INTERNATIONAL JOURNAL OF BEHAVIORAL DEVELOPMENT, 1997, *20* (4), 627–649

Facial and Verbal Measures of Children's Emotions and Empathy

Janet Strayer
Simon Fraser University, Burnaby, Canada

William Roberts
University College of the Cariboo, Canada

The relation between children's emotional expressions and their reported experience of emotions is relevant to multicomponent models of both emotion and empathy development. Children ($N = 73$) in three age groups (5-, 9-, and 13-year-olds) participated in a study of their facial and verbal responses to emotionally evocative videotapes. Children were unobtrusively videotaped while they watched these stimuli, and their facial expressions were coded. Children were also interviewed to determine the emotions they attributed to stimulus persons and to themselves. There was significant but modest convergence between facial and verbal measures of emotion, an important finding given the paucity of independent research in this area. In addition, two nonconvergent indices of empathy derived from facial and verbal measures showed some functional similarities in their relations with other variables. On the basis of similar results from other laboratories, we argue that efforts to find converging measures of emotion and empathy should be redirected to study the semi-independent development of facial and verbal emotional responses to the same events.

The purposes of the present study are to assess the convergence of children's verbal and facial expressive responses to emotions they witness in others; to examine age and gender differences in these responses; and to examine each as an index of empathy, or concordant emotional response to others' emotions. Emotional experience, as assessed by facial and physiological measures, and awareness of one's emotional experience, have been said to

Funding for this research was provided by the Social Sciences and Humanities Research Council of Canada. We would like to express our thanks to the parents and children who participated in this project, and to Marianne Schroeder for her valuable assistance in data collection and anlaysis, and her help on an earlier draft of this paper. Parts of this paper were presented at meetings of the Society for Research in Child Development, 1993.

Requests for reprints should be sent to Dr. Janet Strayer, Department of Psychology, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada.

comprise two major aspects of emotional development (Lewis & Michalson, 1983). Although physiological indices seem promising as measures of emotional experience (Eisenberg et al., 1988a), facial expressions provide more specific information regarding valence and type of emotion than do more general arousal measures (Izard, 1982; Lewis & Michalson, 1983).

Facial expressions can be reliably measured in infants, but assessing awareness of emotional experience requires a verbal subject. For children able to provide both measures, a major and persistent question has been the extent to which verbally reported emotions (i.e. subjective experiences) correspond to other measures of emotional experience. A related issue is whether facial and verbal measures of emotion show similar trends across age and gender. Parallel concerns hold for facial and verbal indices of empathy.

Coherence and Convergence

We know of no theory of emotion that posits the independence of facially expressed and self-appraised emotions. Rather, these are considered to be different aspects or components of an unified, if complex, underlying process (Bowlby, 1982; Laird, 1984; Lewis & Michalson, 1983).

To the extent that facial and verbal measures of emotion are assessing the same process, it follows that they should provide coherent information, that is, they should converge, and if they fail to do so, then their divergent information should be related in lawful, orderly ways. Among such orderly transformations we would include the influence of context (e.g. Schacter, 1964), the consequences of socialisation experiences (e.g. Brody, 1985; Lewis & Michalson, 1983), and coping processes such as denial. Although evidence for several types of coherence was found in the present study, our focus, and the most important for methodological purposes, is convergence.

The convergence of facial and verbal measures is important for our understanding of the construct of emotion and for determining the validity of generalising across methods when investigating emotions and empathy. It would be particularly useful for developmental research on empathy if nonverbal indicators (e.g. facial expressions) could reliably be used to index concordant emotion in ways similar to older persons' verbal reports. There are possible confounding effects in verbal measures due to children's limited introspective and verbal skills, the demand characteristics of verbal interview situations, and dissimulation in verbal reports. Similarly, social constraints for facial expressions also operate for children (Saarni, 1989). Such factors make convergence desirable, but also militate against it.

Convergence is also important because aggregating across methods is essential for construct validity (Cook & Campbell, 1979). A measure of emotion that is tied to a single method is not satisfactory, especially if it is at

variance with another method that theoretically should produce similar results because it is assessing the same process.

In addition, lack of convergence gives rise to an important issue: The discrepancy between public signals of self-state, and self-perception of self-state. It is plausible to suppose that greater congruence between public signals and self-perceptions facilitates the co-ordination and maintenance of co-operative interactions, and that accuracy in self-perception may be related to prosocial behaviour and empathy (Roberts & Strayer, 1996; Strayer, 1989).

Given a fundamental unity of process, then, we expect a positive relationship between children's facial expressions and verbal reports of their own emotions. Findings that identification of one's own emotional experience increases with age (Harris & Lipian, 1989) imply that convergence should also increase with age.

Despite these theoretical considerations, empirical evidence for convergence is not strong. Although findings exist for children's verbally reported and (to a lesser extent) facially expressed emotions assessed singly, there is little information on their joint relations (Casey, 1993; Eisenberg et al., 1988a, 1989). These findings indicate modest convergence between facial and verbal report measures of emotion in response to viewing filmed emotional stimuli. Compared to this work, the present study uses a larger stimulus set (six rather than two stimulus vignettes) presenting a range of different emotions that afford a greater measurement range for both facial expressions and verbal reports. In addition, convergence is examined in ways which extend and clarify previous research.

Emotions: Trends Across Age and Gender

Age. Children's verbally reported emotions are expected to change with age due to increasing cognitive ability to interpret emotions (Harter, 1986). For example, the diversity of emotions experienced in response to emotional stimuli typically increases with age (Feshbach & Roe, 1968; Lennon & Eisenberg, 1987; Strayer, 1993). In addition, consistent with the prevalently dysphoric content of the stimuli used in the present study, we expect a decrease with age in verbal reports of positive emotions (Mood, Johnson, & Shantz, 1978) and an increase in reports of negative or dysphoric emotions, as children come to understand more fully the negative situational and emotional contexts we presented to them, and as cognitive and other changes increase children's ability to tolerate in themselves vicarious responses to others' distress.

We favour this hypothesis over the alternative that facial emotions should decrease with children's age due to socialisation factors that support increasing regulation and control of emotional displays (Cole, 1986; Saarni,

1989). Although regulation is important in some contexts, the present procedure entails no social interaction during the time when unobtrusive facial recordings are made, thus minimising the social meaning of display.

Gender. Traditional gender role socialisation should result in more frequent reports of emotions by girls than by boys (i.e. fewer reports of "neutral" or "nothing much"), of more reported anger by boys, and more reported fear and sadness by girls (Brody, 1985; Edelbrock & Achenbach, 1980; Maccoby, 1980). However, facial expressions should be less affected by gender role socialisation, to the extent that they are less affected than verbal responses by cognitive processes.

Empathy

Although researchers have recognised the need for multiple measures of emotion in the study of empathy (Hoffman, 1982), few studies have used both verbal and facial measures (Eisenberg & Lennon, 1983). Those that have, report equivocal data for adults (Zuckerman, Dorman, Larrence, & Speigel, 1981), and results are even more limited for children. For example, Eisenberg, McCreath, and Ahn (1988b) found no relation between children's facial expressions and verbal reports when these were measured sequentially. Therefore, in the present study, facial responses were recorded while children viewed the stimuli and our interview techniques (in contrast to Eisenberg et al., 1988b) used memory cues to assist children in recalling their emotional responses to the stimuli. In order not to interrupt these responses, interviews about experienced emotion were conducted after all stimuli were viewed.

Empathy, defined as a shared emotion concordant with that of another person (Feshbach & Roe, 1968), is assessed in two ways in the current study: (1) as verbal affect match, or agreement between emotions that the child attributes to self and to the stimulus character; and (2) as facial affect match, or agreement between the dominant facial expressions of the child and the stimulus character. Thus, our focus is primarily on affective, rather than cognitive, aspects of empathy (cf. Strayer, 1993, for a measurement approach that integrates cognitive and affective aspects of empathy).

Because knowledge of and responsiveness to others' psychological states increases with age (Selman, 1980), we expect frequency of facial and verbal empathy to increase with age. Given previous findings of age-related increases in verbal measures of children's empathy (Flapan, 1968; Selman, 1980; Strayer, 1993), similar findings for a facial measure would lend support to its interpetation as a nonverbal index of empathy.

The convergence of facial and verbal empathy is expected because both facial and verbal measures of emotional experience should function

similarly to index empathy. In contrast, findings of nonconvergence between facial and verbal empathy measures might indicate regulation or dissimulation by the child in either the facial or verbal channels, or children's limitations in accessing and identifying their emotions via verbal report. In either case, lack of convergence would have important methodological and theoretical implications.

METHOD

Subjects

A total of 73 children in three age groups participated in this phase of a larger study on socialisation and empathy. Group 1 consisted of 15 boys and 18 girls ($M = 5.13$ years, SD = 0.34). Groups 2 and 3 each consisted of 10 boys and 10 girls. For Group 2, $M = 8.83$ years, SD =0.38; for Group 3, $M = 13.07$ years, SD = 0.42. Children came from predominantly white, middle class backgrounds. The mean age for mothers was 37 years (SD = 5.3), for fathers, 39 years (SD = 6.2).

Measures and Procedures

Children individually viewed six emotionally evocative videotaped vignettes. The vignettes (a brief description is provided in Table 1) portray primarily dysphoric affect, as assessed by adult and child judges (Strayer, 1989). Positive emotions occur briefly in most vignettes, and are prevalent in *Circus*. Details of selection and pretesting are given in Strayer (1993).

Attributions of Emotion. Children were individually interviewed after first watching all vignettes. Each story was cued by a picture, and children described the vignette's content in their own words, as a check on memory and comprehension. They were then asked to report each character's emotion and its intensity, and whether they themselves had felt neutral ("OK", "just regular") or an emotion and its intensity in response to the vignette. The memory check and interview were carried out for each vignette in turn. Eight emotion categories were used across all ages: neutral, happy (including positive surprise), sad, angry, afraid, startle (including negative surprise), concerned-worried, and disgusted.[1]

The stimulus materials appear to be effective elicitors of emotion for most children. Eighty per cent of the sample reported emotions for five or more of

[1]If children's spontaneous attributions of emotion were unclear (e.g. "feels bad"), they were asked which of the following terms best applied: happy, sad, angry, afraid, surprised, disgusted (yucky). We queried "surprised", which was described as most like "afraid" in all vignettes except *Circus*, for which it was described as most like "happy".

TABLE 1
Description of Videotape Stimulus Vignettes

1. *Old House*: Thee children sneak into a yard at night. A boy climbs up creaking stairs to peer through a window into the house. A looming shadow of a man appears above him, and the children run away. (*Source:* commercial film.)
   PROTAGONIST: male
   PREDOMINANT AFFECT: fear
2. *Spilled Milk*: A husband and wife have an angry exchange while their daughter watches TV. The man leaves and the woman shouts at the girl to come to dinner; the girl accidentally knocks over a glass of milk and the mother slaps her. (*Source:* "12½ Cents", National Film Board of Canada.)
   PROTAGONIST: female
   PREDOMINANT AFFECT: sadness
3. *Jeannie*: A young woman talks directly to the viewer about the difficult life she and her children had with her abusive husband. (*Source:* "Loved, Honoured, and Bruised", National Film Board of Canada.)
   PROTAGONIST: female
   PREDOMINANT AFFECT: sadness, anger
4. *Skates*: A girl and boy argue over taking turns on her new skates. The boy calls her names and threatens to tattle. She pushes him down; he runs crying to her parents. The boy lies; the father believes his story and gives the girl's skates to the boy as her punishment. (*Source:* "Our Vines Have Tender Grapes", obtained from Dorothy Flapan, who used them in a 1968 study.)
   PROTAGONIST: female, male
   PREDOMINANT AFFECT: sadness, anger
5. *Canes*: A girl talks pleasantly about her life and the fun she has despite her physical disability. (*Source:* "I'll Find A Way", National Film Board of Canada.)
   PROTAGONIST: female
   PREDOMINANT AFFECT: sadness
6. *Circus*: A father and daughter go to see the circus train. The elephant is let out to perform some tricks. The girl jumps and laughs excitedly, and is lifted up on the elephant's trunk. (*Source:* "Our Vines Have Tender Grapes".)
   PROTAGONIST: female
   PREDOMINANT AFFECT: happy

Note: Films are in black and white. Total viewing time is approximately 30 minutes. Further information is given in Strayer (1993).

the six episodes. In contrast, nearly half (45%) of all "neutral" responses were given by fewer than 10% of the sample.

Facial expressions. Using a ceiling-mounted camera, children were unobtrusively videotaped while they were watched the vignettes. Their facial expressions were scored by coders trained in Izard's Facial Coding System (Izard & Dougherty, 1982). A three-minute baseline tape for each subject was viewed initially to familiarise coders with any idiosyncratic facial

characteristics. Coders judged the child's predominant facial expression during each vignette. The same eight categories were used as for children's reports (neutral, happy, etc.).[2] Details of coding criteria are given in Table 2. Per cent agreement exceeded 80% for all categories of facially coded emotion.

Consistent with verbal reports of emotion, coders found that most children were expressive: 60% of the sample had a facial emotion coded for five or more of the six vignettes. In contrast, more than a third (37%) of all "neutral" codes were ascribed to the least-expressive 10% of the sample.

TABLE 2
Description of Facial Emotion Codes

Emotion	*Description*
Happy	Forehead smooth; cheeks raised; corners of mouth back and up.
Sad	Inner brows drawn together; vertical furrows or bulge between brows; raised inner corners of upper eyelid; eyes squinted; downward mouth corners; lower lip pushed upward by chin muscle (chin puckers).
Anger	Brows sharply down and together; vertical furrows or bulge between brows; nasal root broadened or bulged; eyes narrowed by lowering of brow; rectangular mouth; lips may be pressed together tightly.
Fear	Straight brows slightly raised and drawn together; horizontal lines or bulge on forehead; nasal root narrowed; eyes narrowed or squinted; eyelids raised (white shows more than usual); mouth corners tight or retracted.
Surprise	Brows raised; skin below brow stretched; horizontal furrows across forehead; enlarged, roundish eyes; mouth opened (corners drawn back and up when positive; otherwise just rounded mouth).
Disgust	Brows together and downward; vertical furrow or bulge between brows; nasal root bulged (nostrils enlarged); nasal bridge furrows or bulges; eyes narrowed; cheeks raised; mouth tense (tongue may be forward); lower lip forward.
Concern	Brows drawn together and possible slightly downward; vertical furrows or bulge between brows; eyes narrowed or squinted; cheeks may be raised (but no smile); mouth relaxed; posture may lean forward.
Neutral	Expression as at baseline. Includes both "no expression" and "no clearly identified emotion".

Note: Descriptions are based on Izard, Dougherty, and Hembree (1983). For "Concern", see text and footnote 2.

[2]"Concerned" is not a facial code in Izard's system; rather, it represents a procedural compromise on our part. Judges were trained using MAX pictures (Izard & Dougherty, 1982), which portray full-blown extreme expressions of each emotion—extremes which did not occur in our sample. For example, "startle" in MAX might result from the child's loss of proximal physical support, whereas in the current context it refers to the much milder emotion of "negative surprise". Similarly, "concerned" represents our raters' judgement that "startle" and "fear" were too extreme for what they saw. It therefore reflects a combination of moderate apprehension, negative suprise, and agitated interest on the part of the children so coded. In this sense, we believe that it parallels the use of "concerned" by other researchers.

Empathy. Affective empathy (Feshbach & Roe, 1968) was assessed by two indices, one based on facial measures, the other on verbal. For each index, empathy was scored as present (= 1) during a vignette if there was an exact match in emotion shared by the vignette character and child. Thus, for both indices, scores could range from zero to six.

Verbal empathy was the number of exact matches between emotions the children attributed to themselves and to vignette characters (excluding attributions of "neutral").

Facial empathy was the number of exact matches between the predominant facially expressed emotions of children and characters. [Predominant vignette emotions, listed in Table 1, were identified in pretests by 30 children (5- to 14-year-olds) and 30 adults—see Strayer, 1993.] When there was more than one dominant emotion in the stimulus episode (Table 1, numbers 3 and 4), a match with either was counted as an empathic response. We chose this procedure (matching respondent's facial expression to stimulus person's facial expression) because it gives a more clearly nonverbal measure than the alternative procedure of matching facial expressions to the emotion attributed to the vignette character by the child.

Using these criteria, incidence of verbal empathy was 73% for *Circus*, in which positive affect is predominant, and for the five vignettes in which negative affect is predominant, ranged from 19% for *Spilled Milk* (Table 1) to 49% for *Canes* (mean = 34%), for a grand mean of 40%. Facial empathy for *Circus* was 38%, and for the five negative vignettes ranged from 8% for *Spilled Milk* to 32% for Old House (mean = 20%), for a grand mean of 23%.

These indices of facial and verbal empathy showed reasonable levels of construct validity. As reported elsewhere (Roberts & Strayer, 1996; Strayer & Roberts, 1994), they were related to other measures of empathy, and both showed expected relations with prosocial behaviours. Verbal empathy scores were related to Empathy Continuum scores (Strayer, 1993), and both verbal and facial empathy scores were related to an independent measure of self-reported empathy (Bryant, 1982). In addition, both indices of empathy were positively related to prosocial behaviours across methods and sources. For example, boys' verbal empathy was related to co-operation with a peer during a laboratory task, with parent ratings of co-operativeness, with best friend ratings of prosocial behaviour, and with teacher ratings of helpfulness. Facial empathy was related to co-operation with a peer during a laboratory task, with teacher ratings of co-operativeness, and with best friend ratings of prosocial behaviour.

RESULTS

Results address the four issues raised earlier: To what extent do facial and verbal measures of emotion converge? Are they related in similar ways to

age and gender? Do facial and verbal indices of empathy converge, and are they related in similar ways to age and gender?

Do Facial and Verbal Measure of Emotion Converge?

Verbal and Facial Emotions. As shown in Fig. 1 and Table 3, there was statistically significant, but modest, convergence over methods. Across 436 comparisons (73 children × 6 vignettes), agreement between self-reported and facially coded emotions was 17%, kappa = .06, $P < .001$. (Kappa is the appropriate statistic whenever one is interested in degree of association along the diagonal of a matrix; Brown, 1990.) Convergence above chance levels occurred for "happy" (standardised deviate = 5.09), "afraid" (3.48), and "sad" (2.36). (Standardised deviates are equivalent to z-scores and indicate table cells that diverge significantly from expected values in a chi-square analysis; Brown, 1990, p. 275.) In contrast, other emotions converged only at chance levels.

Contrary to expectations, convergence failed to improve with age. As shown in Fig. 1, there was substantial overlap in the kappa confidence intervals for the three age groups.

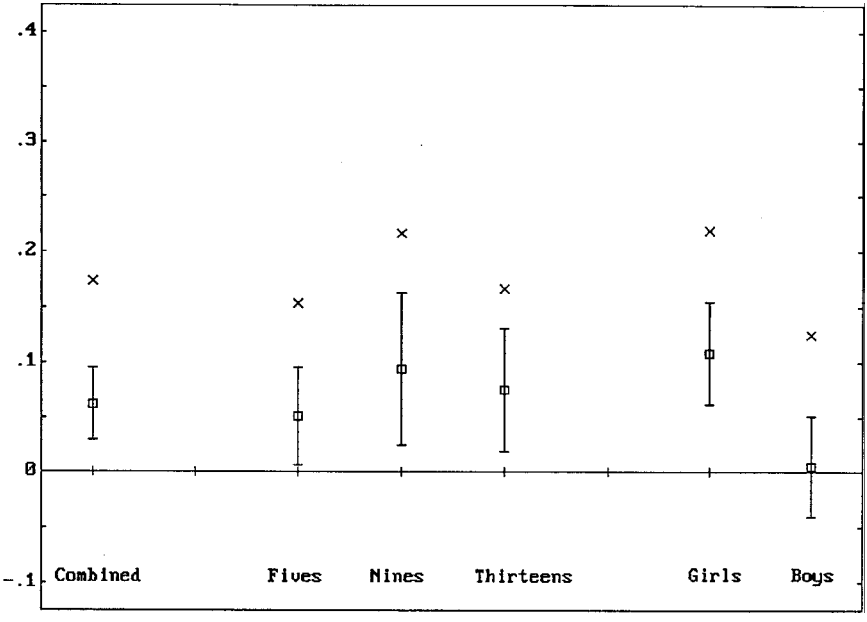

FIG. 1. Facially coded and self-reported emotions across six vignettes: Proportion agreements (×), kappas (□), and 95% confidence intervals for kappa, for the entire sample ($N = 73$) and by age and gender.

TABLE 3
Convergence Between Verbally Reported and Facially Coded Emotions

Verbal	*Facial*								
	Happy	*Sad*	*Angry*	*Afraid*	*Startle*	*Concern*	*Disgust*	*Neutral*	*Total*
Happy	24	14	5	23	1	37	7	43	154
Sad	1	15	4	6	2	26	2	15	71
Angry	0	6	1	9	0	16	0	10	42
Afraid	1	4	1	14	0	12	1	6	39
Startle	1	2	1	0	1	10	1	6	22
Concern	0	2	1	1	1	3	0	7	15
Disgust	0	4	1	4	0	3	1	2	15
Neutral	4	8	3	14	2	25	5	17	78
Total	31	55	17	71	7	132	17	106	436

Note: Values are for 73 children across 6 vignettes.

As also shown in Fig. 1, there was an unexpected gender difference in convergence. Facial expressions and verbal reports converged above chance levels for girls, but not for boys, and their respective kappas differed at $P < .05$. Specifically, girls showed better-than-expected convergence for happy (standardised deviate = 3.98), sad (2.76), and afraid (2.93); in contrast, convergence for boys was better than chance only for "happy" (3.14).

In addition to limited convergence, the two methods produced distinct patterns of overall response. As Fig. 2 and the marginal totals in Table 3 indicate, there was substantial divergence between facial and self-report measures for all emotions except sad and disgusted [for the eight by two table of emotions by method $\chi^2(7) = 209$, $P < .0001$; Cramer's $V = .49$].

Specifically, children were more likely to describe themselves as happy, angry, or startled than they were to be coded as such (standardised deviates were 10.2, 3.4, and 2.8, respectively); whereas their facial expressions were more likely to be coded as neutral, afraid, or concerned than they were so to describe themselves (standardised deviates = 2.3, 3.3, and 10.6). Thus there was both modest convergence and systematic patterns of divergence in facial expressions and self-attributions of emotion.

Are Facial and Verbal Emotions Related in Similar Ways to Gender and Age?

Gender. In contrast to facially coded emotions, which showed no significant gender differences [$\chi^2(7) = 6.63$, $P > .45$; Cramer's $V = .12$], there were marked gender differences for verbally reported emotions, all of them consistent with sex role stereotypes [$\chi^2(7) = 37.8$, $P < .0005$; Cramer's

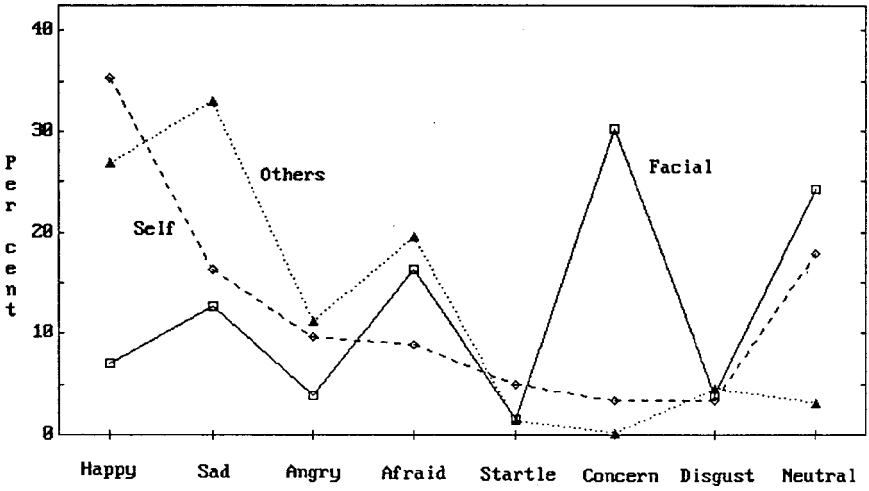

FIG. 2. Facially coded and self-reported emotions (N = 436: 73 children × 6 vignettes), and emotions attributed to vignette characters ("others"; N = 438).

V = .29]. Specifically, as shown in Fig. 3, girls were more likely than boys to describe themselves as sad or afraid (standardised deviates = 3.34 and 2.22), and boys were more likely than girls to describe themselves as feeling neutral or angry (standardised deviates = 4.45 and 2.26). Differences for other emotions were not significant.

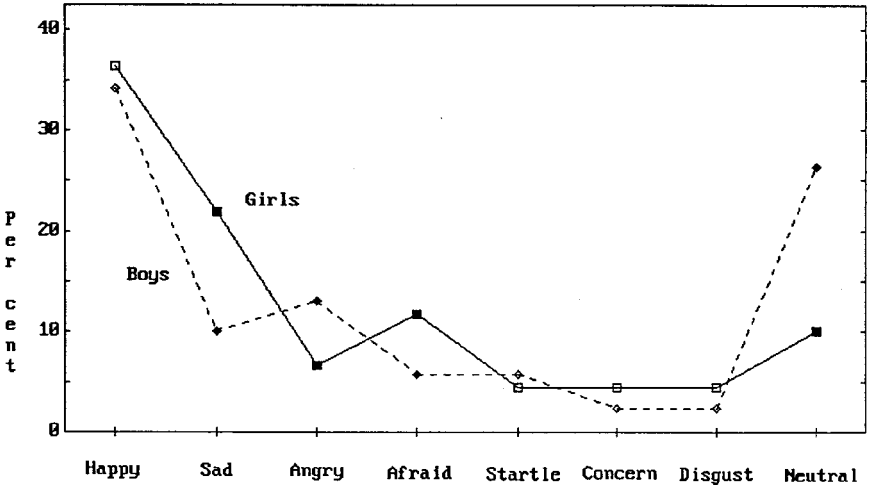

FIG. 3. Frequency of self-reported emotions across six vignettes, by gender. Solid blocks indicate significant differences (see text). Results were unchanged when emotions for the one positive vignette (*Circus*, see Table 1) were omitted.

Age. As hypothesised, children became more facially expressive with age. As shown in Fig. 4, the frequency of neutral facial codes declined with age. [For the eight by three table of emotions by age, $\chi^2(14) = 65.45$, $P < .0001$; Cramer's $V = .27$.] Specifically, standardised deviates indicated that 5-year-olds were coded as neutral more often than would be expected by chance, 13-year-olds were coded as neutral less often, with 9-year-olds intermediate between these extremes (standardised deviates were 3.89, − 1.54, and − 2.79, for 5-, 9-, and 13-year-olds, respectively). Expressions of fear also declined with age (standardised deviates were 2.89, 0.42, and − 3.64, respectively). In contrast, expressions of facial concern increased with age, as one would expect (standardised deviates were − 4.47, − 0.31, and 5.29, respectively). Age differences for facial codes of happy and sad were nonsignificant. Trends for anger, startle, and disgust are not described, given their rarity in this sample (less than 5% of all facial codes, as shown in Table 3).

Consistent with the nature of the stimuli, self-reported emotions showed a shift away from positive affect. (For the eight by three table of emotions by age, $\chi^2(14) = 66.02$, $P < .0001$; Cramer's $V = .28$.) As shown in Fig. 4, reports of "happy" declined for the two older groups (standardised deviates = 7.00, − 3.90, and − 3.90, respectively, for 5-, 9-, and 13-year-olds), and reports of feeling neutral increased (standardised deviates = − 3.03, 2.39, and 0.99). In addition, feelings of sadness showed a steady increase with age (standardised deviates were − 3.37, 0.42, and 3.33, respectively). Age differences for anger and fear were nonsignificant; results for startle, concern, and disgust are not presented due to their infrequent occurrence.

In summary, age changes in facial and self-reported emotions showed a rough correspondence. Children reported fewer positive and more negative emotions with age, as expected given our stimuli. Consistent with this, facial codes of neutral declined and expressions of concern increased. Notwithstanding these similarities, trends for particular emotions always differed by source, suggesting that facial and self-reported emotions may follow somewhat distinct developmental pathways.

Empathy

Verbal and Facial Emotions: Relations to Emotions Attributed to and Shown by Others. Emotions (including neutral) attributed to self and to others (i.e. to stimulus persons) showed moderate levels of agreement: 42% across all vignettes (436 comparisons), kappa = .29, 95% confidence interval = .24 to .33. Agreement between facial emotions expressed by self and stimulus persons was significantly weaker: 20%, kappa = .10, 95% confidence interval = .07 to .13.

These levels of agreement permit substantial differences, as shown in Fig. 2 [attributions to "other" vs. "self"; $\chi^2(7) = 114$, $P < .00001$; Cramer's

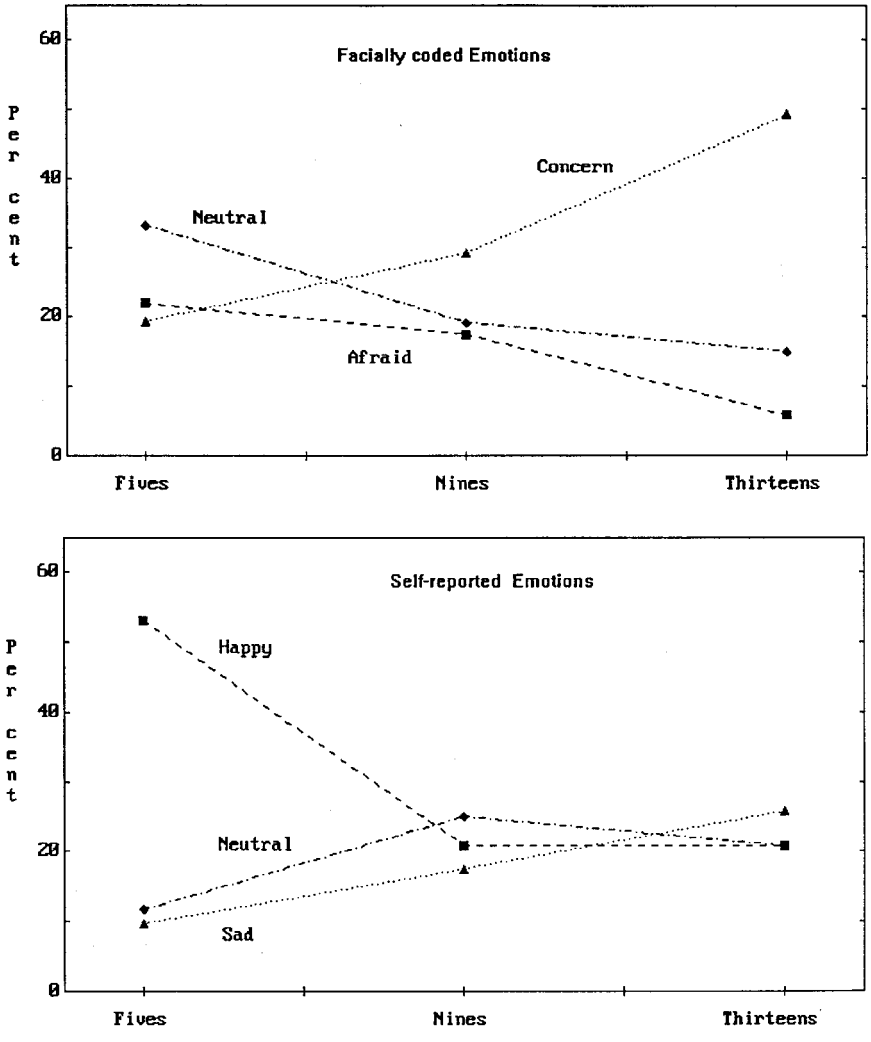

FIG. 4. Age-related changes in frequency of facially coded and self-reported emotions. Omitted emotions were either rare (see Table 3) or had nonsignificant differences across age groups. Results were unchanged when emotions for the one positive vignette (*Circus*, see Table 1) were omitted.

$V = .36$]. Children were more likely to describe themselves as feeling neutral or happy than they were so to describe others, and they were more likely to describe others as feeling sad or afraid (standardised deviates for these comparisons ranged from 2.71 to 7.05). Children were equally likely to describe themselves and vignette characters as angry. (Attributions of

startle, concern, and disgust were relatively rare, and will not be discussed.)

Similarly, facial codes also diverged strongly. Children were frequently coded as expressing concern or no emotion, whereas these were never predominant for the vignette characters. Children were also coded as being happy or sad less often than vignette characters expressed these feelings. (Standardised deviates ranged from 9.62 to 12.47. For the eight by two table of emotions by child and vignette character [$\chi^2(7) = 450$, $P < .00001$; Cramer's $V = .52$]. There were no differences in facial codes of fear. Other facial emotions (startle, disgust, anger) occurred infrequently and will not be discussed.

Thus, although there are important similarities in the emotional experiences of these children and the vignette characters they view, both on a verbal and a facial-expressive level, self-other differentiation is also salient in this data set.

Convergence of Facial and Verbal Empathy. Facial and verbal empathy scores failed to converge, $r(71) = .08$. However, verbal empathy was not completely unrelated to facial emotions. Verbal empathy occurred more often than expected when facial expressions were coded as happy (standardised deviate = 3.98; for the eight by two table of emotions by presence/absence of empathy [$\chi^2(7) = 18.44$, $P < .02$; Cramer's $V = .20$]). Although it has been suggested that facial concern should be associated with higher levels of reported empathy, this did not occur (standardised deviate = −0.07).

Verbal empathy also differed from facial empathy in that it occurred more frequently [$F(1,72) = 20.55$, $P < .0001$]. Mean verbal empathy was 2.4 (standard deviation = 1.7; range = 0–6), whereas mean facial empathy was 1.4 (standard deviation = 1.1; range = 0–4).

Gender, Age, and Empathy. Consistent with reports from other samples, girls had higher average scores than boys for both verbal and facial empathy. For verbal empathy, means were 2.8 and 2.1 for girls and boys, respectively [$F(1,67) = 4.54$, $P < .05$]. For facial empathy, means were 1.7 and 1.1, respectively [$F(1,67) = 6.57$, $P < .05$].

It is possible that these differences arise in part because boys find it more difficult to empathise with female protagonists (portrayed in five of six vignettes; boys were portrayed in two of six—see Table 1). However, several findings suggest that protagonist gender was a relatively unimportant factor in the present case. First, boys were less likely than girls to respond empathically during the vignette that featured a male protagonist. For *Old House*, verbal empathy was present in 42% of girls vs. 23% of boys; for facial empathy, values were 32% and 31%, respectively. (Few children of either sex empathised with the boy in *Skates* who lied about another child.)

Moreover, boys were as empathic to four of five female protagonists in other vignettes as they were to the male protagonist in *Old House*. (Incidence of boys' verbal empathy ranged from 23% to 69% in these vignettes.) Thus, it appears that the gender differences reported here are not simply artefacts of protagonist gender, but reflect more general factors.

As expected, age, treated as a continuous variable, was positively related to verbal empathy [$r(71) = .24$, $P < .05$]. In contrast, age was unrelated to facial empathy [$r(71) = -.05$].

DISCUSSION

Present findings indicate the necessity of distinguishing between facial and verbal measures of emotion and empathy. Although convergence was obtained between facial and verbal emotions, magnitudes were modest, and relations with age and gender sometimes diverged. The implications of present findings are discussed in terms of the issues introduced earlier.

Coherence in Facially Expressed and Verbally Reported Emotions

Although for present purposes convergence is the most important type of coherence between facially and verbally expressed emotions, it is not the only way in which these two measures may be lawfully related. As can be seen in Table 3, there is prima-facie evidence for the presence of denial or dissimulation in the present sample: A third of all verbal responses of "happy" or "neutral" were paired with facial expressions of negative emotions, chiefly fear, concern, and sadness. It is not possible, of course, to say what these children "really" felt (we are not advocating a more basic ontological status for facially expressed emotions), but this pattern of disjunction between verbal reports and facial expression clearly resembles denial as it has been classically defined. (The distinction between denial and dissimulation turns on the degree of awareness of the emotion. As we will argue later, we believe that our experimental procedures militate against deception, although of course it remains a possible response.)

Gender-role socialisation constitutes a second possible source of coherence. Gender differences in verbally reported emotions (Fig. 3) are consistent with greater socialisation pressure for boys to minimise expressions of sadness and fear, and for girls to minimise expressions of anger (Brody, 1985). Such pressures could contribute to lack of convergence in facial and verbal measures of emotion without necessarily involving conscious deception. Rather, as gender-appropriate behaviour is internalised, children come to understand or interpret their emotional experiences in gender-prescribed ways. This appears to have occurred in the current sample, when exposure to the same stimulus materials evoked very

similar emotional responses in girls and boys (as assessed by facial expressions), but resulted in gender-stereotypic differences when described to us verbally.

We can estimate the strength of this effect by taking the difference between boys' and girls' verbal responses for sadness, fear, and anger. By this criterion, about 12% of all responses appear to be influenced by gender-role socialisation pressures.

Overall, then—just as theories of emotion predict—facial and verbal expressions of emotion showed substantial coherence in the present sample: More than 60% of responses can be paired in lawful, orderly ways. What is striking in this sample is that the most straightforward type of coherence, convergence, is relatively infrequent, a finding that raises important issues of method and theory.

Convergence of Facial and Verbal Measures of Emotion

Present findings of modest, although statistically significant, convergence of facial and verbal measures of emotion are comparable to findings reported for other samples (e.g. Chisholm, 1991) and laboratories (Casey, 1993; Eisenberg et al., 1988a, 1989; Fabes, Eisenberg, & Miller, 1990). For example, Eisenberg et al. (1988a) report only 8 significant correlations ($rs >$ 0.23) out of 46 comparisons between self-report and facial measures of emotion. Although binomial tests reject the omnibus null hypothesis that all correlations are zero, $P < .01$, it is clear that the magnitude of correspondence between facial and verbal measures is quite modest.[3] The present study extends this conclusion to considerably broader stimulus and age contexts. We believe that this finding indicates an important issue for both method and theory, because it implies a possible disjunction between self-perceptions of emotional responses (as reflected in our verbal measures) and evaluations of these emotional responses by others, based on facial expressions.

In addition to modest levels of convergence, present findings failed to confirm expected age-related increases in facial-verbal concordance. In this respect, too, our findings and sample do not appear to be unusual. Casey (1993) also failed to find an age-related increase in concordance in her sample of 7- and 12-year-olds. In addition, a recent study of 60 ten-year-olds

[3]Casey (1993) reports more substantial agreement between facial and self-report measures of emotion, but this was achieved by collapsing facial expressions into only two categories, positive and negative. We believe that this procedure is imprecise because sadness, anger, and fear are distinct emotional experiences; thus aggregating in this way masks divergence in facial and verbal measures.

(Chisholm, 1991) reported 28% agreement for facial and verbal emotional responses (kappa = .15), a rate very similar to the 32% agreement (kappa = .14) for our nine-year-olds. These findings suggest that increasing insight into one's own emotions over this age range (if it occurs) may be offset by increasing pressures to conform to social or gender-role expectations. Further research is needed to clarify the size of the effect and the processes at work.

An examination of Table 3 and Fig. 4 suggests some particular ways in which facial expressions and self-attributions diverged—for example, the high number of self-attributions of happiness by five-year-olds, or the high number of facial codes of concern. We believe that such a focus would be mistaken, however. Lack of congruence between facial expressions and self-attributions was a *general* phenomenon in this sample, not the result of idiosyncrasies in a particular age group or in the coding of a particular emotion. Congruence was low for *all* age groups, not just five-year-olds. Congruence was low for *every* emotion, not just "concern". We believe that this discrepancy between facial expressions and self-attributions is important and worthy of further research attention.

Modest convergence between facial and verbal measures of emotion may occur for a number of reasons. One possibility is that children are intentionally deceiving us in their expressions or their verbal reports. Although children of this age range can dissimulate and mask facial expressions (Cole, 1986; Saarni, 1989) and can verbally respond in ways biased by social desirability (Crandall, Crandall, & Katkovsky, 1965), these possibilities seem less likely than other alternatives for the following four reasons: In the present study, facial expressions were monitored unobtrusively in a nonsocial context, thus minimising the social factors typically eliciting facial dissimulation (Yarczower & Daruns, 1982). If masking was an important factor, the number of "neutral" facial codes should have increased with age, because the ability to mask increases with age. However, the number of neutral codes actually decreased. As for possible verbal expectancy sets, present data indicate that children often reported emotions different from those of stimulus persons or reported feeling no emotion at all. Lastly, studies using similar response measures show no relationship between children's social desirability scores and their verbal (Chisholm, 1991; Cohen, 1992) or facial emotion responses (Chisholm, 1991).

A more likely explanation for low facial-verbal convergence is that children, like adults, often find it difficult to identify their feelings accurately (Schacter, 1964). One reason for this is that socialisation, from infancy onwards, often encourages the dissociation of felt and displayed emotion (Campos, Barrett, Lamb, Goldsmith, & Sternberg, 1983). For example, in home observations of 30 families rated as warm and responsive, Roberts and

Strayer (1987) found that 26% of all episodes involving crying resulted in parental demands for emotional control, and that emotional distress in the context of parental directives frequently evoked firm enforcement rather than comforting. Thus it is clear that parents often have other priorities than teaching children to label feelings accurately, and that their actions may lead to some degree of dissociation between what is experienced and what may be acknowledged verbally.

Happiness may be the emotion that is most easy to identify accurately. Given the negative qualities and contexts of dysphoric emotions, it is plausible to expect that happiness would be the emotion most often clearly identified by parents and others, and the one least often subjected to demands for suppression or control. This is consistent with our finding (and that of Casey, 1993) that the greatest convergence of facial and verbal reports occurred for happiness.

In the current sample, only happiness converged above chance levels for boys, whereas girls showed statistically significant levels of convergence for happiness, sadness, and fear. Such differences suggest that socialisation pressures for moderation and control of emotional expession operate more forcefully on boys than girls. Together with gender stereotypes, these factors may account for the greater dissociation observed in the facial expressions and verbal reports of boys in contrast to girls (Fig. 1 and Casey, 1993).

Other reasons for low convergence may be methodological. The relatively moderate emotional intensity that can be aroused appropriately in experimental settings may limit the specificity of observed facial expressions (Fabes et al., 1990). At the same time, this arousal may be sufficient for participants to label their emotion, leading to a lack of correspondence between researchers' coding of facial expressions and participants' own reports. In such a case, one would expect more facial than self-reported codes of "neutral" (as indeed occurred in our sample), and a consequent lowering of convergence.

Notwithstanding, when facial expressions are strong enough to be coded reliably, there should be good convergence with subjects' reported experiences. In the present sample, our stimuli were reasonably good at eliciting emotions: More than 80% of all trials resulted in a verbal report of emotion, and more than 75% in a facial expression clear enough to be coded. In contrast, most neutral responses came from a small subset of the subjects: 45% of all verbal "neutral" responses were given by 10% of the sample; for facial codes, the corresponding value is 37%. Thus, the prevalence of neutral codes seems to reflect individual differences more than the intensity of the stimulus materials per se. Thus, the procedure appears to be sufficient to generate high levels of concordance, if other factors did not intervene.

We do not doubt that more intense emotions would be more easily and accurately perceived by both the person experiencing them and the

observers coding them; but to the extent that there is merit in the arguments we have made concerning socialisation pressures and their consequences for identifying emotions, it follows that the low convergence reported here (and by other researchers) cannot be wholly explained by the relatively moderate level of emotional experiences that one can ethically induce for research purposes. Moreover, it is our impression that the intensities of emotional experience elicited by our procedures are comparable with the intensities aroused in everyday life in children who witness the distress of others. To the extent that this is so, our results indicate clearly a discrepancy in how such witnesses may understand their own reactions and how others may judge their reactions on the basis of their facial expressions. At one extreme, a child could experience himself as responding empathically, whereas others might judge his response as one of indifference. Our results are consistent with impressions that such misattributions of others' emotional reactions are not uncommon.

Facial and Verbal Measures of Emotion: Relations with Age

Age. Present findings indicate that for this stimulus set, positive emotions decreased with age and negative emotions increased. Changes for specific emotions may result from different processes. Increases in reported sadness, for example, accord with the prevalence of sadness in the stimulus materials. There is an increased understanding of this material with age, and this understanding, coupled with children's own accumulating emotional experiences, should result in greater experienced sadness (Eisenberg et al., 1988b; Roberts & Strayer, 1987). Facial fear may decline because older children understand that they themselves are in no danger, or that the stimulus materials are portrayals, not records of real events. The increase in facial concern reflects an increase in mild apprehension or agitated interest, as described earlier. Similar expressions in younger samples have been interpreted by other researchers as indicating children's personal distress in response to stimulus material (Eisenberg et al., 1988b). Such reactions in older children may reflect their distress to portrayed violations of their increasing sense of justice and interpersonal respect (Kohlberg, 1984).

Facial and Verbal Measures of Empathy

Convergence. Facial and verbal empathy were similar in some of their relations with other variables, suggesting that both to some extent reflect similar underlying processes. As reported in Strayer and Roberts (1994), both are correlated with self-reported empathy and with various measures of prosocial behaviour. In addition, girls scored higher than boys on both

measures, a gender difference reported for other measures of empathy (Lennon & Eisenberg, 1987).

In spite of these functional similarities, however, facial and verbal empathy scores correlated at only .08, indicating that scores on the two measures are essentially independent. Verbal empathy scores were also consistently higher than facial empathy scores, and verbal empathy was clearly related to age whereas facial empathy was not. Thus, verbal and facial empathy are distinct enough that results with one method can only cautiously be generalised to the other.

From these data, we cannot tell which measure is a "better" index of empathy. The use of facial expressive measures in studies of empathy may underestimate the extent of concordant emotion that exists. (Eisenberg et al., 1988b, also reported that facial empathy scores were significantly lower than verbal empathy scores.) On the other hand, facial empathy scores may be less subject to distortion by gender-role stereotypes. Because facial measures are particularly useful in the study of very young, nonverbal children, the possibility of systematic differences in facial and verbal data needs to be taken into account when considering measure-specific findings.

Self-Other Distinctions. The specific emotions attributed to oneself and to stimulus persons showed interesting differences. Children more frequently described themselves as happy or neutral than they so described stimulus persons, who were more frequently described as sad. Given that most vignettes were dysphoric in content, children's frequent reports of happy and neutral responses indicate a willingness to report no empathy with stimulus persons. Attributing greater sadness to stimulus persons than to oneself is also an accurate assessment of the impact of events on the participant versus the observer. These findings indicate that children of this age range have little difficulty differentiating emotional consequences for persons directly or indirectly involved in a given situation. It seems appropriate that children, who are generally accurate in attributing emotions to others (Barden, Zelco, Duncan, & Masters, 1980; Strayer, 1986), should distinguish their own from others' emotions. Indeed, such differentiation has been posited as necessary for mature empathy (Hoffman, 1975).

Conclusion

Because of the distinct information they provide, present findings support the inclusion of both facial and verbal measures in the study of emotions and empathy. As social messages, facial expressions influence others' assessment of and reactions to the self. On the other hand, the reflective assessment of

self-feeling (indexed by verbal reports) is thought to be an important component of evaluation and planning, and therefore influential in the construction of behaviour (Bowlby, 1982). Their potential disjunction may therefore have important consequences for social interactions.

Results reported here, together with similar results reported elsewhere (Chisholm, 1991; Eisenberg et al., 1988a), indicate that facial and verbal measures of emotion, although exhibiting substantial levels of coherence and converging at better-than-chance levels, converge modestly. Their distinctness and semi-independence need to be recognised, their connections more fully explored, and the contribution that each makes to the regulation and determination of social interactions (especially when they are inconsistent with one another) needs to be clarified.

Manuscript received August 1994
Revised manuscript received November 1995

REFERENCES

Barden, R., Zelco, S., Duncan, S., & Masters, J. (1980). Children's consensual knowledge about the experiential determinants of emotion. *Journal of Personality and Social Psychology*, *39*, 968–976.

Bowlby, J. (1982). *Attachment and loss: Vol. 1. Attachment.* New York: Basic Books.

Brody, L. (1985). Gender differences in emotional development: a review of theories and research. *Journal of Personality*, *53*, 102–149.

Brown, M. (1990). 4F, Part 2: Analysis of two-way tables. In W. Dixon (Chief Ed.), M. Brown, L. Engelman, & R. Jennrich (Associate Eds.), *BMDP statistical software manual* (Vol. 1, pp. 253–276). Los Angeles, CA: University of California Press.

Bryant, B.K. (1982). An index of empathy for children and adolescents. *Child Development*, *53*, 413–425.

Campos, J.J., Barrett, K.C., Lamb, M.E., Goldsmith, H.H., & Sternberg, C. (1983). Socio-emotional development. In J.J. Campos & M.H. Haith (Eds.), P.H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (pp. 783–915). New York: Wiley.

Casey, R. (1993). Children's emotional experience: Relations among expression, self-report, and understanding. *Developmental Psychology*, *29*, 119–129.

Chisholm, C. (1991). *The relation between verbal and facial measures of emotion, affect match, and empathy in 10-year-old females.* MA Thesis, Simon Fraser University, Burnaby, Canada.

Cohen, D. (1992). *Empathy in conduct-disordered youth.* MA Thesis, Simon Fraser University, Burnaby, Canada.

Cole, P. (1986). Children's spontaneous control of facial expression. *Child Development*, *57*, 1309–1321.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Crandall, V.C., Crandall, V.J., & Katkovsky, W. (1965). A children's social desirability questionnaire. *Journal of Consulting Psychology*, *29*, 27–36.

Edelbrock, C., & Achenbach, T. (1980). A typology of child behavior profile patterns: Distribution and correlates for disturbed children aged 6–16. *Journal of Abnormal Child Psychology*, *8*, 441–470.

Eisenberg, N., Fabes, R.A., Bustamente, D., Mathy, R.M., Miller, P.A., & Lindholm, E. (1988a). Differentiation of vicariously induced reactions in children. *Developmental Psychology*, *24*, 237–246.

Eisenberg, N., Fabes, R.A., Miller, P.A., Fultz, J., Shell, R., Mathy, R.M., & Reno, R.R. (1989). Relation of sympathy and personal distress to prosocial behavior: A multimethod study. *Journal of Personality and Social Psychology*, *57*, 55–66.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, *94*, 100–131.

Eisenberg, N., McCreath, H., & Ahn, R. (1988b). Vicarious emotional responsiveness and prosocial behavior: Their interrelationships in young children. *Personality and Social Psychology Bulletin*, *14*, 298–311.

Fabes, R.A., Eisenberg, N., & Miller, P.A. (1990). Maternal correlates of children's vicarious emotional responsiveness. *Developmental Psychology*, *26*, 639–648.

Feshbach, N., & Roe, K. (1968). Empathy in six- and seven-year-olds. *Child Development*, *34*, 133–145.

Flapan, D. (1968). *Children's understanding of social interaction.* New York: Teachers College Press.

Harris, P., & Lipian, M. (1989). Understanding emotions and experiencing emotions. In C. Saarni & P. Harris (Eds.), *Children's understanding of emotions* (pp. 241–258). New York: Cambridge University Press.

Harter, S. (1986). Cognitive-developmental processes in the integration of concepts about emotions and the self. *Social Cognition*, *4*, 119–151.

Hoffman, M. (1975). Developmental synthesis of affect and cognition and its implications for altruistic motivation. *Developmental Psychology*, *2*, 607–622.

Hoffman, M. (1982). The measurement of empathy. In C. Izard (Ed.), *Measuring emotions in infants and children* (pp. 279–296). New York: Cambridge University Press.

Izard, C. (1982). Measuring emotions in human development. In C. Izard (Ed.), *Measuring emotions in infants and children* (pp. 3–18). New York: Cambridge University Press.

Izard, C., & Dougherty, L. (1982). Two complementary systems for measuring facial expressions in infants and children. In C. Izard (Ed.), *Measuring emotions in infants and children* (pp. 97–126). New York: Cambridge University Press.

Izard, C., Dougherty, L., & Hembree, E. (1983). *A system for identifying affect expressions by holistic judgement (Affex).* Delaware, DE: University of Delaware Press.

Kohlberg, L. (1984). *The psychology of moral development.* San Francisco, CA: Harper & Row.

Laird, J. (1984). The real role of facial response in the experience of emotion: a reply to Torangeau and Ellsworth, and others. *Journal of Personality and Social Psychology*, *47*, 909–917.

Lennon, R., & Eisenberg, N. (1987). Gender and age differences in empathy and sympathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development.* New York: Cambridge University Press.

Lewis, M., & Michalson, L. (1983). *Children's emotions and moods: Developmental theory and measurement.* New York: Plenum.

Maccoby, E. (1980). *Social development: psychological growth and the parent-child relationship.* New York: Harcourt, Brace, Jovanovich.

Mood, D., Johnson, J., & Shantz, C. (1978). Social comprehension and affect matching in young children. *Merrill-Palmer Quarterly*, *24*, 63–66.

Roberts, W., & Strayer, J. (1987). Parents' responses to the emotional distress of their children: Relations with children's competence. *Developmental Psychology*, *23*, 415–422.

Roberts, W., & Strayer, J. (1996). Empathy, emotional expressiveness, and prosocial behavior. *Child Development*, *67*, 449–470.

Saarni, C. (1989). Children's understanding of strategic control of emotional expression in social transactions. In C. Saarni & P. Harris (Eds.), *Children's understanding of emotions* (pp. 181–208). New York: Cambridge University Press.

Schacter (1964). The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1, pp. 49–80). New York: Academic Press.

Selman, R. (1980). *The growth of interpersonal understanding: Developmental and clinical analyses.* New York: Academic Press.

Strayer, J. (1986). Children's attributions regarding the situational determinants of emotions in self and others. *Developmental Psychology*, *17*, 649–654.

Strayer, J. (1989). What children know and feel in response to witnessing affective events. In C. Saarni & P. Harris (Eds.), *Children's understanding of emotions* (pp. 259–289). New York: Cambridge University Press.

Strayer, J. (1993). Children's concordant emotions and cognitions in response to observed emotions. *Child Development*, *64*, 188–201.

Strayer, J., & Roberts, W. (1994, June) *Empathy, expressiveness, and prosocial behavior.* Presented at meetings of the Canadian Psychological Association, Penticton, BC, Canada.

Yarczower, M., & Daruns, L. (1982). Social inhibition of spontaneous facial expressions in children. *Journal of Personality and Social Psychology*, *43*, 831–837.

Zuckerman, M., Dorman, R., Larrence, D., & Speigel, N. (1981). Facial, autonomic, and subjective components of emotion: The facial feedback hypothesis versus the externalizer-internalizer distinction. *Journal of Personality and Social Psychology*, *41*, 929–944.