Beyond Surprise: The Puzzle of Infants’ Expressive Reactions to Expectancy Violation

Klaus R. Scherer, Marcel R. Zentner, and Daniel Stern
University of Geneva

The reactions of 58 infants to expectancy violation by digitally filtering the experimenter’s voice were studied in a cross-sectional design for ages 5, 7, 9, 11–12, and 14 months. The results show that behavioral freezing and changes in gaze direction, but not facial or vocal expression, are reliable responses to expectancy violation. The pattern suggests that a transition in the infant’s capacity for cognitive evaluation of novel and discrepant events may occur around age 9 months. These findings confirm the consistent failure to find prototypical facial surprise reactions in research on novel or impossible situations. Componential theories of emotion, which predict adaptive behavior patterns from appraisal processes, may provide clues for underlying mechanisms and generate hypotheses on age-related changes in emotional expression.

Psychologists often study affective reactions to situations or events by requesting the individual’s verbal report concerning the nature of the emotional experience. This request is not an option for preverbal infants, in which case psychologists studying emotional development need to resort to nonverbal indicators of emotion such as facial expressions or body movement. Much of the work in this area has been informed by the theoretical notion, first suggested by Tomkins (1962) and further developed by Ekman (1972, 1989) and Izard (1971, 1994), of “affect programs” for a limited number of basic emotions characterized by prototypical expression patterns. On the basis of this notion, developmental psychologists studying infant affect expected to find prototypical surprise reactions in response to novel or “impossible” situations in which the infants’ expectations are violated through clever manipulations (e.g., Baillargeon, 1993; Baillargeon, Spelke, & Wasserman, 1985; Cashon & Cohen, 2000; Hiatt, Campos, & Emde, 1979; Lewis, Alessandri, & Sullivan, 1990; Shapiro, Fagen, Prigot, Carroll, & Shalan, 1998). If one assumes that the novel situations created by these experimenters produced the basic emotion of surprise, affect program theories would predict prototypical surprise reactions in the infant, particularly in the form of facial expressions characterized by raised brows, widened eyes, and open mouth (Camras, 1992; Izard, 1971, 1994). Note that the expectation of prototypical surprise expression in the face as a consequence of an expectancy violation rests on the following two hypotheses: (a) Expectancy violation will produce a comparable reaction in all infants exposed to the manipulation, and this reaction will consist of a single basic emotion, in particular, surprise; and (b) this emotion will manifest itself in a prototypical pattern of facial expression.

Researchers have had a hard time providing empirical support for these predictions and demonstrating full-fledged prototypical surprise expressions in infants confronted with expectancy violations. Camras et al. (2002) reviewed this literature and concluded...
that “empirical attempts to identify surprise events via facial coding have met with limited success” (p. 180), and “no study to date has found that the majority of infants in a presumed surprise-inducing situation actually produce a surprise-related facial expression” (p. 188). In Camras et al.’s own empirical study with European American, Chinese, and Japanese infants, using a covert toy-switch procedure to violate expectancies, they also failed to find a significant increase in surprise-related facial expressions in an expectancy-violation condition as compared with a baseline condition. Instead, they reported an increase in gaze at the object and sizable amounts of “facial sobering” and “bodily stilling.”

Camras et al. (2002, pp. 180–181) provided several explanations for this lack of isomorphism between assumed emotion and facial expression such as Ekman’s (1998) suggestion that emotions may be “open programs” or proposals by Barrett and Campos (1987) and Camras (1992) of viewing emotions as contextualized functional systems in which facial expression production is determined by both emotional and nonemotional factors. Arguing that facial expression is not a gold standard for detecting the presence of a particular emotion, they recommended a multimethod approach to coding emotions that takes many different behavioral modalities into account.

Although investigating different response modalities is certainly preferable to single modality measurement, it does not provide a solution to the underlying theoretical problem: What behavioral cues are to be considered as bona fide markers of discrete basic emotions? In fact, a multimodality approach can actually render the issue even more complex. What if there are conflicting cues in different modalities? The issue of valid markers remains obscure as long as there are no theoretical predictions for concrete mechanisms that underlie the externalization of affect. Although multimodal response profiles might well become the gold standard for the existence of an underlying emotional state, more theoretical effort needs to be invested in developing falsifiable theoretical predictions on the response profiles that are considered signatures for specific basic emotions. In the absence of such predictions, it is difficult to envisage systematic research on this issue. It is, of course, likely, as Camras et al. (2002) assumed, that facial expression production is determined by both emotional and nonemotional factors. However, if the occurrence of a facial action, or another type of expressive behavior, cannot be used to infer reliably the presence of an underlying basic emotion, and if the nonoccurrence of such behaviors cannot serve as a reliable index of the absence of such an emotion, then predictions of emotion-specific expression profiles cannot be empirically tested.

An alternative theoretical framework to study infants’ reactions to novel situations is offered by appraisal approaches to emotion. The application of this line of theorizing to the study of emotional development has been less prominent than affect program theories of basic emotions, even though many of the pioneering contributions to the study of emotional development (e.g., Campos & Stenberg, 1981; Lewis, Sullivan, & Michalson, 1984; Sroufe, 1979) have highlighted the important role of the infant’s appraisal of objects and events as a central aspect of the emotion process (see also Campos, Barrett, Lamb, Goldsmith, & Stenberg, 1983, for an early summary of this literature).

Building on this line of theorizing, we propose an appraisal-based component process model (CPM) as a complementary theoretical approach to inferring the affective state of nonverbal infants from expressive behavior. Theorists in this tradition hold that the majority of emotion episodes are elicited by cognitive evaluation of the significance of an event for an individual, and that the nature of the ensuing emotional reaction is shaped by the result of the respective appraisal processes (see Scherer, Schorr, & Johnstone, 2001, for a comprehensive overview of the history, the different appraisal theories, the empirical evidence to date, and a critical assessment of this approach).

Rather than assuming the existence of a limited number of basic emotions characterized by emotion-specific patterns of expression, this theoretical approach suggests that expressions (and other response components) are direct, efferent products of the results of appraising a stimulus event for specific evaluation dimensions such as novelty, expectedness, goal conduciveness, and coping potential. Applied to the present context, this approach means that appraisal theorists would interpret the expressive behavior of an infant in response to an expectancy violation as a result of the infant’s evaluation of the event for some

---

1 It should be noted that the study by Camras et al. (2002) was actually conducted after the research project reported here. In their article, these authors cited extensively from an earlier version of the present article that was sent to one of the authors for commentary. Given changes in statistical analyses suggested by reviewers, some of the results they cited are now reported in a slightly different form. Because their article was published earlier in Emotion, we base some of our discussion on their data and conclusions.
of these criteria rather than a read-out of a specific basic emotion such as surprise. Appraisal theorists highlight individual differences in appraisal, thus accounting for the fact that the same event may provoke very different emotional reactions in different individuals. In the present context of expectancy-violation experiments, appraisal theorists would predict that different infants’ reactions to the same manipulation could be variable (Ellsworth, 1991; Roseman & Smith, 2001). Most important, the explanation of reaction patterns in motor expression and physiological responses differs greatly (Scherer, 1992, 2001; Smith, 1989). Whereas affect program theorists (Ekman, 1972, 1989; Izard, 1971, 1994) assume that basic emotions are characterized by prototypical expression patterns (which Tomkins, 1962, attributed to innate “neuro-motor programs”), appraisal theorists expect an emergent pattern based on the dynamic unfolding of appraisal results (see Wehrle, Kaiser, Schmidt, & Scherer, 2000, for a detailed discussion concerning facial expression).

In the context of his component process model of emotion, Scherer (2001) made detailed predictions about the effects of different stimulus evaluation checks (the appraisal dimensions in the CPM) on expressions in face, voice, and body as well as on neurophysiological reaction patterns. Specifically, for a “novel” outcome of the novelty check, which is directly pertinent to the novel or impossible situation paradigm, the following predictions on expression were made: (a) general effects: orienting response, interruption of ongoing activity; (b) face: Action Units (AUs) 1 and 2 (brows up), 5 (lids up), or 4, 7 (frown, scanning), 26 (jaw drop, open mouth), 38 (open nostrils), gaze directed; (c) voice: interruption of phonation, ingressive (fricative) sound with glottal stop; and (d) body: interruption of ongoing instrumental action, raising head, straightening posture (see Scherer, 2001, Table 5.3, p. 109).

In this study, we apply these predictions to the novel or impossible situation paradigm and examine whether we are able to find evidence for a novelty appraisal by the infants in accordance with the theoretical predictions by measuring gaze behavior, facial and vocal expression, and total body reactions. In addition, we were interested in the effects of infants’ developmental progression on the appraisal process, as inferred from expressive behavior, because appraisal theory predicts that the affective states produced by appraisal depend on the level and complexity of cognitive capacity (Scherer, 1979, 1984). We chose a cross-sectional design to examine the period between 5 and 14 months of age in 2-month increments. Given the importance of this age window for cognitive and affective development (Barrett & Campos, 1987; Case, 1991; Fischer, Shaver, & Carnochan, 1990; Sroufe, 1996), examining the existence of systematic shifts in expressive reactions to expectancy violations over this age range is of major interest. On the basis of neo-Piagetian theory (Mounoud, 1986), we chose the age levels of 5 months, 9 months, and 14 months to represent transition points in development and 7 as well as 11–12 months as plateaus. We were particularly interested in the pivotal role of the transition point at around 9 months, given the major neurological and psychological reorganizations that take place during this period (see reviews in Herschkowitz, Kagan, & Zilles, 1997; Mandler, 1998; Meltzoff, 2002), which is sometimes referred to as the “9-month revolution” (Rochat & Striano, 1999). Starting at this age, infants can match facial and vocal expressions to the portrayed underlying emotion, and they can categorize and generalize across varied productions of emotional expressions (Walker-Andrews, 1997).

These issues were analyzed in a cross-sectional study across five age groups of infants by examining responses to an expectancy violation in the form of a sudden change (brought about by digital filtering) in the voice quality of a female experimenter playing with the child. The assumption is that this paradigm produces both novelty (voice change) and expectancy discrepancy (see Scherer, 2001, p. 96) through the sudden appearance of a disjunction between the visual and auditory input into the infant’s schema for the person who, as the experimenter, has been established through prior interaction. The predicted effects on expression were measured through microcoding of facial AUs, vocalizations, and gaze direction as well as through judgments of freezing (interruption of ongoing activity).

Method

Experimental Violation of Expectancies in Infants of Different Ages

Participants. Fifty-eight infants (27 boys, 31 girls) completed (with missing observations in some cases) the experimental program as a whole. The
number and gender of infants studied in each age group were as follows: 12 5-month-olds (6 boys, 6 girls); 11 7-month-olds (5 boys, 6 girls); 12 9-month-olds (3 boys, 9 girls); 13 11- to 12-month-olds (8 boys, 5 girls); 10 14-month-olds (5 boys, 5 girls).

General procedure. Parents were recruited by placing advertisements in a major Geneva newspaper. On arrival at the laboratory, the experimenter and the accompanying caregiver played with the child in a room adjacent to the laboratory for approximately 20–30 min to familiarize the child with the experimenter and the laboratory environment. The child was then placed in a specially prepared chair in the laboratory, with the caregiver sitting behind the child. Two video cameras were positioned in such a way as to allow a split-screen take of both the experimenter’s face and a close-up shot of the upper body and head of the infant. To allow the manipulation of the infant’s perception of the experimenter’s voice quality in one of the violation situations, the experimenter’s voice was fed into a computer circuit and played back through two small loudspeakers that were inserted into the sides of the headrest of the experimental chair. It can be assumed that the infants accepted the voice coming out of the two loudspeakers in the headrest as the experimenter’s voice because they interacted naturally with the experimenter and did not attempt to investigate the loudspeakers. Because of the use of two loudspeakers, one for each ear, there was no change in the apparent direction of the voice when the infant shifted in the seat.

Several brief tasks were conducted during the experimental session. The voice-change expectancy-violation task was always conducted after two other, unrelated manipulations, one involving grasping behavior, the other involving visual tracking. Although the experimental method of presenting an “impossible,” or unexpected, event was used in all three tasks, each had a different purpose, and there was little theoretical interconnection. The order of the three tasks was not varied for practical and theoretical reasons. Given the age of the infants studied and the widely divergent nature of the tasks, it was felt that there was little danger of carryover effects, and that it would be preferable to keep the sequence of the tasks and, in general, the experimental context stable across infants.

Voice-filtering procedure. Following these two initial experimental situations (about 10–15 min into the session), the experimenter brought out a new toy (a big plastic snail that the infant had not seen before) and kept talking about it to the child, maintaining her gaze fixed on the infant’s face. By surreptitiously flipping a switch, the experimenter could activate a computerized digital filter built into the voice amplification circuit described earlier. The filter characteristics had been set in such a way as to augment the energy in the frequency range between 2 and 4 kHz of the voice spectrum. It is known from the literature that proportionally higher energy in this spectral range makes the voice sound sharp or metallic.

After approximately 12–14 s of talking with a normal voice (first baseline condition), the experimenter, while continuing to talk, activated the filter and talked for 8–10 s with the filter on (first filtered-voice condition). Then the experimenter switched the filter off and talked normally for 8–10 s. This sequence of baseline and filter-violation conditions was repeated twice, creating a total of three experimental periods, with the filtered-voice condition serving as a violation of the infant’s expectancies relative to the experimenter’s vocal characteristics. The first baseline condition was allowed to last somewhat longer to acquaint the child with the new toy and the specific style of interaction in this situation.

Coding of Expressive Behavior

Coding facial expression. Facial expressions were coded with the help of the Baby Facial Action Coding System (FACS) (Oster, 1999). Three female coders who had learned FACS and passed the reliability test, as administered by Ekman’s laboratory, were trained in the use of Baby FACS by Harriet Oster. Baby FACS, rather than the Maximally Discriminative Facial Movement Coding System (which is based on prototypical facial expression configurations; Izard, 1980), was chosen because the questions described earlier require the measurement of the activation of individual facial AUs to examine, for example, partial realizations of prototypic patterns or individual AUs as markers of a particular appraisal result.

Facial AUs, as specified in Baby FACS (Oster, 1999), were coded with the exact onset and offset times to allow duration measurements. However, in the analyses reported in the following paragraphs, only those AUs that had their onset in either the baseline or the violation conditions were used because the occurrence of AUs that have their offset in the condition, or that continue throughout, cannot be attributed to an effect of the manipulation.

Initial perusal of the data showed that many of the individual AUs coded with Baby FACS (Oster, 1999) had very low frequencies, prohibiting systematic sta-
tistical analysis. Therefore, the decision was made to regroup AUs likely to co-occur as functional configurations (see Wallbott & Ricci-Bitti, 1993). In particular, AUs 1, 2, and 5 (brow raising); 3 and 4 (brow lowering); 6 and 12 (smile); and 25, 26, and 27 (mouth open) were regrouped.

**Coding vocal expression.** The exact timing of onset and offset of all vocalizations was determined. No further coding was performed at that point.

**Coding gaze direction.** Gaze behavior was assessed with a simple coding system developed especially for the present study. On the basis of pilot analyses, four major categories of gaze direction were identified: (a) toward the object (toy), (b) toward the experimenter, (c) elsewhere in the room, or (d) toward the caregiver sitting behind the experimental chair. Coders then identified the onset and offset of each of these four types of gaze direction. Gaze at the caregiver, which required the baby to turn around, was only coded for control purposes. As expected, the latter occurred infrequently; therefore, it was not analyzed in detail.

**Computerized coding procedure.** The infants’ facial expression, vocalizations, and gazing behavior during each of the three pertinent periods of this situation were coded on the basis of the videotapes, with the help of a specially developed computer-based coding program that uses a vertical time code (written directly onto the videotape of each infant’s session during the initial recording).

**Reliability.** The reliability of the three coders was assessed by asking them to code independently a subset (10 infants from different age groups) of the complete video records and to compare the resulting time line records as produced by the coding program. The reliability of the assignment of categories within each coding system was assessed by computing percentage agreement, chance agreement, and Cohen’s kappa (based on Bakeman & Gottman, 1987). Vocalizations occurred too infrequently to compute reliability coefficients. For gaze, agreement was virtually perfect, with a percentage agreement of 92% and a kappa of .81. The reliability of the Baby FACS (Oster, 1999) coding (classification of relatively rare events distributed over 25 categories) had to be determined by percentage of agreement only, as the kappa coefficients could not be interpreted. As is well documented in the literature, kappa yields low coefficients despite high actual agreement in cases in which there are many categories with low base rates and an uneven distribution over categories (Feinstein & Cicchetti, 1990; Landis & Koch, 1977; Uebersax, 1987). The percentage of agreement of coefficients for the assignment of categories varied somewhat over AUs, with a minimum of 44% and a maximum of 98%. The median percentage agreement across 25 coded AUs was 88%, the mean was 83%, and the standard deviation was 15%. The lower agreement coefficients were mostly the result of confusions between similar facial movements, in particular AUs 3 versus 4 and AUs 25 versus 26 versus 27. Once these individual AUs were combined in functional groups, as described earlier, percentage of agreement reached or exceeded .80. Because neither onset-offset nor duration information was used in the data analysis, the reliability of the placing of onsets or offsets on the time line (which is a difficult task because of the large number of zeros) was not determined.

**Measurement of Freezing Behavior**

**General approach.** A review of the literature at the time of the analysis did not yield any information on the existence of a reliable coding scheme to measure frequency and duration of behavioral freezing in infants. As several attempts were made to use the definition of freezing, described earlier, for the develop-
ment of microcoding techniques similar to FACS coding by using trained coders to assess the incidence and duration of freezing bouts. However, it became apparent that going back and forth on a tape to determine the exact video frame locations for onset and offset of behavioral freezing breaks the dynamic continuity of behavior and makes it more difficult to observe freezing. Given that the onset and offset of freezing bouts can best be detected on the background of viewing dynamic, continuously changing behavior, it was decided to use an approach in which freezing bouts were to be measured online and in real time. Pilot studies with trained and untrained coders showed that there was no difficulty in obtaining reliable identification for the occurrence of freezing bouts in real-time observation. However, given the nature of the phenomenon, there was less agreement about the exact onset and offset of these freezing periods, even for trained coders. Because a relatively precise measure of freezing bout onset and offset was preferred (which was needed to measure duration and to assess the synchronization of behaviors across different modalities), the decision was made to use a judgment approach on the basis of observer agreement and define beginning and ending of a bout probabilistically from a cutoff point constituted by a minimal number of judges.

**Judges.** Twenty-four judges (22 female and 2 male students; age range = 20–24 years) were recruited from the participants of a first-year psychology course. They were paid for their services.

**Stimuli.** Video excerpts of all baseline and violation periods (about 1 min each in duration) for 50 infants for whom all three experimental periods had been recorded were copied, in random order, onto three different judgment tapes. The respective segments were separated on the judgment tapes by empty blue screens of 5-s duration.

**Judgment procedure.** Judgment sessions were conducted in six groups, one group with five judges, four groups with four judges, and one group with three judges. Given the large number of stimuli and the degree of concentration needed, each judge group saw only approximately two thirds of all stimuli. To avoid confounding of judge group with stimulus tape, each group was asked to judge two different stimulus tapes in an overlapping design in such a way that each stimulus was judged by 16 judges.

After arrival at the laboratory, judges were seated in a half circle around the monitor, separated by curtains. Written instructions provided a detailed description of the behavioral aspects constituting the freezing response: “Interruption of exploratory behavior and/or decrease in general activity level and/or fixity of gaze and/or postural rigidity and/or still face.” Judges were asked to indicate the presence of a freezing bout when they had the subjective impression of the infant freezing rather than to try to ascertain the presence of all criteria listed (which would have been impossible in real time). They were not given any information about the experimental conditions. Only the video signal was shown, with the sound turned off; thus, the nature of the condition could not be inferred from the excerpts. Judges were asked to push a button on a remote monitoring unit whenever the infant to be coded showed an onset of a freezing state, as defined in the instructions, and to release the button at the offset of the freezing state. The button-press signals were stored as multiple time lines on a central computer. The sessions lasted about 1 hr, with a break after approximately 30 min.

**Operational definition of freezing bouts.** Figure 1 shows a representative sample excerpt of the raw data plots for the freezing judgments. The three major peaks in the sample represent states of the infant that were identified as freezing, with very high agreement between judges (generally, well over two thirds of the judges pressed the button during this period). As Figure 1 shows, in addition to these clearly defined peaks, there are periods in which only a subset of the judges pressed the button. This pattern represents the well-known “pedestal” phenomenon in online signal processing, which requires cleaning of the signal before further analysis. A separate reliability analysis was used for each infant as the basis to remove the pedestals from the signal. The on–off states (0–1) across all time points for all pairs of judges for a particular infant were intercorrelated, and judges who had a mean correlation of < .05 with all other judges were excluded for the respective infant. This operation concerned 10 of the 24 judges. The maximum number of judges per infant was 15; the minimum was 11 ($M = 14.04, SD = 0.84$). The precise breakdown is as follows: For 25 infants, 1 of the 16 judges was excluded; for 5 infants, 2 of the 16 judges were excluded; for 16 infants, 3 of the 16 judges were excluded; for 3 infants, 4 of the 16 judges were excluded; and for 1 infant, 5 of the 16 judges were excluded. The fact that none of the judges showed

---

3 There had been several incomplete sessions in the 5-month-old age group. In order to keep the stimuli for freezing judgments comparable, freezing data were not obtained for the respective infants.
consistently lower agreement levels for all infants suggests that there were no qualitative differences between judges with respect to ability or conscientiousness. The reasons for lower agreement are likely to be variable and may depend on the general movement level of an infant.

After judges with low agreement coefficients had been excluded, the intercorrelations and effective reliabilities for the remaining judgments were recomputed, again separately for each infant. The variables used in the analyses were computed from the respective means. Over all 50 infants, the following central tendencies for the average intercorrelations across judges were found: mean $r = .31$, median $r = .34$, $SD = .11$. On the basis of the median intercorrelation of $r = .34$, the effective reliability for 4–5 judges is approximately $R = .70$ (see Rosenthal, 1987, Table 2.1, p. 11), suggesting that the one-third criterion leads to a very reliable estimate of the on–off time points. In any case, a potential estimation bias would not be of any incidence for the present study because freezing duration is not interpreted parametrically, but only with respect to control-violation differences. Because both conditions would be affected by a bias, it would cancel out in the statistical analyses.

Variables. The raw frequency of freezing bouts was divided by the length of the respective condition, given the slight variations in duration over infants and conditions. Thus, this variable can be interpreted as the likelihood of a freezing bout occurring within a standard time frame. Similarly, the average duration of bouts was divided by condition duration in order to allow the interpretation of this variable with respect to a standard time frame.

Results

Facial Expression

We used occurrence of AUs in a condition rather than their frequency or duration for the analyses to be
reported here because we were interested in the nature of the reaction to the experimental manipulation rather than in the detailed temporal patterning of the expressions. Because preliminary analyses had shown that there were only a limited number of AUs and that restricted variance would not allow the use of parametric statistical techniques, we computed variables for four AU combinations (from theoretical considerations concerning functionally important configurations; see Camras, Lambrecht, & Michel, 1996; Camras et al., 2002): brow raising (AUs 1, 2, and 5), brow lowering (AUs 3 and 4), smiling (AUs 6 and 12), and mouth opening (AUs 25, 26, and 27). These combinations were based on the presence of one or two of the respective individual AUs. Over all age groups combined, the following percentages of occurrence were observed for the four AU combinations (baseline/violation): brow raising (22.4%/31.0%), brow lowering (6.9%/5.2%), smiling (8.6%/12.1%), and mouth opening (17.2%/8.6%). Both brow raising and mouth opening combined (the predicted prototypical pattern for the basic emotion of surprise) were observed in 1.7% (baseline) versus 3.4% (violation) of all cases. Tests of the differences between two correlated proportions (percentage of AU combinations shown) did not show any significant differences between baseline and violation conditions. There were too few facial expressions to systematically analyze age trends.

Vocal Expression

Infants showed little vocal activity, with a few exceptions, and the experimental manipulation did not systematically produce vocalizations.

Gaze Direction

Gaze variables are expressed as a percentage of the total duration of a specific condition during which the infant gazed at the experimenter, the object (the toy snail), the caregiver, or the ambient environment. On average, and across all conditions and age groups, infants directed their gaze to the experimenter 27.4% of the time, to the object 52.0% of the time, to the caregiver 2.5% of the time, and to the ambient environment 18.1% of the time. Given that the violation emanates from the experimenter, we focused on the variable gaze direction to the experimenter.

We performed detailed data analyses for Experimental Period 1 only because a preliminary analysis of variance (ANOVA) had shown a significant drop-off in gaze direction changes following the manipulation, indicating habituation. A 2 × 5 × 2 repeated measures ANOVA for the raw percentage scores for experimenter-directed gaze, with condition (2) as a within-subjects factor and age group (5) and sex as between-subjects factors, yielded a significant main effect for condition or change, $F(1, 48) = 18.818, p < .01, \eta^2 = .282$, and significant Sex × Change, $F(1, 48) = 9.262, p < .01, \eta^2 = .162$ (male: baseline $M = 21.5$, $SD = 19.4$, violation $M = 30.3$, $SD = 24.6$; female: baseline $M = 13.0$, $SD = 12.0$, violation $M = 40.1$, $SD = 34.1$) and Age Group × Change, $F(4, 48) = 2.792, p < .05, \eta^2 = .189$, interaction effects. Given the small N per subgroup, we did not further interpret the interaction effects involving sex. The Age Group × Change interaction, illustrated in Figure 2, can best be explained by comparing the 11- to 12-month-olds with the remaining age groups, $F(1, 48) = 6.035, p < .05, \eta^2 = .112$ (Weights −1, −1, −1, 4, and −1). However, as can be easily seen in Figure 2, the amount by which gaze toward the experimenter increases from baseline to violation is very similar for 11- to 12- and 14-month-olds. Testing for a contrast (Weights −2, −2, −2, 3, and 3) between the three younger groups and the two older groups in an ANOVA of the change scores for gaze toward the experimenter (violation–baseline) did indeed yield a significant result, $F(1, 48) = 6.035, p < .05, \eta^2 = .112$, suggesting that an important transition occurs after 9 months and before 11 months of age.

In order to check what gaze categories decreased as a consequence of the increase in gaze toward the experimenter, we computed the same types of ANOVAs for gaze toward the toy snail and toward other direc-

Figure 2. Mean proportion (Prop.) of gaze toward the experimenter by age groups for the baseline (solid bars) and violation (gray bars) conditions. mo. = month.
We conducted a 2 (condition) × 5 (age groups) × 2 (sex) repeated measures ANOVA, with a planned contrast estimate. A fourth-order effect represents a \( \wedge \)-shape.

**Associations Between Measures of Expressive Behavior**

We examined the possibility that infants showing reactions to the experimental manipulation in one expressive domain showed corresponding changes in another domain by using various types of association measures. None of the respective coefficients came close to significance.

**Discussion**

The experimental manipulation did not produce a systematic increase in surprise-specific facial expression patterns (brow raising and mouth opening), nor of other classes of facial or vocal expressions following expectancy violation. These results are consistent with previous findings, including those obtained by Camras et al. (2002, Table 1). On the whole, there is no evidence for a prototypical surprise expression in

For gaze toward the toy snail, we found a significant main effect for condition or change, \( F(1, 48) = 11.642, p < .01, \eta^2 = .19 \), consisting of a general decrease following violation, and significant Sex × Change interaction effects, \( F(1, 48) = 5.053, p < .05, \eta^2 = .09 \) (male: baseline \( M = 54.4, SD = 4.5 \), violation \( M = 49.6, SD = 6.6 \); female: baseline \( M = 69.7, SD = 3.9 \), violation \( M = 46.5, SD = 5.7 \)). Testing for a contrast (Weights \( -2, -2, -2, 3, \) and \( 3 \)) between the three younger groups and the two older groups yielded a trend, \( F(1, 48) = 3.504, p < .10, \eta^2 = .07 \), showing that the two older groups decreased gaze toward the toy snail more than the three younger groups did. This is a mirror image of the pattern of increase in gaze toward the experimenter. None of the effects for gaze toward other directions reached significance. Thus, the increase in gaze toward the experimenter is accounted for by a decrease in gaze toward the toy.

**Freezing**

As in the case of gaze, we performed detailed data analyses for Experimental Period 1 only because a preliminary ANOVA had shown a significant drop-off in freezing following the manipulation, indicating habituation. We conducted a \( 2 \times 5 \times 2 \) repeated measures ANOVA for the time-standardized scores for frequency of freezing bouts, with condition or change (2) in Experimental Period 1 only (with age group and sex as between-subjects factors). The results showed a significant main effect for condition or change, \( F(1, 40) = 9.858, p < .01, \eta^2 = .198 \), indicating a general increase from baseline (\( M = .24, SD = .22 \)) to violation (\( M = .47, SD = .44 \)) in the proportion of freezing bouts occurring in a standard time frame. None of the other effects reached significance.

The same analysis for the average duration of individual freezing bouts during a condition also yielded a significant main effect for condition or change, \( F(1, 40) = 15.204, p < .01, \eta^2 = .275 \), indicating an increase in the relative duration of freezing from baseline (\( M = .11, SD = .14 \)) to violation (\( M = .28, SD = .30 \)).

Although the overall Age Group × Condition interaction effect for freezing duration was not significant, the data suggests that the 9-month-old group showed less increase than the other groups in the duration of freezing because of the manipulation (see Figure 3). In order to test this effect, which corresponds to the expected pivotal status for the 9-month age group, we conducted a \( 2 \) (condition) \( \times 5 \) (age groups) \( \times 2 \) (sex) repeated measures ANOVA, with a planned contrast for age groups (Weights 1, 1, −4, 1, and 1). The analysis yielded a significant main effect for condition or change, \( F(1, 40) = 15.204, p < .01, \eta^2 = .275 \), and a trend for the contrast, \( F(1, 40) = 2.966, p < .10, \eta^2 = .069 \). As a check, we also examined the polynomial contrasts. As one would expect from the pattern shown in Figure 3 and the results of the planned comparison, we obtained a trend for a fourth-order polynomial effect, \( F(4, 40) = 1.173, p < .10 \) (Order 4 contrast estimate). A fourth-order effect represents a \( \wedge \)-shape.

**Figures**

**Figure 3.** Mean relative duration of freezing bouts by age groups for the baseline (solid bars) and violation (gray bars) conditions. mo. = month.
reaction to novelty, as would be predicted by a strong version of an affect program theory. In both Camras et al.’s results and ours, about one third of the infants showed facial actions predicted as cues to novelty appraisal (brows up and mouth open). However, there is no difference between the baseline and violation conditions in either study. It is possible, of course, that the infants showing these facial actions appraised stimuli and events in both the baseline and violation conditions as novel. In our study, the experimenter produced a new toy when starting the session, and the period of play before the experimental voice change may have been too brief to produce sufficient habituation, leading to novelty expressions even in the baseline condition (especially for younger infants). Future research needs to control for this possibility, for example, by extending the baseline situation individually until there are no more novelty reactions.

Voice

Few of the infants produced vocalizations. This may mean that the threshold for the production of vocal reactions to particular appraisal results, as observed in some cases, may be high. In addition, more fine-grained analyses, including the measurement of respiration patterns, may be required.

Gaze

Both the present study and the study by Camras et al. (2002) found a sizable increase in gaze directed at the object of expectancy violation as compared with baseline. To our knowledge, affect program theories do not make any concrete predictions with respect to gaze direction as a concomitant of specific fundamental emotions. In contrast, appraisal models, adopting a functionalist approach (identifying what types of behaviors have adaptive value in the service of obtaining more extensive information to allow realistic appraisal or to prepare coping responses), predict a change of gaze direction toward novel, unexpected stimuli (see Scherer, 2001, pp. 109–112). The results in infant studies of expectancy violation that measured gaze direction confirm this prediction. In our case, the pertinent stimulus is the experimenter and, we find, as predicted, a strong main effect showing an increase of gaze directed at the experimenter after violation for all age groups.

In addition, the present findings suggest that there is a significant increase in this tendency after 9 months of age. One possible explanation is that after this important transition, infants acquire the ability to use gaze direction strategically to appraise the causal- ity of events. This may coincide with a developmental change between 9 and 15 months in infants’ attentiveness to and understanding of others’ goals and intentions, for which there is now converging evidence from a number of studies (Meltzoff, 2002, p. 18; Tomasello, 1995). This result is of particular importance in understanding the close link between cognitive and emotional development in the sense of the acquisition of the cognitive capacities that are required for complex appraisal.

An alternative explanation for the increase in gaze direction toward the experimenter after 9 months of age is social referencing, a competence that also emerges during this developmental window (see Campos & Stenberg, 1981; Klinnert, Emde, Butterfield, & Campos, 1986). In the present study, it is difficult to evaluate the plausibility of the two alternative explanations because the experimenter is at the same time the object of the violation and the only easily available social reference person (the mother being placed behind the infant’s seat). In the study by Camras et al. (2002), the object (uncovered toy) and the social reference person (the mother) were separate. Adopting a social referencing explanation of changes in gaze direction, one would have expected the 11-month-old infants in that study to increase their gaze direction to the mother when the expectancy violation occurred. However, as shown in Table 1 in Camras et al. (2002), this was not the case. Rather, the duration of the gaze directed at the object increased. Thus, it seems more parsimonious to explain both sets of findings as an attempt to determine the causal mechanisms underlying the unexpected event.

Freezing

CPM predicts interruption of ongoing activity and of instrumental behavior as a consequence of a novelty appraisal. This is borne out by the results for behavioral arrest or freezing, which shows a strong main effect for violation across all ages. Camras et al. (2002) found similar results for their variables, facial sobering and bodily stilling.

To our knowledge, affect program theories do not make formal predictions about the role of freezing in basic emotions (although P. Ekman, personal communication, November 22, 2003, sees it as a sign of fear). CPM predicts freezing as a response to novelty because of its function as an interruption—the organism has to reorient to a potentially changed situation. In consequence, we propose that freezing be considered
a marker of an appraisal process that has not yet yielded a concrete motor action tendency because of an incapacity to fully evaluate an uncertain or ambiguous situation. The inability to appraise the situation adequately and to select an appropriate, adaptive action tendency can be expected to yield the evolutionarily prepared, nonspecific response to uncertainty, which is freezing. This may include situations of potential danger (or situations interpreted by infants as threatening), which is why freezing may sometimes occur as part of a fear response. In other words, freezing might be considered as expressing a state in which the organism cannot decide on, or does not have available, an appropriate behavioral reaction to deal adequately—at least with some probability of success—with a potentially threatening situation. Consequently, in situations in which appropriate appraisal is impossible or inconclusive (because of the complexity of the situation or cognitive processing limitations), freezing seems an appropriate fall-back response that allows the organism to await further information, or hope for the danger situation to clear, or to discover that there was a false alarm. One can assume that in many situations, very young infants do not yet have the necessary cognitive mechanisms (nor the stored experiences) to conclusively appraise highly unusual events and to prepare appropriate action tendencies. This explanation does not exclude the possibility that an orienting response is involved (indeed, an orienting response is predicted by CPM as a consequence of a novelty appraisal; Scherer, 2001, p. 109). What is important is that the efference that usually follows appraisal is suspended for the duration of freezing, and no motor action tendencies are generated.

As to developmental changes, we find a tendency for a \( \sqrt{ } \)-shaped relationship, with the 9-month age group at the low point. This is consistent with an assumption generally found in the literature and with our theoretical expectation about major transition points in development (see the introduction section for the discussion of the choice of age levels), namely, that the period around 9 months may be pivotal in the capacity to evaluate the event and/or to respond with appropriate action. The result is statistically not sufficiently powerful to attempt a more detailed interpretation. One possibility is that, because of the major cognitive restructuring taking place during that age period, more ambiguity or discrepancy is tolerated than during more stable age periods. However, replication of the finding and more detailed evidence will be required to evaluate the underlying mechanisms.

## Conclusion

Two issues seem worth considering in conclusion. First, the results of the present study confirm the general pattern of findings in this research area. Novel, impossible, or unexpected situations or events do not consistently elicit the prototypical facial expression for surprise in infants as originally predicted by affect program theories. A shift from a strong version of these theories to a weak one has been suggested, invoking "open programs" and multiple determination of expressive behavior. We have raised the problem of indeterminacy with this solution and suggest using a different theoretical approach, based on appraisal theory. This has two advantages.

1. If the detailed predictions can be borne out in future research, expressive behavior might allow one to infer the operation of specific cognitive processes, which would not only be extremely useful for understanding the operation of appraisal processes in general but also would contribute to the study of the role of cognitive development in interaction with emotional development. For example, if the cognitive mechanism required for certain types of appraisals is not yet available, the individual should be unable to experience the emotions for which the evaluation of particular criteria is a necessary condition (Scherer, 1979, 1984; see also Case, 1991; Fischer et al., 1990; Sroufe, 1996, for similar proposals based on a competence approach). The detailed study of how the acquisition of well-documented developmental milestones benefits certain types of appraisal could provide important insights into the underlying mechanisms. This is particularly true for the study of the age window around 9 months, the importance of which is again underlined by the results of the present study. Because many different transitions take place during this period, an important task for further research is the unpacking of the developmental milestones that are essential for different types of appraisal. Thus, one might profitably distinguish between the appraisal of novelty (fairly automatic detection of a discontinuity in stimulus features or sudden onsets) and discrepancy (comparison of elements of a schema or corresponding schemata) of a stimulus event, requiring different types of cognitive competencies.

2. The affect program approach is centered on a small number of basic emotions (while allowing for blends and family relationships). Appraisal theories, and the CPM in particular, are much more open with respect to the variability and fuzziness of different types of emotions, both for emotion-specific appraisal.
profiles and for reaction patterns. This openness allows a more differentiated approach (which has its cost in terms of parsimony) and provides room for discovery (as compared with confirmation of the existence of a limited number of relatively fixed emotion programs). This may be particularly valuable for developmental approaches because it is possible that emotions develop from more rudimentary to more complex processes. It also allows one to identify members of an emotion family that may be distinguished by important differences. Thus, on the basis of our results, we propose to distinguish between surprise and stupefaction. We suggest that stupefaction, characterized mostly by freezing, is produced by a situation in which the individual is stunned or dazed by an event that is completely outside of any established or imaginable set of expectations and in which there are no schemata available for appraisal. In contrast, we suggest use of the term surprise, primarily expressed by facial responses such as brow raising, for cases in which the individual appraises the discrepancy of a stimulus or event on the basis of an established set of expectations.

Although research on emotional development has often been focused on facial expression, relatively neglecting other expressive and behavioral modalities, the results of the present study suggest a special role for freezing. We have suggested that freezing be viewed as a holding operation in the process of appraisal, keeping action suspended, as long as novel or discrepant events are analyzed or further information is sought. Taken together, the present data and reflections suggest that adoption of componential models may hold considerable promise for stimulating and advancing future research on cognition–emotion interaction in infant development.

References


Received July 11, 2002
Revision received April 30, 2004
Accepted April 30, 2004

---

**Call for Nominations**

The Publications and Communications (P&C) Board has opened nominations for the editorships of *Clinician’s Research Digest, Emotion, JEP: Learning, Memory, and Cognition, Professional Psychology: Research and Practice*, and *Psychology, Public Policy, and Law* for the years 2007–2012. Elizabeth M. Altmaier, PhD; Richard J. Davidson, PhD, and Klaus R. Scherer, PhD; Thomas O. Nelson, PhD; Mary Beth Kenkel, PhD; and Jane Goodman-Delahunty, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2006 to prepare for issues published in 2007. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations also are encouraged.

Search chairs have been appointed as follows:

- **Clinician’s Research Digest**: William C. Howell, PhD
- **Emotion**: David C. Funder, PhD
- **JEP: Learning, Memory, and Cognition**: Linda P. Spear, PhD, and Peter Ornstein, PhD
- **Professional Psychology**: Susan H. McDaniel, PhD, and J. Gilbert Benedict, PhD
- **Psychology, Public Policy, and Law**: Mark Appelbaum, PhD, and Gary R. VandenBos, PhD

Candidates should be nominated by accessing APA’s EditorQuest site on the Web. Using your Web browser, go to [http://editorquest.apa.org](http://editorquest.apa.org). On the Home menu on the left, find Guests. Next, click on the link “Submit a Nomination,” enter your nominee’s information, and click “Submit.”

Prepared statements of one page or less in support of a nominee can also be submitted by e-mail to Karen Sellman, P&C Board Search Liaison, at ksellman@apa.org.

The deadline for accepting nominations is **December 10, 2004**, when reviews will begin.