The Role of Person Familiarity in Young Infants’ Perception of Emotional Expressions

Ronit Kahana-Kalman and Arlene S. Walker-Andrews

This research investigated the role of person familiarity in the ability of 3.5-month-old infants to recognize emotional expressions. Infants (N = 72) were presented simultaneously with two filmed facial expressions, happy and sad, accompanied by a single vocal expression that was concordant with one of the two facial expressions. Infants’ looking preferences and facial expressions were coded. Results indicated that when the emotional expressions were portrayed by each infant’s own mother, infants looked significantly longer toward the facial expressions that were accompanied by affectively matching vocal expressions. Infants who were presented with emotional expressions of an unfamiliar woman did not. Even when a brief delay was inserted between the presentation of facial and vocal expressions, infants who were presented with emotional expressions of their own mothers looked longer at the facial expression that was sound-specified, indicating that some factor other than temporal synchrony guided their looking preferences. When infants viewed the films of their own mothers, they were more interactive and expressed more positive and less negative affect. Moreover, infants produced a greater number of full and bright smiles when the sound-specified emotion was “happy,” and particularly when they viewed the happy expressions of their own mothers. The average duration of negative affect was significantly longer for infants who observed the unfamiliar woman than for those who observed their own mothers. These results show that when more contextual information—that is, person familiarity—was available, infants as young as 3.5 months of age recognized happy and sad expressions. These findings suggest that in the early stages of development, infants are sensitive to contextual information that potentially facilitates some of the meaning of others’ emotional expressions.

INTRODUCTION

An important aspect of human interaction is the exchange of emotional information. Even for young infants, communication is based on an ability to extract meaning from their caregivers’ emotional expressions and the ability to respond to these signals adaptively (Stern, 1974). Developmental research on the emotional life of young infants suggests that dramatic and rapid changes occur in the first 2 years of life with respect to infants’ abilities to perceive and respond to the emotions of others (see Walker-Andrews, 1997, for a review). Such developmental changes contribute to infants’ ability to become active participants in social interactions (Malatesta, Culver, Tesman, & Shepard, 1989; Tronick, 1989).

Research on infants’ perception of emotion has focused primarily on their ability to discriminate emotional expressions, with particular emphasis on charting the developmental course of discrimination of facial expressions (for reviews, see Nelson, 1987; Oster, 1981; Walker-Andrews, 1988, 1997). Results indicate that at approximately 3 months of age, infants can discriminate among facial expressions of happiness, anger, fear, surprise, anger, and disgust (e.g., Barrera & Maurer, 1981; Kuchuck, Vibbert, & Bornstein, 1986; Nelson, 1987; Serrano, Iglesias, & Loeches, 1992; Young-Browne, Rosenfeld, & Horowitz, 1977). By 5 months of age, infants can discriminate among different vocal emotional expressions, but only when they are concurrently presented with a facial expression (A. J. Caron, 1988; Walker-Andrews & Grolnick, 1983; Walker-Andrews & Lennon, 1991). Although these results are informative with respect to infants’ ability to distinguish among different expressions, they tell us little about infants’ ability to extract any meaningful information from the expressions they appear to distinguish (see R. F. Caron, A. J. Caron, & Myers, 1982; Nelson, 1987; Oster, 1981; Walker-Andrews, 1988, 1997). Furthermore, findings from several studies suggest that infants’ discrimination of facial expressions is based merely on simple featural differences rather than on the recognition of the affective information (R. F. Caron, A. J. Caron, & Myers, 1985; Oster, 1981; but see Kestenbaum & Nelson, 1990). For example, R. F. Caron et al. (1985) found that 4- to 8-month-old infants discriminate “toothy” smiles from closed-mouth smiles and closed-mouth anger, but not from toothy anger.

In an attempt to determine whether infants merely discriminate featural aspects of expressions or detect
meaning in those expressions, Walker-Andrews (Walker, 1982; Walker-Andrews, 1986) presented infants with two films portraying a woman with two different expressions along with a single vocal expression corresponding to one of the facial expressions. In a series of experiments, both 5- and 7-month-old infants looked more to all facial expressions (happy, sad, angry, and neutral) when they were sound specified than when they were not. Even when the soundtrack was asynchronous with the facial displays, 7-month-old infants still looked preferentially at the affectively concordant display, demonstrating that synchrony is not the only correspondence between facial and vocal expressions detected by infants (Walker, 1982). In another experiment, only those infants who viewed facial expressions in an upright orientation looked appropriately to the sound-specified film, even though synchrony relations were retained in all displays. In addition, when Walker-Andrews (1986) showed infants synchronous happy and angry displays with part of the models’ faces concealed so that the synchrony between lip movements and vocalization was disguised, 7-month-old infants (but not 5-month-olds) still looked preferentially to the sound-specified emotion films. Walker-Andrews (Walker, 1982; Walker-Andrews, 1986, 1988) interpreted these findings as evidence that, at least by 7 months, infants detect information that is invariant across optic and acoustic displays of a single affective expression, thereby perceiving the bimodally presented expressions as unitary, meaningful events.

Soken and Pick (1992), using the same technique, further investigated the variables that may contribute to infants’ ability to match facial and vocal expressions. They created films that eliminated synchrony and rate information, yet retained the affective correspondence. Infants demonstrated the usual preference for the modified “point light” visual display (Bassili, 1979), which affectively matched the sound-specified emotion films. Walker-Andrews (Walker, 1982; Walker-Andrews, 1986, 1988) concluded that 7-month-old infants can detect the correspondence between facial and vocal expressions based on their affective meaning, and that they can discriminate between happy and angry expressions based on motion information alone.

Thus, experiments conducted on infants’ intermodal perception of emotions demonstrate that 5- to 7-month-old infants look preferentially to a dynamic facial expression accompanied by its characteristic vocal expression even when many of the relations are distorted or eliminated (Soken & Pick, 1992; Walker, 1982; Walker-Andrews, 1986, 1988). Many mother–infant studies using live interaction paradigms, however, suggest that infants recognize the emotional expressions of their own caregivers and respond to them meaningfully as early as 2 to 3 months of age (e.g., Beebe & Gerstman, 1980; Bigelow, 1998; Cohn & Ellmore, 1988; Cohn & Tronick, 1988; Field, 1977; Malatesta & Haviland, 1982; Stern, 1971, 1974; Trevathan, 1979).

One common method for evaluating young infants’ ability to engage in communicative exchanges with their mothers has been the still-face paradigm (Tronick, Als, Adamson, Wise, & Brazelton, 1978). In a typical still-face study there are three brief phases: (1) the mother interacts in her normal way, using facial and vocal expressions and touch; (2) the mother assumes a neutral and nonresponsive still face; and (3) the mother resumes normal interaction. Results from a host of studies have shown that during the still-face period, infants show increased looking away, increased arousal, less smiling, and more distress (Gusella et al., 1988; Lamb, Morrison, & Malkin, 1987; Stack & Muir, 1990, 1992; Toda & Fogel, 1993; Tronick et al., 1978; Tronick, Ricks, & Cohn, 1982). These patterns of response suggest that the cessation of maternal movement and expressiveness in the context of prior continuous maternal activity has a mildly stressful effect on infants as young as 2 to 3 months of age (Gusella et al., 1988; Legerstee, Pomerleau, Malcuit, & Feider, 1987; Matias & Cohn, 1993; Muir & Hains, 1993). Furthermore, the effect of the still face has been replicated with fathers (Braungart-Rieker, Murphy Garwood, Powers, & Notaro, 1998), an adult stranger (Kisilevsky et al., 1998; Stack & Muir, 1992), and even a televised image of the mother (Gusella et al., 1988).

Findings from several other mother–infant interaction studies suggest that by 3 months of age, communications between the mother and infant are quite sophisticated, with both partners trying to understand the intent of the other and respond appropriately. Haviland and Lelwica (1987) examined the responses of 10-week-old infants to their mothers’ live presentation of happy, sad, and angry emotional displays. Coding of infants’ facial responses demonstrated that 10-week-olds responded differentially and contingently to maternal emotional displays. Specifically, when mothers expressed joy, infants expressed more joy and interest. When mothers presented sad expressions, infants’ expressions of joy decreased, their mouthing behaviors increased, and they were more likely to gaze downward. During maternal presentations of anger, infants demonstrated increased anger expressions and their movements appeared to freeze. More recently, Izard et al. (1995) examined mother–infant interactions longitudinally from 2.5 to 9.5 months of age. Even at the youngest age tested, mothers’ expressions in positive conditions elicited more interest and joy from infants,
whereas negative conditions elicited more negative expressions from infants (see also Weinberg & Tronick, 1994). Similar results have emerged in studies of the effect of maternal vocalization (Fernald, 1992). For example, Fernald et al. (1989) showed that at 5 months of age, infants expressed more positive affect in response to expressions of approval and more negative affect in response to expressions of prohibition.

Furthermore, several studies of early mother–infant face-to-face interaction have documented early emotional responsiveness among dyads and suggest that mothers use this mode of interaction as a context for socialization of infants’ expressiveness according to cultural demands (Malatesta & Haviland, 1985). Specifically, Malatesta and Haviland (1982) found that mothers of 3- to 6-month-old infants displayed and reinforced primarily infants’ positive affect, a finding that supports the implicit premium on the expression of positive affect found in our culture. Although infant males and females displayed the same type and rate of facial expression change (with the one exception that infant females displayed more interest), the mothers responded differentially to the two genders. The mothers showed a significant increase, over infant age, in contingent smiling to the smiles of infant boys, and a decrease in contingent smiling, to the smiles of infant girls. The authors speculate that this difference may reflect the increased interest mothers may have in reinforcing positive affect in their male infants, who seem to have a greater predisposition to irritability. In addition, mothers imitated more male than female expressions, and they followed more female, as opposed to male, expressions with a different expression (see also Malatesta, Culver, Tesman, & Shepard, 1989).

In sum, infants’ affective communications with their caregivers have been examined predominantly in the context of face-to-face interaction situations. Results suggest that by 3 months of age, infants have already developed “expectations” about their mother’s behavior during social interactions and respond to violations of those expectations with meaningful affective changes (Gusella et al., 1988; Izard et al., 1995; Tronick et al., 1978). Furthermore, these studies have demonstrated that infants display a wide range of emotional expressions as early as 3 months of age, including interest, enjoyment, surprise, sadness/distress, anger, and discomfort/pain (Haviland & Lelwica, 1987; Izard, 1971; Malatesta & Haviland, 1982). The finding that young infants have such a wide repertoire of expressions combined with the fact that they display these emotional expressions contingent upon their mothers’ affect, suggests to us that infants recognize the meaning of their mothers’ expressions as early as 3 months of age. In contrast, findings from experimental investigations of infants’ recognition of emotional expressions suggest that only at around 7 months of age do infants match vocal and facial expressions of the same emotion (Soken & Pick, 1992; Walker-Andrews, 1986) or recognize that different examples of the same expression belong to the same category (e.g., Ludemann & Nelson, 1988).

Several differences between interaction studies and experimental studies might explain this apparent age difference. First, in the majority of laboratory studies, the presentation of emotional expressions is restricted to either the visual or auditory domain, whereas in face-to-face interaction studies, infants are exposed to multimodal emotion presentation in face, voice, touch, and gesture. Several investigators have argued that crucial affective information is conveyed in the dynamic motion of an emotional expression, but lost in static photographs (R. F. Caron et al., 1985; Ekman, Friesen, & Ellsworth, 1972; Fogel, 1993). Indeed, when A. J. Caron, R. F. Caron, and MacLean (1988; see also Walker-Andrews, 1986) used a standard habituation procedure but with dynamic expressions, 4-month-olds who were habituated to a happy dynamic expression (i.e., films) recovered their attention when a novel sad expression was presented; 7-month-olds also responded to differences between happy and angry expressions. A second difference that may contribute to the reported age difference is that in most studies the emotional expressions are portrayed by an unfamiliar actress, whereas in interaction studies they are typically portrayed by a familiar person (i.e., mothers). Indeed, Barrera and Maurer (1981) showed that 3-month-olds find it easier to discriminate among facial expressions portrayed by their own mothers than among the same expressions portrayed by a stranger.

What might mediate such early sensitivities in infants’ perception of emotions? First, several investigators have proposed that for young infants, dynamic, naturalistic, and multimodal displays may be the optimal stimuli (R. F. Caron et al., 1985; Fogel, 1993; Walker-Andrews, 1988, 1997). Infants’ auditory, visual, and haptic systems are well coordinated at birth, so they can attend to intermodal correspondences—that is, relations that remain constant across stimulus transformations—at an early age (Bower, 1974, 1977; Gibson, 1969; Bushnell, 1981; Walker-Andrews, 1997). Second, it may be that younger infants recognize emotions more readily when they encounter highly salient and familiar examples of these emotions, as would be the case when emotions are portrayed by their own mothers (Watson, 1985). This explanation is consistent with the finding that in face-to-face interactions, 4- to 5-month-old infants responded more contingently to their own mothers’ vocalizations and smiles than to the same frequency of smiling and vo-
calizations expressed by a female stranger (Bigelow, 1998).

This study was designed to examine the role of contextual information, such as person familiarity, in young infants’ perceptions of emotions in a controlled laboratory situation. Specifically, we examined whether person familiarity plays a role in the ability of young infants to perceive happy and sad emotional expressions. To test the possibility that person familiarity influences young infants’ ability to extract meaning from emotional expressions, we tested 3.5-month-old infants in an intermodal preference paradigm and compared their ability to detect the correspondence between vocal and facial expressions when portrayed by their own mothers versus an unfamiliar woman. Overall, we hypothesized that 3-month-old infants learn to detect the relation between the vocal and facial emotional expressions of their own mothers earlier than they detect such information portrayed by unfamiliar others.

Furthermore, in addition to infants’ attentional patterns, we examined their affective responsiveness to the emotional displays. Our interest in using affective responsiveness parameters, such as the intensity and duration of infants’ positive and negative facial expressions, is based on the assumption that the infants’ facial expressions of emotion may reflect their internal feeling states (Izard, 1971, 1977; Izard & Malatesta, 1987; Stern, 1985). Izard and Malatesta (1987) argued that the concordance between emotion expression and feeling state is a basic and adaptive relationship; one that evokes others’ caregiving interactions. This concordance is particularly important for preverbal infants who must communicate their feeling states through facial and vocal expression. Using both global and fine-grained coding systems, still-face procedural studies have demonstrated that by 3 months of age, infants respond to maternal affective displays with consistent and appropriate facial affect (e.g., Cohn & Tronick, 1988; Haviland & Lelwica, 1987; Izard et al., 1995; Tronick et al., 1978; Weinberg & Tronick, 1994).

Finally, according to differential emotions theory (Izard, 1971, 1977; Izard & Malatesta, 1987), emotional expressions of one member of an interpersonal affective relationship, particularly a mother–infant attachment relationship, tend to elicit emotions in the other member. A number of empirical studies with young infants support this proposition. For example, Termine and Izard (1988) demonstrated that 9-month-old infants displayed more joy and interest and looked longer at their mothers when the mothers expressed joy in a 2-min emotion-induction period. In contrast, infants displayed more sadness, anger, and gaze aversion when their mothers expressed sadness. Similar findings were documented with 10-week-old infants (Haviland & Lelwica, 1987; Izard et al., 1995), and also in the work of Stern (1985) and Tronick (e.g., Tronick et al., 1982) using the still-face paradigm. Thus, young infants appear to respond differentially to the discrete emotional expressions of their mothers. Based on the empirical studies and theoretical positions reviewed, we predicted that in our paradigm, the model’s emotional expression could be a stimulus or releaser of an emotional response from the infant. This response would vary according to the emotion displayed (happy versus sad) with an appropriate vocalization, and with the experiences that the infant has had with the person displaying the emotion (mother versus stranger).

To summarize, our study addressed three specific research questions with respect to the role of person familiarity in young infants’ perception of emotion: (1) Would 3.5-month-old infants detect the correspondence between facial and vocal displays of happy and sad expressions when the emotions were displayed by their own mothers? (2) Would infants detect the correspondence between maternal facial and vocal displays, even if synchrony relations between the face and voice were disrupted (i.e., when lip movements were asynchronous with the speech)? and (3) Would infants exhibit different patterns of affective responses across emotional displays performed by their own mothers versus those performed by unfamiliar women? To address these questions, we compared infants’ ability to match the facial and vocal expressions of their own mothers versus those of an unfamiliar woman. In addition, we examined infants’ affective reactions to different emotional displays. Infant affect measures that are typically used in live interaction studies were used in this study to provide more information with respect to the meaning of infants’ visual preferences.

**METHOD**

The three research questions guided the design and analyses of this study. First, to examine whether person familiarity influenced 3.5-month-old infants’ matching of vocal and facial displays, intermodal matching of a group that viewed an unfamiliar woman and a group that viewed their own mothers were compared. To explore the robustness of matching based on person familiarity, we tested another group of infants with films of their own mothers in which the facial and vocal displays were asynchronous (lip movements were asynchronous with the speech). Finally, infants’ patterns of affective responsiveness across all three infant groups were compared (unfamiliar, familiar synchronous, and
familiar nonsynchronous) to examine whether infants displayed different facial affect patterns across mother and stranger emotional displays and across the emotions displayed with appropriate vocalizations (i.e., happy versus sad). Given the role of infant gender previously reported in early face-to-face interaction studies (e.g., Malatesta & Haviland, 1982), this variable was taken into consideration in all comparisons.

Participants

Participants were 72 term infants (40 females and 32 males, gestational age >38 weeks, birthweight >2,500 g) tested between 14 and 18 weeks postnatally ($M = 114.92$ days, $SD = 10.41$). The sample included 88% White ($n = 63$), 5% biracial ($n = 4$), 3% Asian ($n = 2$), 3% Hispanic ($n = 2$), and 1% Black ($n = 1$) infants from English-speaking working- and middle-class intact (100% of parents were married) families. Approximately 18% of the parents had a high school degree or less (5.6%, mothers; 12.9%, fathers), about half had some college education or a college degree (57.7%, mothers; and 52.8%, fathers), and about one third had some postgraduate training or a postgraduate degree or less (5.6%, mothers; 12.9%, fathers), about half had some college education or a college degree (57.7%, mothers; and 52.8%, fathers), and about one third had some postgraduate training or a postgraduate degree (36.6%, mothers; 34.3%, fathers). The mothers’ average age was 32.9 years ($range = 25–42$ years); fathers’ average age was 34.72 years ($range = 26–54$ years). Half of the infants were firstborns (48.6%), 38.9% had only one sibling, and the remainder (12.5%) had two or more siblings.

Infants were recruited from birth announcements published in the local newspaper. Fourteen additional infants were tested but were excluded from the analysis because of experimental error ($n = 2$), infant crying ($n = 1$), infants’ failure to meet looking-time criteria ($n = 5$), or failure of maternal films ($n = 6$) to meet criteria (see Measures and Coding section for details). Attrition rates did not vary across experimental groups. The infants had no acute or chronic pre- or perinatal medical complications, and no apparent visual or auditory deficits; all were in good health.

Procedure

Infants were tested in the laboratory at Rutgers University. After infants were acclimated to the laboratory and the mothers signed the consent forms, infants were seated in an infant seat. The infant seat was situated approximately 85 cm away from two adjacent 20-inch (50.8-cm) color TV monitors. The plane of the monitors was parallel to the infants’ eyes. The soundtrack came from a centrally located speaker. This ensured that infants could not localize the sound to either the right- or left-hand side. A videocamera positioned above the monitors and behind a screen recorded infants’ faces for later coding of their looking patterns and facial affect. Blinds (of a neutral color) surrounded the monitors to obscure infants’ views of the experimenters and mothers. The lights in the experimental room were dimmed and a blinking line of red lights was flashed between the monitors to attract infants’ visual attention to the center of the apparatus before the presentation began.

Two videotaped facial expressions (one happy, one sad) presented simultaneously on the two adjacent monitors. When shown to the infants, the images filled the 50.8-cm monitor; the size of each face was 18 cm wide at the widest point and 28.5 cm long at the longest point. When viewed from a distance of 85 cm, each of the target faces on average subtended a visual angle of 12.18 degrees. The films presented a female from the shoulders up portraying each of the two selected emotions, happy and sad. In the unfamiliar infant group, the experimental tapes presented an unfamiliar female expressing the selected emotions. In the familiar synchronous and familiar nonsynchronous groups, the infants’ own mothers presented the same two emotions. The vocal expression affectively matched only one of the two visual events. Infants were presented with four 25-s trials. The sound-specified emotion alternated sides on each trial. A 2-s blank screen interval occurred between test trials. Infants were assigned to one of the three groups. Each of the three infant groups ($n = 24$) was counterbalanced in terms of sound-specified emotion (happy or sad), and the side on which the sound-specified display first appeared.

Stimulus Materials

All women (i.e., unfamiliar female models and mothers of the infants) were filmed at their homes expressing the target emotions (happy and sad) using a camcorder. All the women (models included) were mothers of young infants who were roughly the same age (2 to 6 months of age), with similar demographic characteristics. Throughout the taping, they were instructed to maintain a still body and head posture and to direct their gaze at the camcorder. They were filmed from the shoulders up against a plain white background under normal lighting.

To induce the target emotion, the women were asked to think about situations that would make them happy or sad. This method of inducing affect has been used in previous studies (e.g., Haviland & Lelwica, 1987). The experimenter briefed the women about the important facial muscle changes that are commonly used to express each emotion based on de-
scriptions by Ekman and Friesen (1978). Prior to the videotaping, the women were asked to practice their facial expressions using a mirror. With respect to their vocal expressions, they were instructed to try and speak continuously to the camera, as they would to a young baby about any event. Specifically, the women were asked to use a high-pitched and rhythmical speech when portraying the happy expression, and to avoid both long pauses and a very low pitched voice for the sad vocal expression. They were also asked to avoid raising or lowering their speech loudness abruptly or dramatically, and to refrain from using their infants’ names or making any nonverbal vocalizations such as tongue clicking.

To create the stimuli films for the group of infants presented with an unfamiliar female, we initially filmed four different women. Segments of their emotional expressions were viewed silently by 10 adults who were naive to the purposes of the study. The adults were asked to identify each emotional expression segment and to rate the valence of each. To provide infants with the exemplar, the model whose emotional expressions were rated most accurately was selected for the portrayal of all the experimental tape trials. Next, to select the exemplars of this model’s emotion portrayals, 150 college students who were naive with respect to the purposes of the study were asked to identify each emotional expression segment and to rate the intensity (on a scale of 0 to 5) of each of the emotions. The undergraduate coders coded by two observers who were naive to the gender of the mothers for the particular emotion. Based on these criteria, six mother–infant dyads were replaced.

Measures and Coding

Infant Looking Time

Coding tapes were made from the original tape to eliminate the vocal track. This ensured that coders would be naive with respect to the identity and side of the sound-specified emotion. Infants’ looking was coded by two observers who were naive to the gender of the infants and the identity of the female portraying the emotions (i.e., unfamiliar female or the infant’s own mother). These two observers coded tapes for four pilot participants together and then for an additional five infants separately until 90% interrater reliability was established.

For each of the four 25-s trials, the observers judged each look as directed toward the right monitor, toward the left monitor, or away. A time code recorded on the videotapes allowed the judgment of time within 1/30 of a second. The time duration of infants’ looking toward each film (left or right) or away from the films was recorded and expressed as the proportion of total looking time (PTLT) toward the films across all four trials. Intermodal preference was defined as the total number of seconds spent looking at the sound-specified displays divided by the total number of seconds spent looking at both displays across all four 25-s trials. The larger this percentage, the stronger the preference for the sound-specified visual display (e.g., Soken & Pick, 1992; Walker, 1982; Walker-Andrews, 1986, 1988).

Based on looking-time coding, data were omitted if, across all four 25-s trials: (1) the infant looked at the
Infant Facial Affect

Real-time coding of global affect measures. To obtain a measure of overall responsiveness to the stimuli, the infants’ videotapes were first viewed in real time with the sound turned off, after which observers answered two questions on a 5-point scale: (1) How engaged (i.e., interested, interactive) do you think the infant was (1 = not at all, 5 = totally)? and (2) What is your overall perception of the infant’s emotional state (1 = experiencing negative or unpleasant affect [distress, stress, agitated], 5 = experiencing positive and pleasant affect [happiness, joy, excitement])? The mean interobserver percent agreement for Questions 1 and 2 was .86% (κ = .72) and .93% (κ = .83), respectively.

Frame-by-frame analysis of affect-specific measures. Infant positive and negative affect was further scored from the videotapes by two additional observers who were naive to group (familiarity) and condition (emotion). Six different facial affect behaviors were scored. These affect codes were based on previously developed scales by Thompson and Lamb (1984) and the AFFEX System (Izard, Dougherty, & Hembree, 1983). Positive affect codes included faint, full, and bright smiles, generally defined as upward movements of the corners of the mouth with raised cheeks, smooth forehead, and neutral brows. Negative affect codes included frowning, low-level distress and cry faces, generally defined as downward movements of the corners of the mouth, raising of the inner corners of the brows, and narrowing of the eyes (see the Appendix for complete code descriptions).

The duration of each affect behavior was coded from the onset of the expression through its apex to the time any facial configuration changed or returned to a neutral state. Coding was performed with the James Long Video Coding System (VCS; Long, 1996). This hardware system connected to a videotape recorder and a compatible PC that placed a time code onto VHS-format videotapes. The software of this system created a time-coded data file in which the onset and offset of mutually exclusive and exhaustive coding units were recorded. The six target affect codes were organized into a set of VCS codes (see Appendix).

Coding was performed by two researchers who, prior to scoring, studied the AFFEX training videotapes. Reliability was assessed by having both coders scoring the same affect code in the same .1 s (i.e., >1 ms difference was considered a disagreement). This criterion required agreement in both side and duration of each infant look. The average kappa statistics calculated on the duration and side (left, right, or away) of each look was .93 (range = .67–1.00, SD = .09). The mean percent agreement was 98% (range = 81–100%, SD = 4.55).

RESULTS

Intermodal Matching: Looking Time

Experiment 1A: The Role of Person Familiarity

Infants in all groups found the visual displays compelling and spent an average of 92.8% (SD = 9.78) of total available time looking at the events. We conducted a 2 × 2 × 2 (Gender × Order × Emotion: happy versus sad) ANOVA with PTLT toward the sound-specified events as the dependent measure. There were no effects for order or gender and no interactions between them, ps < .10. The data from both genders and across order of presentation were therefore combined for all subsequent analyses; however, emotion was retained as a factor throughout because this variable revealed a significant main effect, F(1, 47) = 19.30, p < .00, but no significant interactions with either order or gender. The main effect of emotion indicated that overall, infants looked longer at happy facial expressions than at sad facial expressions.

One research question asked whether 3.5-month-old infants are able to detect information that is invariant across the visual and acoustic presentations of a single affective expression, and whether there are
any effects of person familiarity. More specifically, do infants look longer toward facial expressions that are sound specified when they are exhibited by their own mothers? To address this question, we compared data from infants who watched their own mothers perform happy and sad expressions with data from infants who watched the same two expressions performed by an unfamiliar woman. A 2 × 2 (Group: familiar versus unfamiliar × Emotion: happy versus sad) ANOVA was conducted with PTLT toward the sound-specified events as the dependent measure to examine the effect of familiarity. The analysis revealed a main effect for group, F(1, 47) = 5.22, p < .03, and a main effect of emotion, F(1, 47) = 19.30, p < .00. Inspection of Table 1 shows that the PTLT looking toward the sound-specified facial expression was higher when infants were presented with their own mothers versus an unfamiliar female. Also, in both groups, infants looked longer at happy facial expressions than at sad facial expressions.

Two separate one-way ANOVAs were conducted to compare infants' visual preferences for the sound-specified facial expressions within each of the three groups. The primary between-subjects analysis was the average PTLT toward a facial expression (happy or sad) when it was sound specified versus when it was silent. Such a comparison examines whether a vocalization influences or biases infants' looking more toward the visual event that it matches than toward the one that it does not match (e.g., Spelke, 1979; Walker-Andrews, 1982, 1986). For the group of infants who were presented with their own mothers,

<table>
<thead>
<tr>
<th>Table 1 Mean PTLT toward Filmed Facial Expression While Silent and When Accompanied by an Appropriate Soundtrack by Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound Specified Film</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Unfamiliar**</td>
</tr>
<tr>
<td>Happy</td>
</tr>
<tr>
<td>Sad</td>
</tr>
<tr>
<td>Familiar synchronous*</td>
</tr>
<tr>
<td>Happy</td>
</tr>
<tr>
<td>Sad</td>
</tr>
<tr>
<td>Familiar nonsynchronousb</td>
</tr>
<tr>
<td>Happy</td>
</tr>
<tr>
<td>Sad</td>
</tr>
</tbody>
</table>

** Twenty-four infants participated in Experiment 1A.

b Twenty-four infants participated in Experiment 1B.

*p < .01; ***p < .001.

Experiment 1B: The Role of Temporal Synchrony

To determine whether infants were matching the facial and vocal expressions on the basis of affective correspondence (i.e., between vocal and facial expressions) rather than on the basis of temporal synchrony (i.e., mouth movement and vocal relations), an additional group of 24 same-age infants (M = 118.87 days, SD = 8.79) were tested. These infants were recruited and tested using the same procedures as in Experiment 1A. There were no demographic or experimental differences between this group of infants and the groups who participated in Experiment 1A.

The discrepancy between the vocalization and lip movements was created by inserting a 5-s delay between the vocal and facial displays. The delay between the vocal and sound-specified facial display ensured
that neither soundtrack matched the video on the basis of lip movements. Once again, the primary focus was on whether infants looked longer at happy or sad facial expressions when these were accompanied by an affectively matching soundtrack than when they were silent. A one-way ANOVA conducted to measure the PTLT toward Happy Films × Condition (happy versus sad) revealed a main effect for condition, \( F(1, 23) = 20.40, p < .001 \), indicating that infants looked significantly longer at happy facial expressions when these were accompanied by an affectively matching soundtrack than when they were silent. Similarly, an ANOVA conducted on the PTLT toward Sad Films × Condition (happy versus sad) revealed a main effect for condition, \( F(1, 23) = 20.40, p < .001 \), indicating that the primary focus was on whether infants looked significantly longer at happy facial expressions when these were accompanied by an affectively matching soundtrack than when they were silent. Similarly, an ANOVA conducted on the overall PTLT toward the sound-specified film as the dependent measure and emotion (happy versus sad) as the between-subjects effect yielded overall effects for group, \( F(4, 116) = 6.48, p < .001 \), condition, \( F(2, 58) = 3.20, p < .048 \), Condition × Gender interaction, \( F(2, 58) = 3.63, p < .033 \), and a Group × Condition interaction, \( F(4, 116) = 4.58, p < .002 \). The univariate \( F \) tests for each dependent measure were inspected (see Table 2) and post hoc Tukey tests \( (p < .05) \) were performed to identify differences across groups in infants’ global ratings. Results showed that when happy was the sound-specified emotion, infants across groups were rated as experiencing more positive affect and as more interested and engaged than when sad was the sound-specified emotion. More importantly, infants from the groups who observed their own mothers were rated as more positive and more engaged when happy was the sound-specified emotion. The infants who observed the unfamiliar female received global scores that were both stable across emotions and lower than those of infants who observed their own mothers (see Table 3 for means).

### Table 2 Significant Univariate Effects for Global Affective Quality Measures

<table>
<thead>
<tr>
<th>Source</th>
<th>( df )</th>
<th>Interactivity</th>
<th>Emotional State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2, 71</td>
<td>11.70***</td>
<td>7.20*</td>
</tr>
<tr>
<td>Condition</td>
<td>1, 71</td>
<td>4.48*</td>
<td>4.64*</td>
</tr>
<tr>
<td>Group × Condition</td>
<td>2, 71</td>
<td>7.91**</td>
<td>5.08*</td>
</tr>
<tr>
<td>Condition × Gender</td>
<td>1, 71</td>
<td>ns</td>
<td>ns</td>
</tr>
</tbody>
</table>

\( *p < .05; **p < .01; ***p < .001. \)

### Table 3 Means for Measures of Global Affective Quality by Group and Condition

<table>
<thead>
<tr>
<th>Group</th>
<th>Engage.</th>
<th>Emotional State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Happy</td>
<td>Sad</td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>2.33(a)</td>
<td>3.09(a)</td>
</tr>
<tr>
<td>Familiar synchronous</td>
<td>4.67(b)</td>
<td>3.17(b)</td>
</tr>
<tr>
<td>Familiar nonsynchronous</td>
<td>4.33(b)</td>
<td>3.54(b)</td>
</tr>
</tbody>
</table>

Note: Judgments were made on 5-point scales: for engagement, 1 = not at all, 5 = totally; for emotional state, 1 = negative affect, 5 = positive affect. Means in the same column that do not share subscripts differ at \( p < .05 \) in the Tukey significance difference comparison.

### Responsiveness: Facial Affect

#### Global Affect Measures

**Quality of affect.** Two measures that reflect the overall valence and interactiveness of infants were calculated (see real-time coding in Method section), engagement and emotional state, and found to be correlated, \( r(71) = .57, p < .001 \). Therefore, we conducted a \( 3 \times 2 \times 2 \) (Group: unfamiliar, familiar synchronous, familiar nonsynchronous × Gender × Condition: happy versus sad) MANOVA, using the two global measures of emotional quality. The analysis yielded overall effects for group, \( F(4, 116) = 6.48, p < .001 \), condition, \( F(2, 58) = 3.20, p < .048 \), Condition × Gender interaction, \( F(2, 58) = 3.63, p < .033 \), and a Group × Condition interaction, \( F(4, 116) = 4.58, p < .002 \). The univariate \( F \) tests for each dependent measure were inspected (see Table 2) and post hoc Tukey tests \( (p < .05) \) were performed to identify differences across groups in infants’ global ratings. Results showed that when happy was the sound-specified emotion, infants across groups were rated as experiencing more positive affect and as more interested and engaged than when sad was the sound-specified emotion. More importantly, infants from the groups who observed their own mothers were rated as more positive and more engaged when happy was the sound-specified emotion. The infants who observed the unfamiliar female received global scores that were both stable across emotions and lower than those of infants who observed their own mothers (see Table 3 for means).

In sum, infants as young as 3.5 months of age demonstrated the ability to match facial and vocal expressions of their own mothers, even when voice–lip synchrony was disrupted. In contrast, no evidence of matching or visual preference was found when infants were presented with the same emotions portrayed by an unfamiliar female.
Quantity of affect. Three measures reflecting the overall quantity of infant facial expression were calculated (see frame-by-frame analysis in Method section): (1) facial expression time (that is, the percentage of the coding interval during which the infant expressed facial affect different from neutral, either positive or negative), (2) number of alternating expressions (that is, the number of times the infants’ facial expression changed during the testing interval), and (3) variability of expressions (that is, the number of different facial configurations produced during the testing interval out of six possible coded expressions recorded). These measures were selected to represent previously described differences in infant affect in earlier studies of infant emotion expressiveness (e.g., Camras et al., 1998; Thompson, 1990). Pearson correlations between these measures ranged from .75 to .88 (\(p < .001\)). We therefore conducted a 3 \( \times \) 2 \( \times \) 2 (Group: unfamiliar, familiar synchronous, familiar nonsynchronous \( \times \) Gender \( \times \) Condition: happy versus sad) MANOVA using all three quantity-of-expression measures as the dependent variables. There was a significant multivariate effect for condition, \(F(3, 58) = 4.08, p < .01\), and the univariate \(F\) tests were significant for all three dependent measures (see Table 4 for means and summary of significant results). These effects demonstrate that when happy was the sound-specified emotion, infants expressed emotion more frequently (\(M = 21\%\) versus 11\%), changed their expressions more (\(M = 9.06\) versus 5.56), and showed increased variability of expressions (\(M = 3.00\) versus 2.03) relative to when sad was the sound-specified emotion. This overall pattern of results suggests that infants’ expressions were both held longer and changed more often when happy was the sound-specified emotion. The MANOVA also revealed a significant Condition \( \times \) Gender interaction, \(F(3, 58) = 3.19, p < .03\), that showed significant univariate effects for facial expressive time, \(F(1, 71) = 4.70, p < .034\), and expressive variability, \(F(1, 71) = 8.52, p < .005\). This indicates that female infants responded differently to happy than to sad, whereas males were intermediate and stable.

Affect-Specific Measures

To further examine the infants’ facial behavior we considered the six affect codes obtained in the frame-by-frame analysis in several ways. Following procedures used in several previous studies (e.g., Camras et al., 1998; Kagan et al., 1994), we determined the number of each type of positive and negative facial configuration (i.e., three positive affect codes and three negative affect codes as described in the Appendix), as well as the latency to the first positive affect code and to the first negative affect code. Given the Condition \( \times \) Gender interaction reported above, all analyses were carried out first with gender as a between-subjects factor. Because no further main effects or interactions were found with gender, the next set of analyses are presented collapsed across this variable. All patterns of results were otherwise identical.

Positive affect codes: Smiles. Positive affect codes included three types of smiles (described fully in the Appendix): (1) faint smiles, (2) full smiles, and (3) bright smiles. In the literature, smiles accompanied by cheek raising (i.e., bright smiles in our coding) have been referred to as Duchenne smiles. These are hypothesized to reflect genuine positive emotion, whereas non-Duchenne smiles are hypothesized to be social smiles (Camras et al., 1998; Dickson, Walker, & Fogel, 1997; Ekman, Davidson, & Friesen, 1990). With respect to differences in smiling during infancy, Fox and Davidson (1988) demonstrated that in 10-month-old infants, Duchenne smiles were associated with mother’s approach, whereas non-Duchenne smiles were associated with a stranger’s approach. For infants and toddlers, open-mouth smiles are more common while playing with an attentive playmate, whereas closed-mouth smiles occur more frequently during toy-centered play (Demos, 1982; Jones & Raag, 1989). In our study, we explored whether these different types of smiles would occur in different frequencies across conditions and familiarity of the person displaying the emotion. Given previous findings, we speculated that infants would display more open-mouth smiles (i.e., full smiles) and smiles involving cheek raising (i.e., bright smiles) when happy was the sound-specified emotion and when the expressions were portrayed by their own mothers.

Two dependent measures were calculated with respect to the positive expressions: (1) duration (propor-

---

Table 4 Significant Univariate Effects and Means for Global Affective Quantity Measures by Condition

<table>
<thead>
<tr>
<th>Affective Quantity Variable</th>
<th>Univariate (F)</th>
<th>(M)</th>
<th>(SD)</th>
<th>(M)</th>
<th>(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Facial expression time</td>
<td>6.73**</td>
<td>.21</td>
<td>.19</td>
<td>.11</td>
<td>.13</td>
</tr>
<tr>
<td>2. Number of alternating</td>
<td>6.54**</td>
<td>9.06</td>
<td>6.54</td>
<td>5.56</td>
<td>5.37</td>
</tr>
<tr>
<td>Expressions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Variability of expression</td>
<td>12.52***</td>
<td>3.00</td>
<td>1.12</td>
<td>2.03</td>
<td>1.30</td>
</tr>
<tr>
<td>(range = 0–6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) All variables were calculated based on facial affect (both positive and negative) expressed during the testing interval.

\(^b\) Mean proportion of seconds during which infants expressed facial affect different from neutral.

\(* \ p < .01; ** \ p < .001.\)
tion of total presentation time in which smiling occurred) and (2) latency (in seconds) to the emergence of the first smile. Each measure was analyzed in a $3 \times 2$ (Group × Condition) ANOVA. For smile duration there were significant main effects for group, $F(2, 71) = 3.82, p < .027,$ and condition, $F(1, 71) = 8.24, p < .006,$ as well as a Condition × Group interaction, $F(2, 71) = 6.15, p < .004$ (see Tables 5 and 6 for significant univariate effects and means). When happy was the sound-specified emotion, infants who watched their own mothers spent more time smiling at the films than did infants who watched the unfamiliar female in the same condition. Infants spent less time smiling when sad was the sound-specified emotion, but did not differ in the duration of smiling according to whether they viewed their own mother or an unfamiliar female. These findings mirror the findings previously reported using the real-time global ratings. There were no significant main effects or interactions in terms of latency to the first smile (see Table 5). Nevertheless, inspection of the means in Table 6 shows that the shortest latency to the first smile occurred when infants watched their own mothers’ synchronous happy expressions.

Next, we examined the number of each type of smile coded (i.e., faint, full, or bright smiles). Because these were all correlated, $r$ ranged from .46 to .62, $p < .01,$ they were entered as dependent measures in a $3 \times 2$ (Group × Condition) MANOVA. This yielded a multivariate main effect for group, $F(6, 126) = 2.17, p < .05;$ and the univariate tests (summarized in Table 5) showed that infants in the two familiar groups expressed a greater number of full and bright smiles than did infants in the unfamiliar group; the groups did not differ in the number of faint smiles. Post hoc Tukey tests showed that infants in the synchronous familiar group expressed a significantly greater number of full and bright smiles than did infants who watched the films of the unfamiliar female ($p < .01$). The same trend was apparent with respect to infants who viewed their own mothers, but with nonsynchronized vocalization; however, these differences did not reach significance, $p < .24$ and $p < .09$ for full and bright smiles, respectively.

The MANOVA also yielded a significant multivariate effect for condition, $F(3, 63) = 5.63, p < .002,$ that showed that infants exhibited a greater number of full and bright smiles when happy was the sound-specified emotion (see Table 5 for significant univariate tests). Finally, a Group × Condition interaction, $F(6, 126) = 1.98, p < .02,$ followed by significant univariate effects for the number of faint, full, and bright smiles (see Table 5), suggested that infants who watched their own mothers expressed a greater number of smiles of all intensities, and smiled more often when happy was the sound-specified emotion. In contrast, infants who

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Smiling Duration</th>
<th>Latency to First Smile</th>
<th>Number of Faint Smiles</th>
<th>Number of Full Smiles</th>
<th>Number of Bright Smiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2, 70</td>
<td>3.82**</td>
<td>ns</td>
<td>ns</td>
<td>4.31*</td>
<td>4.46*</td>
</tr>
<tr>
<td>Condition</td>
<td>1, 70</td>
<td>8.24*</td>
<td>ns</td>
<td>ns</td>
<td>11.58***</td>
<td>11.10***</td>
</tr>
<tr>
<td>Condition × Group</td>
<td>2, 70</td>
<td>6.15**</td>
<td>ns</td>
<td>3.40*</td>
<td>3.84*</td>
<td>3.23*</td>
</tr>
</tbody>
</table>

*p < .05; ** p < .01; *** p < .001.

Table 6 Means for Positive and Negative Affect Measures by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Happy Duration</th>
<th>Happy Latency to First Smile</th>
<th>Sad Duration</th>
<th>Sad Latency to First Smile</th>
<th>Total Duration</th>
<th>Total Latency to First Smile</th>
<th>Happy Duration</th>
<th>Happy Latency to First Smile</th>
<th>Sad Duration</th>
<th>Sad Latency to First Smile</th>
<th>Total Duration</th>
<th>Total Latency to First Smile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfamiliar</td>
<td>3.03a</td>
<td>7.91a</td>
<td>5.50a</td>
<td>17.51</td>
<td>17.40</td>
<td>17.46a</td>
<td>17.85</td>
<td>5.87</td>
<td>11.86</td>
<td>44.20</td>
<td>26.67</td>
<td>36.98a</td>
</tr>
<tr>
<td>Familiar synchronous</td>
<td>22.67b</td>
<td>5.79a</td>
<td>14.23b</td>
<td>9.20</td>
<td>20.08</td>
<td>12.83b</td>
<td>3.44</td>
<td>2.89</td>
<td>3.16b</td>
<td>49.37</td>
<td>25.53</td>
<td>36.54b</td>
</tr>
<tr>
<td>Familiar nonsynchronous</td>
<td>15.76b</td>
<td>5.38a</td>
<td>10.57b</td>
<td>12.25</td>
<td>11.70</td>
<td>14.23b</td>
<td>3.19</td>
<td>1.30</td>
<td>2.25b</td>
<td>32.84</td>
<td>31.45</td>
<td>30.18a</td>
</tr>
</tbody>
</table>

Note: Means in the same column that do not share subscripts differ at $p < .05$ in the Tukey significance difference comparison.
watched the unfamiliar female expressed significantly fewer smiles (faint, full, or bright), and the number of their smiles did not differ across the emotion that was sound specified.

Negative affect codes: Sadness/distress. Negative affect codes included three types of sadness/distress expressions (described fully in the Appendix): (1) frowning, (2) low-level distress, (3) cry faces. Two dependent measures were calculated with respect to negative expressions: (1) duration (proportion of total presentation time in which a negative affect occurred) and (2) latency (in seconds) to the emergence of the first negative expression (e.g., Thompson 1990). Each measure was analyzed in a $3 \times 2$ (Group $\times$ Condition) ANOVA. For duration, there was a significant main effect for group, $F(2, 71) = 4.01$, $p < .023$. Post hoc Tukey tests revealed that the mean duration of negative affect was significantly longer for infants who observed the unfamiliar female than for those who observed their own mothers ($ps < .05$; see Table 6 for means). There were no significant main effects or interactions in terms of latency to the first negative expression; however, observation of the means suggests that it took longer for infants to show a negative expression when happy was the sound-specified emotion (see Table 6 for $Ms$ and $SD$s). Whereas none of the infants in the familiar synchronous group showed any full cry faces, there were 11 cry faces among the infants who observed the unfamiliar female, and 7 among the infants who observed their mothers but with unsynchronized voices, $\chi^2(2, N = 18) = 8.00$, $p < .018$.

In sum, both real-time coding and frame-by-frame analysis showed that infants who observed their own mothers expressed more positive affect and less negative affect, especially when maternal happy facial expressions were sound specified. Smiles directed at mothers were also more intense and involved more cheek raising and mouth opening. In contrast, infants who watched the unfamiliar woman showed more negative affect overall, and they did not show the same increases in positive affect when the happy expression was sound specified. In addition, infants in this group did not exhibit many full or bright smiles.

**DISCUSSION**

In this study, we examined the role of person familiarity in infants’ early recognition of emotional expressions in an intermodal preference procedure. As expected, the results of this experiment indicated that when emotional expressions were portrayed by their own mothers, infants as young as 3.5 months of age detected the intermodal correspondences between the facial and vocal displays, as demonstrated by their increased looking time to sound-specified films. Young infants perceived the components of the happy and sad expressions of their own mothers as part of a unified, multimodal expression, even when the maternal facial and vocal expressions were presented out of temporal synchrony. Infants also looked preferentially toward the happy facial expression when happy and sad films were presented simultaneously and in synchrony with the vocalization, indicating that infants found this filmed emotion the more compelling of the two films. In contrast, infants did not show by their looking patterns that they could detect the correspondence between vocal and facial displays of the same two emotions when these were portrayed by an unfamiliar woman.

The affective responsiveness of the infants also differed across emotions and familiarity with the person portraying the emotions (i.e., mother versus unfamiliar woman). Global affective measures showed that when happy was the sound-specified emotion, infants were more expressive: they showed greater variability of affective expression and increased the number of alternating expressions. They were also rated as experiencing more positive affect and as more interested and engaged, particularly when the emotion displays were portrayed by their own mothers. Furthermore, more affect-specific measures of infants’ smiles and negative expressions showed that the infants who viewed their own mothers spent more time smiling at the films than did infants who viewed the unfamiliar woman; the groups did not differ in the duration of smiling when sad was the sound-specified emotion. Moreover, infants produced a greater number of full and bright smiles when happy was the sound-specified emotion, particularly when they viewed the happy expressions of their own mothers. Finally, the average duration of negative affect was significantly longer for infants who observed the unfamiliar woman than for those who observed their own mothers. These results speak both to the importance of multimodal, dynamic presentations in infants’ perception and to their early sensitivity to maternal expressions.

**Recognition of Emotional Expressions**

In general, the results of this study suggest an early sensitivity to expressions, especially when these are experienced as dynamic, multimodal, and familiar events. The finding that richer, more naturalistic displays of emotions facilitate discrimination at younger ages is consistent with previous findings from studies using dynamic films in the intermodal preference paradigm (see Walker-Andrews, 1997, for a review). What is par-
particularly interesting in this study is that infants detected and responded to the affective correspondences in their own mothers’ facial and vocal expressions, even when synchrony relations between the face and voice were disrupted. They did not, however, match the synchronous vocal and facial expressions of an unfamiliar woman. Such a pattern implies that what accounts for infants’ ability to detect correspondences between facial and vocal affective displays is their ability to extract a common meaning from their mothers’ affective displays. This explanation builds on and extends proposals with respect to intermodal perception of emotional expressions previously advanced by Walker-Andrews (1997). It is based on what is known about infants’ perception in general and underscores the importance of perception for adaptation (see Lewkowicz & Lickliter, 1994; Rose & Ruff, 1987). The initial unity of the sensory systems is considered to be related to the detection of amodal properties, such as shape, size, substance, and intensity, given that these properties can be picked up by different sensory systems without prior coordinated experience (Walker-Andrews, 1997; Walker-Andrews & Gibson, 1984). Development in all modalities (haptic and visual, see e.g., Gibson & Walker, 1984; Meltzoff & Borton, 1979; visual and auditory, see e.g., Kuhl & Meltzoff, 1982; Spelke, 1976) is based on the extraction of invariants and the differentiation of distinctive features (but see Piaget, 1952 for a different account).

Given this early ability of young infants to detect multimodal relations, Walker-Andrews (1997) suggests that through the detection of intermodal invariants, infants also discover the meaning of emotional expressions. She asserts that perceptual development is characterized by increasing differentiation of the information for affect: “infants may first recognize the affective expressions of others as part of a unified multimodal event that has a unique communicative affordance” (Walker-Andrews, 1997, p. 449), only later recognizing the same information for affect in vocal expressions and then facial expression alone. As stated earlier, the results of this study provide further evidence that supports this proposal. That is, the ability to detect the correspondence between visual and acoustic displays of emotion has been demonstrated in this study with infants as young as 3.5 months of age, when more contextual information was available.

Early Sensitivity to Maternal Expressions

Additional “contextual information” was used in this study, in the form of the infants’ own mothers’ ability to portray the emotional displays. The looking-time data demonstrate that infants detected intermodal correspondences only for their own mothers’ renditions of happy and sad emotions. This intermodal preference for the sound-specified emotion film of mothers was replicated even when temporal synchrony between the vocal and visual displays was eliminated. Infants’ affective responses to the sound-specified expressions of their own mothers dovetail with this evidence to show infants’ earlier recognition and responsiveness to their own mothers’ affective signals compared with those of an unfamiliar female. Three factors may help explain infants’ proficiency: (1) mothers’ expertise in presenting expressions, (2) increased exposure by an infant to maternal expressions, and (3) the context in which infants experience maternal expressions. Each of these will be considered.

First, one could argue that the affective behaviors expressed by mothers, especially their vocal intonations, included prosodic features that were more salient to infants. Parents often modify their speech when addressing young infants. They typically do so by elevating pitch, exaggerating pitch changes, slowing tempo, simplifying vocabulary and syntax, increasing the number of repetitions, and lengthening pauses between words (Fernald et al., 1989; Fernald & Simon, 1984; Grieser & Kuhl, 1988). This speaking style, commonly referred to as infant-directed (ID) speech, has been observed with infants from birth to 2 years of age (McLeod, 1993), and several studies have demonstrated that it elicits stronger visual and affective responding from infants than does adult-directed (AD) speech (Cooper & Aslin, 1990; Fernald, 1985; Fernald & Kuhl, 1987; Pegg, Werker, & McLeod, 1992; Werker & McLeod, 1989). It is possible that infants in the present study were more attentive to the emotion displays portrayed by their own mothers because the facial and vocal expressions were more infant directed. We are not convinced by this explanation, however, for several reasons. First, all women were instructed to use ID speech during videotaping, and the unfamiliar female model was also a mother of a young infant. Second, infants in both conditions (i.e., familiar versus unfamiliar) were equally attentive to the displays and looked at the films 90% of the time available. Finally, adults’ ratings of the facial expressions of the mothers and the unfamiliar female did not reveal any differences with respect to valence or intensity.

An alternative explanation for the infants’ performance is that by 3.5 months of age, infants simply have been exposed more often to their own mothers’ expressions and they have learned to associate her voice and face more readily. During the early months of life infants are frequently engaged in affective exchanges with their own mothers. The period of 3 to 6 months of age has also been identified as a peak time
in dyadic face-to-face play, and as an important context for the socialization of emotion (Lewis & Michelson, 1983; Malatesta & Haviland, 1985). In their second-to-second analysis of maternal and infants’ expressions, Malatesta and Haviland (1982) found that mothers displayed an average of eight emotional expressions per minute. Assuming that mothers spend about 3 hours a day in face-to-face interaction with their infants, this rate implies that young infants are exposed to hundreds of maternal emotional signals every day.

Whereas mere exposure may be tenable, it may be only partially useful. We prefer an explanation that involves familiarity as an intrinsic contextual factor that adds meaning to the task of early recognition of emotion (Baddeley, 1982). That is, we propose that maternal affective expressions are not only more common in a young infant’s environment, but also more informative with respect to ensuing actions. Gibson (1969; Gibson & Rader, 1979) often refers to “the perceiver as performer.” In her view, humans, as a species, are inherently motivated to discover, explore, attend, extract information, and differentiate objects, all in order to learn about their environments. There are, however, goals and needs specific to each task or situation. For example, Gibson points out that a mountain climber is more attentive to where each step is placed than is someone who is taking a leisurely stroll (Gibson & Rader, 1979). Hence, perception is motivated by goals that are important to an individual. With respect to emotion, infants may have been more motivated to attend to the affective behaviors of their own mothers, because these may foreshadow more specific outcomes for them. That is, typically, maternal smiles are likely to be followed by positive caretaking interactions or experiences of being held; in contrast, maternal negative expressions may be frequently followed by episodes when the infant is left alone. Thus, it is possible that infants more readily recognized the affective signals of mothers because these entailed more idiosyncratic patterns of ensuing interactions.

The intermodal preference obtained in the current study for maternal expressions partially substantiates such a claim. We also found corroborating evidence from the data on infants’ affective reactions to emotional expressions. First, infants who watched their own mothers showed significantly more positive and less negative affect than did infants who watched the unfamiliar female. Furthermore, the finding that infants expressed particularly more positive affect when maternal happy facial expressions were sound specified suggests that they recognized their own mothers’ happy expressions and responded appropriately with more positive affect themselves. Although infants exhibited less positive affect when maternal sad facial expression were sound specified, we did not find increases in distress when maternal sad expressions were sound specified. We did find, however, that infants looked longer at happy facial expressions than at sad expressions. Perhaps, as has been documented in previous live interaction studies (see Haviland & Lelwica, 1987), infants responded to maternal sound specified sad expressions by looking away. Others have proposed that infants’ gaze aversion serves as a regulator of arousal (Field, 1981; Stern, 1974), and it is often discussed in the context of negative emotionality (Cohn & Tronick, 1983; Waters, Matas & Stroufe, 1975). For example, in a study of mother–infant interaction during which mothers simulated depression, Cohn and Tronick (1983) found that infants alternated between negative facial expressions, protest, and looking away. Presumably, infants avert their gaze from the stimulus to regulate arousal, inhibit crying, and preserve the motivation to reapproach the event (Rothbart, 1989). Moreover, several studies suggest that perhaps the ability to intersperse periods of attention with gaze aversions may be the most effective means of regulating arousal while remaining involved with a stimulus (Field, 1981; Stifter & Moyer, 1991).

Second, when infants watched their mothers, they exhibited a greater number of full and bright smiles. Other studies suggest that these smiles (Duchenne smiles) reflect genuine positive emotion, whereas non-Duchenne smiles are hypothesized to be social smiles (Camras et al., 1998; Dickson et al., 1997; Ekman et al., 1990). The finding that infants expressed a greater number of smiles of all intensities more frequently when maternal happy facial expressions were sound specified corroborates our proposition that infants discriminated between their own mothers’ happy and sad expressions and responded more appropriately to maternal expressive displays. In contrast, infants who watched the unfamiliar female did not show increases in the number of smiles across conditions. To further substantiate this proposal, future research will examine the differences in terms of infants’ emotional responsiveness toward their own mothers versus an unfamiliar female, using several other emotions, including anger, fear, and surprise.

In conclusion, the finding that younger infants can detect the relation between the vocal and facial emotional expressions of their own mothers earlier than they detect such information presented by unfamiliar others is consistent with previous findings from studies that examined infants’ recognition of emotion in the context of mother–infant interactions (Cohn & Tronick, 1988; Haviland & Lelwica, 1987; Papousek & Papousek, 1987; Stern, 1971, 1974; Toda & Fogel, 1993). Further, these findings point to an early sensitivity to affective expressions that are dynamic, multimodal,
and contextually relevant. Given that infants are frequently exposed to their caregivers’ emotional displays and further presented with opportunities to view the affordances (Gibson, 1959, 1979) of those emotional expressions, we propose that the expressions of familiar persons are meaningful to infants very early in life. This proposal should be more clearly established by continued investigations on the contingent patterns of infants’ responsiveness to others’ emotions that should include more microanalytic measures of infants’ facial expressiveness, their gaze aversion, and also physiological changes. Furthermore, the next step, derived from the current findings, should involve experimentation with additional emotional expressions and different pairings of emotions, as well as testing infants’ recognition of emotional expressions of other familiar caregivers (e.g., fathers), and across more heterogeneous backgrounds and ethnic identities.

ACKNOWLEDGMENTS

This article is based on a dissertation submitted by R. Kahana-Kalman in partial fulfillment of the requirements for the doctoral degree from the Department of Psychology at Rutgers University. Portions of this research were presented at the Biennial International Conference on Infant Studies, Atlanta, GA, in April 1998. This research was supported by a grant from the National Science Foundation (SBR-9408993). R. Kahana-Kalman thanks the members of her dissertation committee, M. Lewis, L. McCune, and C. Tamis-LeMonda, for their helpful comments and suggestions. The authors also thank the infants and mothers for their generous participation in the study; M. Cappozoli, D. Montague, M. Ryder, and C. Wiggins for their assistance; and Holly A. Ruff for her careful reading and insightful comments on an earlier version of this manuscript.

ADDRESSES AND AFFILIATIONS

Corresponding author: Ronit Kahana-Kalman, Albert Einstein College of Medicine/The Children’s Hospital at Montefiore, Department of Pediatrics, Kennedy Center 221, 1300 Morris Park Avenue, Bronx, NY 10461; e-mail: kalman@aecom.yu.edu. Arlene S. Walker-Andrews is at Rutgers University, New Brunswick, NJ.

APPENDIX

Facial Affect Codes and Reliability Scores

<table>
<thead>
<tr>
<th>Affect Code</th>
<th>Definition</th>
<th>Reliability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>κ</td>
<td>Agreement</td>
</tr>
<tr>
<td>Positive affect codes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Faint smile, grin</td>
<td>Small, closed-mouth smile in which there is some upturning but no pulling back of the corners of the mouth and hence no cheek puffing. Brows remain neutral.</td>
<td>.74</td>
<td>.32</td>
</tr>
<tr>
<td>2. Full smile</td>
<td>Full, open-mouth smile with lips pulled back and turned up, usually showing teeth in older infants. Eyes focused and attentive. Brows neutral.</td>
<td>.79</td>
<td>.24</td>
</tr>
<tr>
<td>3. Bright smile</td>
<td>Lips pulled back and turned up, mouth open, cheeks raised. Eyes narrow and sometimes crow’s feet crinkling at outer edges of eyes. Eyes convey bright and animated pleasure.</td>
<td>.90</td>
<td>.09</td>
</tr>
<tr>
<td>Negative affect codes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Wariness</td>
<td>Eyebrows drawn close together and slightly raised where they then meet at center of the forehead. No other signs of distress in the baby’s face.</td>
<td>.83</td>
<td>.13</td>
</tr>
<tr>
<td>5. Low-level distress (pouting)</td>
<td>Mild turning down of corners of the mouth and/or jutting lower lip. Mouth either open or closed. May be accompanied by brow furrowing described above.</td>
<td>.68</td>
<td>.05</td>
</tr>
<tr>
<td>6. Cry faces</td>
<td>Eyes narrow and compressed by brow-furrowing resulting in nondirected and squinted gaze. Mouth open with corners pulled downward; nasolabial folds apparent; eyebrows lowered and narrowed; nose wrinkled. Infant looks clearly distressed.</td>
<td>.70</td>
<td>.05</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>.77</td>
<td>.08</td>
</tr>
</tbody>
</table>
REFERENCES


