite valence, which was also true (the happy face was virtually never called *angry*, *sad*, *scared*, or *disgusted*). The position-of-facial-features dimension indicates that it was more likely that the children would confuse faces with similar feature positions, which they did (e.g., calling the scared face *happy*).



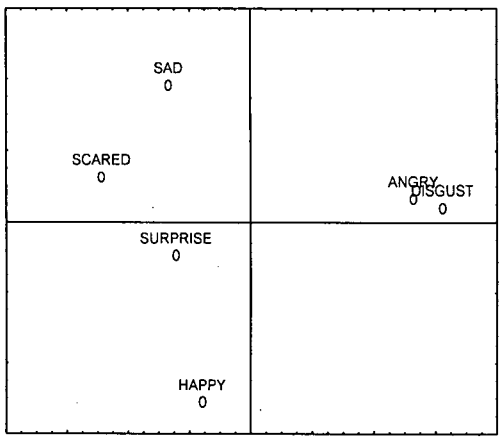
2-year-olds



3-year-olds



4-year-olds



5-year-olds

Figure 7. Two-dimensional solutions for each age group in multidimensional scaling analyses of children's responses to facial expressions.

General Discussion and Conclusions

Past researchers often found that children had difficulty labeling facial expressions and that performance improved only gradually with age. In interpreting such results, researchers discounted the poor performance as due to the incidental verbal demands of the task. For those who assumed that children had recognized and understood facial expressions in terms of discrete emotions from infancy (e.g., Izard, 1971), the children's difficulty would otherwise have been perplexing, and the difficulty of the task may have seemed to be the only possible explanation for the results.

What do the results of the current study indicate about this explanation and the development of how children label facial expressions? First, the children's overall level of performance on the facial expression labeling trial again improved gradually with age, as it did in previous facial expression free-labeling studies (Gates, 1923; Izard, 1971; Harrigan, 1984; Wiggers & van Lieshout, 1985; Markham & Adams, 1992). In the current study, the 2-year-olds' performance was poor overall. By 3 years of age, the children produced correctly valenced labels for facial expressions. The proportion of correct specific emotion category labels increased with age.

Second, the children's level of performance varied with facial expression. The 3s', 4s', and 5s' proportion of correct specific emotion category labels was very high and quickly approached ceiling level as age increased for the happy and angry faces; for the sad, scared, and surprised faces, proportion correct also increased with age; the proportion correct remained at floor level for the disgust expression. The difference between the proportion of responses with the correct valence and those of the correct specific emotion category produced for the angry, scared, and sad faces decreased as age

increased, indicating that the children's emotion categories were narrowing with age, as predicted by Bullock and Russell's (1986) model. (The 2s rate of responding to the faces was very low, overall, and there were minimal differences between the proportion of correct-valence and correct emotion-category responses.)

Third, the children's responses to the faces reflected two underlying dimensions: pleasure, and position of facial features (especially the mouth). In the multidimensional scaling analysis, the children's two dimensional solutions tended to locate *happy* opposite the emotions with negative valence; and faces that shared similar mouth positions tended to be clustered together (e.g., *angry* and *disgust* were clustered together). The finding of a dimension based on pleasure also offers further support to Bullock and Russell's (1986) findings and the finding of a dimension based on the position of the facial features offers partial support to Bormann-Kischkel et al.'s (1990) findings.

Finally, was the free-labeling task the source of children's poor performance in labeling facial expressions, as suggested by Izard (1971) and others (e.g., Markham & Adams, 1992)? As noted in the introduction, there is ample evidence that young children can free-label a variety of visual stimuli (Johnson, 1992; Johnson & Clark, 1988; Cycowicz et al., 1997). Furthermore, from the results of the animal labeling trial, it is clear that even 2-year-olds can freely produce correct labels for some stimuli. The 2s produced correct broad category labels for all six animal pictures on the majority of the trials ($\geq .85$), and their production of correct specific labels was at ceiling for the cat, dog, and horse. On the facial expressions trial, the 3s produced labels of the correct valence or correct specific emotion category at near-ceiling levels for the happy, sad, and angry faces. Evidently, children can freely produce labels for various types of stimuli,

including facial expressions. Thus, Izard's conclusion that the children's poor performance in labeling facial expressions was the result of task difficulty clearly does not explain all of the results of the current study.

Overall, the results of the current study indicate that there is a developmental course in how children label facial expressions. Given the children's excellent performance on the animal labeling trial, the results of the current study suggests that this developmental course is not merely a reflection of their increasing ability to perform the free-labeling task itself, as suggested by some (e.g., Izard, 1971). Rather, this developmental course reflects, at the very least, children's increasing association of facial expressions with discrete emotion terms.

Children's Emotion Lexicon

What do the results of the current study indicate about children's use and understanding of emotion terms? First, the proportion of children who produced emotion labels, be they the target term for the face or not, far exceeded those who could apply the target labels to the expressions. For example, 76% of all the children produced an emotion word in response to the sad face, even though only 54% were synonyms of *sad*. Shown a prototypical facial expression of disgust, 3% of the 3s responded with a synonym of *disgust*; another 62% responded with *angry*, *sad*, or *happy*.

If we require a child to know the appropriate face for a given emotion term before crediting the child with full knowledge of the meaning of the word, then many 2s and 3s do not have the full meaning of some of the emotion terms they produce. This finding implies that, although observational research has found that 2-year-olds produce a

number of emotion terms in appropriate contexts (Brown & Dunn, 1991; Dunn et al., 1987; Bretherton & Beeghly, 1982; Ridgeway et al., 1985), they do not seem to produce that term in response to an appropriate facial expression. This result suggests that children's implicit definitions of the emotion terms they produce are not initially linked to facial expressions, which is a surprising finding if children's knowledge of emotions is based on their early, possibly innate, understanding of facial expressions (e.g., Plutchik, 1994; Denham, 1998; Harris, 1989, 1983; Field & Woodson, 1982).

Clearly, further research into children's acquisition and understanding of emotion terms is required. For example, a replication of observational studies (e.g., Ridgeway et al., 1985) which reports not only the prevalence of children's use of emotion terms, but also whether some children produce many such terms while others produce very few would be interesting and indicate whether children generally acquire emotion terms at the same rate, or if there are particular children who appear to especially seek out terms for emotions. In addition, studies that investigate how children label other aspects of emotions (e.g., the causes and effects of emotions) would offer indications of what children's early implicit definitions of emotion terms are linked to.

Second, between the ages of 2 and 6 years, the children's emotion categories narrowed, as predicted by Bullock and Russell's (1986) model. Specifically, the proportion of correct specific emotion category labels increased with age, while the proportion of correct-valence-but-incorrect-specific-emotion-category labels decreased with age. One interpretation of this result is that children's association of particular facial expressions with particular labels was developing. While they labeled the happy and angry faces virtually perfectly before their sixth birthday, for the other faces (e.g., sad,

disgust, scared, and surprised facial expressions) the process of associating specific emotion terms with facial expressions was far from complete.

Children's Understanding of Faces

Bullock and Russell's (1986) model of the development of children's concepts of specific emotions and their understanding of facial expressions proposed that before children associate facial expressions with discrete emotions, they understand them in terms of valence. Thus, from Bullock and Russell's perspective, young children are more likely to mislabel faces that are similar in valence (e.g., calling the disgust face *angry*) than they are to mislabel faces of the opposite valence (e.g., calling the happy face *angry*). For the most part, this model describes the results of the current study.

However, there was a surprising result that is not predicted by any model of children's understanding of emotions and facial expressions. The children, especially the 2s and 3s, tended to produce *happy* in response to the scared face. This was the only face for which a remarkable proportion (.21) of the incorrect specific emotion category responses had the opposite valence of the target emotion. It is possible to explain this result in terms of the underlying dimensions identified in the multidimensional scaling analysis: while the pleasure dimension predicts that the children should not have confused the scared face with an expression of happiness, the facial features dimension might explain why it occurred: both the scared and happy faces have similar mouth positions.

The happy and scared faces were not grouped together in the 4s' or 5s' two dimensional solutions, while these two faces were grouped together in the 3s'solution.

This difference between the 3s' two dimensional solution and the 4s' and 5s' solution, suggests that the older children utilized the position of the mouth in their judgements of the expressions (e.g., the clustering of *disgust* and *angry*), but that these two groups gave precedence to the pleasure dimension, while the 3s gave precedence to the facial features dimension alone. Thus, it is possible that younger children's judgements of facial expressions are based on the surface features, and that with development they become more sensitive to the levels of pleasure in the different facial expressions. Perhaps a social referencing study, in which the parent was instructed to express fear with Ekman's (1976) prototypical fear expression and happiness with Ekman's happy expression, could assess whether young children's early judgements of facial expressions are based on surface features. That is, if 1-year-olds tended to proceed with some ongoing activity when each of these expressions was displayed, it would suggest that they were basing their responses on the surface features of the facial expressions.

However, there is another possible explanation for the mislabeling of the scared face with *happy*. Specifically, it is possible that before children recognize the valence of a facial expression, they will offer any emotion label. For example, the 3s produced emotion labels on 76% of their opportunities to label a facial expression, but only 52% of these responses were of the correct specific emotion category. Thus, 3-year-olds understood what was being asked of them (i.e., produce an emotion label), and had the necessary vocabulary, but when they did not know the correct specific emotion label they offered any emotion label. A similar trend occurred for the turkey picture on the animal-labeling: the 3s produced animal labels on 50% of opportunities to label the turkey, but only 33% of these labels were of the correct specific animal category, while the

remaining 67% were other bird labels.

Thus, it is possible that the developmental course of how children label facial expressions is somewhat different than the one proposed by Bullock and Russell (1986). Initially, when children do not know the "correct" emotion label for a facial expression, they may simply to offer any emotion term when asked to label facial expressions. As children's recognition of the valence of a facial expression develops, the tendency to offer any emotion term is replaced by the production of an emotion term of the correct valence, which is itself replaced by the production of the target label as they learn to associate facial expressions with discrete emotions. (The suggestion that children have an early willingness to offer any emotion label for facial expressions may apply only to children's early understanding of facial expressions, rather than to their emotion concepts in general.)

It is also possible that children have a preference for certain (perhaps more common) emotion terms, such as *happy* and *mad*, which occurs both in their initial tendency to offer any emotion term and their later offering of terms of the correct valence. This suggestion is based on the observation that when children produced labels of the incorrect specific emotion category, .44 were from the *angry* category, and .31 were from the *happy* category. If this observation has generality, it raises questions about the children's early high proportion of correct specific emotion category responses for the happy and angry faces. Specifically, did the children recognize and correctly label these two faces, or was some proportion of them offering their favorite emotion term of the correct valence? This possibility clearly requires further investigation, but if it and the suggestion that children's earliest facial expression-labeling behavior is the result of their

willingness to offer any, or some favorite, emotion term, are borne out, current assumptions about children's concepts of emotion and their understanding of facial expressions will have to be scrutinized and perhaps reconceived.

There are several avenues for further investigation. Are the results of forced choice studies the result of children's understanding of valence (i.e., being able to discriminate happy facial expressions from not-happy ones)? What pattern do children's responses in the forced choice procedure follow when all the expressions are of the same valence? What is the effect of priming on children's free labeling of facial expressions? Is the advantage offered by priming greater for free labeling than for other facial expression tasks? Such further investigation of prior findings and conclusions would illuminate our understanding of how children understand emotion, and how this understanding changes with age.

Bibliography

Anderson, P. A. & Guerrero, L. K. (1998). Principles of communication and emotion in social interaction. In P. A. Andersen & L. K. Guerrero (Eds.), Handbook of communication and emotion: Research, theory, applications, and contexts (pp. 49-96). San Diego: Academic Press.

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.

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.

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

Bretherton, I. & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. Developmental Psychology, 18, 906-921.

Brown, J. R. & Dunn, J. (1991). 'You can cry, mum': the social developmental implications of talk about internal states. British Journal of Developmental Psychology, 9, 237-256.

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 axial 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.

Camras, L. A., Grow, J., & Ribordy, S. (1983). Recognition of emotional expression by abused children. Journal of Clinical Child Psychology, 12, 325-328.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. Journal of Experimental Child Psychology, 65, 171-237.

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

Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. Developmental Psychology, 23, 132-139.

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.

Ekman, P., & Friesen, W. V. (1978). Facial action coding systems. Palo Alto, CA: Consulting Psychologists Press.

Field, T. M. & Woodson, R. (1982). Discrimination and imitation of facial expressions by neonates. Science, 218, 179-181.

Field, T. M., et al. (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.

Gates, G. S. (1923). An experimental study of the growth of social perception. Journal of Educational Psychology, 14, 449-461.

Gnepp, J. (1983). Children's social sensitivity: Inferring emotions from conflicting cues. Developmental Psychology, 19, 805-814.

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. (1983). Children's understanding of the link between situation and emotion. Journal of Experimental Child Psychology, 36, 490-509.

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

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

Johnson, C. J. (1992). Cognitive components of naming in children: Effects of referential uncertainty and stimulus realism. Journal of Experimental Child Psychology, 53, 24-44.

Johnson, C. L. & Clark, J. M. (1988). Children's picture naming difficulty and errors: Effects of age of acquisition, uncertainty, and name generality. Applied Psycholinguistics, 9, 351-365.

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.

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.

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. & de Haan, M. (1997). A neurobehavioral approach to the recognition of facial expressions in infancy. In J. A. Russell, J. M. Fernandez-Dols (Eds.), The psychology of facial expression: Studies in emotion and social interaction, 2nd series (pp. 176-204). New York: Cambridge University Press.

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

Plutchik, R. (1994). The psychology and biology of emotion. New York: Harper Collins.

Reichenbach, L. & Masters, J. C. (1983). Children's use of expressive and contextual cues in judgements of emotion. Child Development, 54, 993-1004.

Ridgeway, D., Waters, E., & Kuczaj II, S. A. (1985). Acquisition of emotion-descriptive language: Receptive and productive vocabulary norms for ages 18 months to 6 years. Developmental Psychology, 21, 901-908.

Russell, J. A. (1989). Culture, scripts, and children's understanding of emotion. In C. Saarni & P. Harris (Eds.), Children's understanding of emotion. Cambridge: Cambridge University Press.

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

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. & Widen, S. C. (1999). Words versus Faces in Evoking Children's Knowledge of the Causes of Emotions. Submitted for publication.

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.

Stifter, C. A. & Fox, N. A. (1987). Preschool children's ability to identify and label emotions. Journal of Nonverbal Behavior, 11, 43-54.

Walker-Andrews, A. S. (1988). Infants' perception of the affordances of expressive behaviors. In C. Rovee-Collier & L. P. Lipsitt (Eds.), Advances in infancy research, Vol. 5 (pp. 173-221).

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.

Appendix I

Animal Labels Produced by the Children

Table I-1

Labels Included in Each Specific Animal Category

	Cat	
cat	kitten	kitty
	Dog	
bow-wow	dog	puppy
small woofer	woof-woof	
	Horse	
donkey	horse	neigh
	Cow	
calf	cow	moo
	Turkey	
gobble-gobble	turkey	
	Goose	
duck	geese	goose
quack		

Responses that Did Not Fit into One of the Animal Categories

animal	flamingo	pig
bird	frog	seagull
a cock	hen	sheep
chick	must be a horn	swan
cock-a-doodle-doo	not a bird	white bird
chicken	ostrich	rooster
eagle	peacock	

Table I-2

Specific Category Judgements of Children's Free-Labeling Responses to Animals

		Specific category						
Animal	Age (years)	Cat	Dog	Horse	Cow	Turkey	Goose	None/ Non-Response
Cat	2	**.97**	.00	.00	.00	.00	.00	.00
	3	**1.00**	.00	.00	.00	.00	.00	.00
	4	**1.00**	.00	.00	.00	.00	.00	.00
	5	**1.00**	.00	.00	.00	.00	.00	.00
	Mean	**.99**	.00	.00	.00	.00	.00	.00
Dog	2	.00	**.97**	.00	.00	.00	.00	.00
	3	.00	**1.00**	.00	.00	.00	.00	.00
	4	.00	**1.00**	.00	.00	.00	.00	.00
	5	.00	**1.00**	.00	.00	.00	.00	.00

	Mean	.00	**.99**	.00	.00	.00	.00	.00
Horse	2	.00	.00	**.97**	.00	.00	.00	.03
	3	.00	.00	**.97**	.00	.00	.00	.03
	4	.00	.00	**1.00**	.00	.00	.00	.00
	5	.00	.00	**1.00**	.00	.00	.00	.00
	Mean	.00	.00	**.99**	.00	.00	.00	.02
Cow	2	.00	.00	.08	**.84**	.00	.00	.08
	3	.00	.00	.00	**1.00**	.00	.00	.00
	4	.00	.00	.00	**.95**	.00	.00	.05
	5	.00	.00	.00	**.97**	.00	.00	.03
	Mean	.00	.00	.02	**.94**	.00	.00	.04
Turkey	2	.00	.00	.00	.00	**.18**	.45	.37
	3	.00	.00	.00	.00	**.16**	.08	.76
	4	.00	.00	.00	.00	**.25**	.05	.70
	5	.00	.00	.00	.00	**.41**	.03	.74

	Mean	.00	.00	.00	.00	**.24**	.15	.61
Goose	2	.00	.00	.00	.00	.00	**.77**	.23
	3	.00	.00	.00	.00	.00	**.90**	.10
	4	.00	.00	.00	.00	.00	**.85**	.15
	5	.00	.00	.00	.00	.00	**.89**	.11
	Mean	.00	.00	.00	.00	.00	**.85**	.17

Note. N = 155. Bold cells indicate the correct specific category for each animal.

Appendix II

Emotion Labels Produced by the Children

Table II-1

Labels Included in Each Specific Emotion Category

Happy		
'Cause she's happy	Happy	Nice
Excited	Happy like a goose	Really super happy
Going to play	He's thinking about someone she likes	Better
Good		Yay

Sad		
Sad	Hurt feelings	She got an owie
Sad ... like lonely	Upset	Nervous
She looks upset	When someone hurts you	Shy

Angry		
Angry	Mad	Frustrated
She's very angry	Maddy	Grouchy
A little bit angry	Real mad	Grumpy
A little mad	Cross	

Surprised

Surprise	She's surprised	Shocked
Surprised face	Hap-um-surprised	
Surprised	'Prised!	

Disgusted

Yucky	Yucky face	Gross

Scared

Scared	Frightened	Nervous

Responses that Did Not Fit into One of the Emotion Categories

A Batman is coming	Calling	Feeling bad
A bit smile	Can't say it	Feels that her…
A harebrain	Cleaning her teeth	Fine
Ahhh	Crazy	Frowned
Awed	Cry	Funny
Bad	Crying	Goofy
Beautiful!	Don't know	Grumpy and sad
Blowing her nose	Embarrassed	Ha- smile- because she's smiling
Boogie face	Exhausted	
Bored	Eyes black	Hard to think
Bunny-her nose	Feeling	Her head

Her teeth

Hi

His teeth are going up.

How's she feeling?

Hurt

I can't remember

I can't remember this one

I don't know

I don't know that one

I don't know these pictures

I don't want to

I know (but wouldn't say)

I think she's crying

Itchy?

Kinda sick

Laughing

Let's look at next one.

Like a monster. Kinda like a monster.

Like a piggy

Loud

Mask

Maybe she didn't want her mom

Mommy!

Monster

Mouth

Naughty

No

No feeling right now

Not feeling

Not good

Not happy

Not sick

Oh

Oh that!

Oooh

Oopsy daisy

Open her mouth

Opening his teeth

Pig

Proud

Puzzled

Saying "please"

She doesn't care

She doesn't want her mom.

She feels...she's making funny faces. Funny faces are feelings, too

She's putting her nose like that

She's saying, "cheese"

She's saying "huh!"

Shivering

Showing teeth

Sick

Silly

Sleepy

Smelling something

Smile

Smiley

Smiling

Sniffle

Squealy

Squichy

Stinky

Stressed

Talking

Teasing

Teeth

Teeth are together

Tired

Uh, oh!

Umm, kinda spooky

What is she doing?

What's she smelling?

When something bad happen I just tell the teacher

Whew

Why is she doing that?

Why is she smiling?

Witch is coming

Witch is coming with a broom… and maybe take her away

Table II-2

Specific Emotion Category Judgements of Children's Free-Labeling Responses to Facial Expressions

		Specific emotion category						
Face	Age (years)	Happy	Angry	Sad	Scared	Surprised	Disgusted	Non-Response/ Uninterpretable
Happy	2	**.38**	.00	.00	.00	.00	.03	.60
	3	**.90**	.00	.00	.00	.00	.00	.10
	4	**.95**	.00	.00	.00	.00	.00	.05
	5	**.98**	.00	.00	.00	.00	.00	.03
	Mean	**.80**	.00	.00	.00	.00	.01	.19
Angry	2	.00	**.33**	.00	.00	.00	.03	.65
	3	.00	**.83**	.05	.00	.00	.00	.13
	4	.00	**.88**	.00	.00	.00	.00	.13
	5	.00	**.98**	.00	.03	.00	.00	.00
	Mean	.00	**.75**	.01	.01	.00	.01	.23

Sad	2	.03	.10	**.28**	.00	.00	.03	.58
	3	.00	.33	**.50**	.00	.00	.00	.18
	4	.00	.13	**.78**	.00	.00	.00	.10
	5	.00	.20	**.65**	.05	.00	.00	.10
	Mean	.01	.19	**.55**	.01	.00	.01	.24
Scared	2	.20	.00	.00	**.03**	.03	.03	.73
	3	.43	.18	.08	**.10**	.00	.00	.23
	4	.15	.05	.20	**.30**	.05	.00	.25
	5	.10	.05	.13	**.45**	.13	.00	.15
	Mean	.22	.07	.10	**.22**	.05	.01	.34
Surprised	2	.10	.00	.00	.03	**.00**	.03	.85
	3	.35	.03	.05	.13	**.10**	.00	.35
	4	.13	.03	.05	.15	**.25**	.00	.40
	5	.15	.05	.03	.03	**.65**	.00	.10
	Mean	.18	.03	.03	.08	**.25**	.01	.43

Disgusted	2	.05	.03	.03	.00	.00	**.03**	.88
	3	.10	.48	.08	.00	.00	**.03**	.33
	4	.08	.40	.03	.00	.00	**.03**	.48
	5	.03	.53	.05	.00	.00	**.10**	.30
	Mean	.06	.36	.04	.00	.00	**.04**	.49

Note. n = 40, N = 160. Bold cells indicate the correct specific emotion category for each emotion.