Journal of Cognition and Development

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hjcd20

Fifteen-Month-Old Infants Match Vocal Cues to Intentional Actions

Elena Hoicka [a] & Su-hua Wang [b]

[a] University of California, Santa Cruz, and University of Stirling, United Kingdom

[b] University of California, Santa Cruz

Version of record first published: 08 Aug 2011.

To cite this article: Elena Hoicka & Su-hua Wang (2011): Fifteen-Month-Old Infants Match Vocal Cues to Intentional Actions, Journal of Cognition and Development, 12:3, 299-314

To link to this article: http://dx.doi.org/10.1080/15248372.2010.542215

JOURNAL OF COGNITION AND DEVELOPMENT, 12(3):299–314
Psychology Press
Taylor & Francis Group

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2010.542215

Fifteen-Month-Old Infants Match Vocal Cues to Intentional Actions

Elena Hoicka
University of California, Santa Cruz, and University of Stirling, United Kingdom

Su-hua Wang
University of California, Santa Cruz

Fifteen-month-old infants detected a violation when an actor performed an action that did not match her preceding vocal cue: The infants looked reliably longer when the actor expressed a humorous vocal cue followed by a sweet action or expressed a sweet vocal cue followed by a humorous action, than when the vocal cue was followed by a matching action. The infants failed to detect the mismatch when one person expressed the vocal cue and another performed the action. The results suggest that by 15 months of age, infants are capable of distinguishing between two types of vocal cues and actions along the positive emotional spectrum: humor and sweetness. Furthermore, they match humorous vocal cues to humorous actions and sweet vocal cues to sweet actions only when the cues and actions are made by the same person.

Affective cues in speech are a salient source of information for young children in the context of social interaction. Past research has demonstrated early use of affective cues in children's communication and in their understanding of intentional actions. For example, at as young as 6 months, infants already differentiate among positive, neutral, and negative speech

Correspondence should be sent to Elena Hoicka, Psychology, School of Natural Sciences, University of Stirling, Stirling, FK9 4LA, UK, or Su-hua Wang, Department of Psychology, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA. E-mail: elena.hoicka@stir.ac.uk; suhua@ucsc.edu

(e.g., Singh, Morgan, & Best, 2002). After their first birthdays, infants take into account affective cues to predict or interpret others' goal-directed actions. For example, 14-month-olds expect a person to grab the object to which she previously displayed positive emotion, but not the object to which she displayed either no emotion (Phillips, Wellman, & Spelke, 2002) or negative emotion (Barna & Legerstee, 2005). Slightly older, 18-month-old infants begin to use affective cues to infer others' desires: They attend to whether an adult displays positive or negative emotion and use this information to infer which food he or she desires (Repacholi & Gopnik, 1997).

Additionally, young children appear to use affective cues that are coupled with linguistic expressions to distinguish between different intentions underlying similar actions. For example, 14-month-olds were more likely to copy actions that were followed by the expression, "There," than by the expression, "Whoops" (Carpenter, Akhtar, & Tomasello, 1998), and 2-year-olds imitated wrong actions cued by laughter but corrected the same actions cued by "Whoops" (Hoicka & Gattis, 2008). Even children younger than 1 year old seem able to differentiate intentions underlying similar actions based on complex affective cues. For example, 9-month-olds were more persistent in asking for a toy (e.g., they reached more often and looked away less often) when an experimenter was unwilling to give the toy (marked by smiling) than when the experimenter was unable to do so (marked by negative or neutral expression; Behne, Carpenter, Call, & Tomasello, 2005).

The above findings could be taken as evidence that young children interpret the intentions underlying vocal cues given by an actor and that they react differently based on their interpretation. However, the goal-directed actions used in the aforementioned research could all be differentiated by affective valence alone. Positive (e.g., saying, "There," laughing, smiling) versus nonpositive emotions (e.g., saying, "Whoops," negative or neutral affect) were provided to children to distinguish intentional from unintentional actions, jokes from mistakes, and unwillingness from inability (e.g., Behne et al., 2005; Carpenter et al., 1998; Hoicka & Gattis, 2008). Thus, it may be the case that children rely on affective valence alone, and not their interpretation of the actor's intention, to differentiate different intentions underlying similar actions. One way to tap into this issue is to compare two types of vocal cues expressed along the same emotional spectrum.

The present research focuses on infants' understanding of two *positive* vocal cues: humorous and sweet cues (Hoicka & Gattis, 2011). Throughout this article, we use "humorous" to refer to the attribute that provokes positive feelings derived from unexpected actions, and "sweet" to refer to the attribute that provokes nice, warm-hearted feelings derived from caring or loving canonical actions. Specifically, we examined whether infants differentiated between humorous and sweet vocal cues, and whether they

matched humorous vocal cues to humorous actions and sweet vocal cues to sweet actions, despite that humor and sweetness cannot be differentiated by affective valence alone.

In naturalistic settings, parents readily distinguish humorous from sweet cues for their children. When mothers read books to their 18- to 24-month-olds, they used distinctive acoustic profiles to differentiate humorous from sweet speech (Hoicka & Gattis, 2011). Specifically, mothers spoke with a higher pitch and more loudly when telling jokes than when uttering sweet sentences. These features attract attention and help infants understand what has been said (e.g., Fernald, 1989; Newman, 2003). Furthermore, when expressing jokes, mothers increased their pitch throughout the utterance, which resulted in a significantly different contour than when parents expressed sweetness. This rising contour suggests that an utterance is open to interpretation (Gussenhoven, 2004) or signals contradiction (e.g., Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985), which could indicate to infants that jokes should not be taken seriously. In contrast, sweet utterances followed no specific contour. Empirically, infants have been shown to be sensitive to sweet or "loving" voices. For example, Trainor, Clark, Huntley, and Adams (1997) found that when mothers sang to their infants, they used a higher pitch and greater intensity variation than when they sang alone. The infant-directed songs were rated as more loving (i.e., sweet) by adults, and 4- to 7-month-olds' song preference was correlated to adult ratings of lovingness (Trainor, 1996).

Additionally, we chose humorous actions because they contrast starkly to typical goal-directed actions. Humor is defined as something incongruous, unexpected, or out of the ordinary (e.g., McGhee, 1979; Shultz, 1976). In conducting humorous actions, actors typically couple incorrect actions with positive affect (Hoicka & Gattis, 2008). The fact that both humor and sweetness are expressed with positive affect allowed us to disentangle affective valence from intention.

In the present experiment, we used the violation-of-expectation method to examine whether 15-month-old infants differentiated between humor and sweetness, and whether they matched humorous and sweet vocal cues to the corresponding goal-directed actions. An experimenter first gave a vocal cue. The cue was either laughter followed by a sentence spoken in a humorous way (see "Events"), or the expression, "Awww," followed by the same sentence spoken in a sweet manner. Next, the same experimenter performed either a humorous or a sweet action (same-agent condition). The choice of age was based on the findings that 15-month-olds were sensitive to actions that violated a pretense scheme (Onishi, Baillargeon, & Leslie, 2007), a concept similar to humor (Hoicka, Jutsum, & Gattis, 2008). Additionally, 10-month-olds laughed at incongruous actions performed by their mothers such as putting socks in their mouths (Sroufe & Wunsch, 1972), whereas 19- to 24-month-olds

not only appreciated but also reproduced humorous actions (Hoicka & Gattis, 2008). The age of 15 months represents the midpoint of the above two age groups that have demonstrated comprehension or production of humor.

We reasoned that if the 15-month-olds in our experiment correctly linked the vocal cues to the actions, they should detect the mismatch and look significantly longer when the humorous or sweet scheme was violated than when it was not. One possible interpretation for such results could be that the infants detected the experimenter's intention expressed by the vocal cues (i.e., intent to be funny or sweet) and expected her immediate actions to be consistent with it. Recent research showed that 13-month-olds used cues such as grasping and mild positive emoting to infer the goal-directed action of the same but not a different actor (Buresh & Woodward, 2007). Based on this finding, when the vocal cues and actions are provided by different agents, infants should not expect the vocal cues to match with the actions—after all, one person's intention may or may not be shared by another person. However, another interpretation could be that infants' understanding of humor and sweetness guided them to perform cross-modal matching. That is, infants might have linked the auditory cue to the visual event that belonged to the same category (humor or sweetness) without considering the intention that bridged the two. If this is the case, infants should still detect a mismatch between the vocal cues and subsequent actions performed by different agents.

To examine these two possible interpretations, a separate group of 15-month-olds heard an experimenter voice the vocal cues and then watched another experimenter perform the actions (different-agent condition). If the infants matched vocal cues to the corresponding actions without considering who gave the vocal cues, they should look reliably longer at the mismatch than at the match event, just like those in the same-agent condition. However, if the infants considered the intention underlying the vocal cues and used it to predict subsequent actions, they should do so only within the same person. As a result, those in the different-agent condition should not be intrigued by the mismatch between vocal cues and subsequent actions and should thus look about equally at the mismatch and match events.

METHOD

Participants

Thirty-two healthy full-term 15-month-olds participated ($M = 15$ months, 13 days; range = 14 months, 25 days to 16 months, 9 days; 18 boys). An additional 12 infants were eliminated due to fussiness ($n = 1$, same agent) or distraction/inattentiveness (e.g., due to an external noise, looking at

feet/hands/the ceiling light, babbling/talking; $n = 4$, same agent; $n = 3$, different agent), or because they looked at both events for the maximum amount of time allowed ($n = 4$, different agent). Sixteen infants were assigned to the same-agent condition ($M = 15$ months, 12 days), and 16 to the different-agent condition ($M = 15$ months, 14 days). Participants were recruited through birth announcements or parent–infant groups and classes. Parents were offered a small gift or travel reimbursement but were otherwise not compensated for their participation.

Apparatus

A wooden display box (98 cm wide × 105 cm high × 62 cm deep) was mounted on a table about 78 cm above the room floor. The infant faced an opening (98 cm × 60 cm) in the front of the box; between trials, a curtain was lowered in front of this opening. An experimenter sat behind a window (47 cm × 60 cm) centered in the back wall; the window extended from the bottom of the wall. The stimuli included a toy cat (30 cm long × 19 cm high), a doll (38 cm × 24 cm), and a doll shoe (11 cm × 5 cm).

Each infant was tested in a brightly lit room. Four 60-watt lamps attached to the apparatus provided additional light. Two fabric-covered frames (each 71 cm wide, 182.3 cm deep) stood at an angle on either side of the apparatus, blocking the infant's view from the rest of the room.

Events

The events consisted of two components, vocal cues and subsequent actions. In the following sections, the numbers in parentheses indicate the time taken to perform each component. A metronome beat softly once per second to help the experimenter(s) follow the scripts of events.

Vocal cues. In making the humorous cue, the experimenter laughed (2 seconds) and said, "I'm going to play with the kitty/baby," with a humorous acoustic profile (3 seconds). Specifically, she increased her pitch and loudness overall and increased her pitch incrementally across the sentence, so that it began at the middle range and ended at the high range (i.e., a rising contour; see Figure 1, bottom panel, dotted line). This pattern is similar to asking a question and has been shown to be a typical way in which parents express humor to their children (Hoicka & Gattis, 2011).

In making the sweet cue, the experimenter uttered, "Awww" (2 seconds) and said the sentence, "I'm going to play with the kitty/baby," with a sweet acoustic profile (3 seconds), while smiling to the same extent as when she uttered the humorous cue. Specifically, she decreased her pitch and loudness

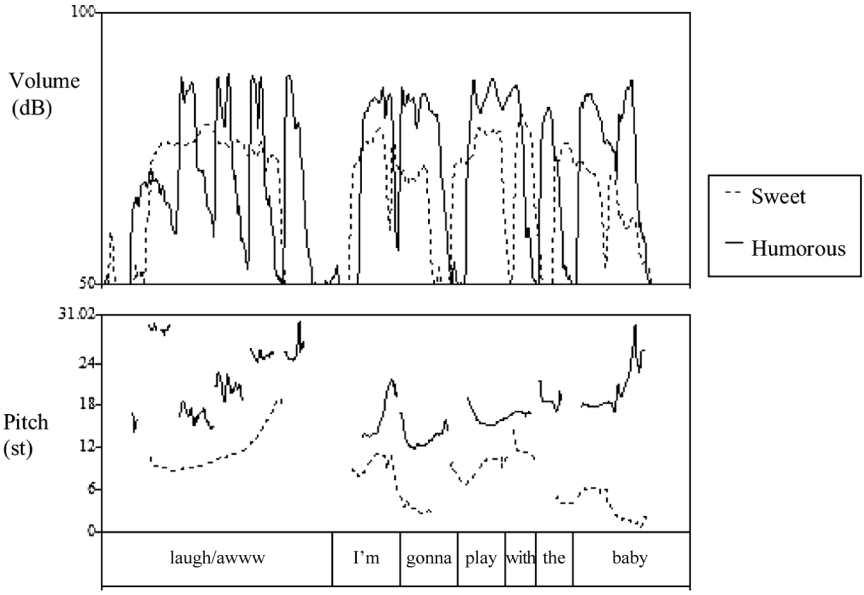

FIGURE 1 Examples of volume and pitch contours for humorous and sweet sentences. These contours are based on two randomly selected recordings from the experiment.

overall, and the sentence was spoken with a rise–fall pattern such that the pitch began at the low range, increased to the middle range, and then fell again to the low range (Figure 1, solid line). Katz, Cohn, and Moore (1996) found that parents often used rise–fall contours with their infants but that this contour had no specific pragmatic function. Note that the content of the sentence was identical across humorous and sweet cues.

Three steps were taken to ensure consistent prosody across participants. First, the same experimenter gave the vocal cues for all participants. Second, we examined a subset of audio tracks that were clearly recorded ($N = 15$), using Praat acoustic analysis software, to visually inspect the contours of each cue. The results confirmed that humorous cues followed a rising contour, whereas sweet cues followed a rise–fall contour. Finally, we compared the mean fundamental frequency (*F0*, pitch in semitones, *st*) and the mean energy (loudness, in decibels, *dB*), using a $2 \times 2 \times 2$ analysis of variance (ANOVA) with cue (humorous or sweet) and order (whether the cue was given first or second during the trial) as between-subject factors and trial type (match or mismatch) as a within-subject factor. The analysis indicated that humorous cues were made with a significantly higher mean *FO* ($M = 18.77$ st, $SD = 1.68$ st) than sweet cues ($M = 12.42$ st, $SD = 0.79$ st),

$F(1, 13) = 93.19$, $p < .001$, $\eta_p^2 = .878$. However, there was no difference between the mean energies of humorous ($M = 63.63$ dB, $SD = 5.00$ dB) and sweet cues ($M = 62.98$ dB, $SD = 4.34$ dB), $F(1, 13) = 0.70$, $p = .795$, which may be due to insensitivity of the recording device. There were no main effects of trial type on mean F0, $F(1, 13) < 0.01$, $p = .964$, or mean energy, $F(1, 13) = 0.07$, $p = .794$. The cue × trial type interaction was not significant either, for mean F0, $F(1, 13) = 1.39$, $p = .259$, or for mean energy, $F(1, 13) = 2.60$, $p = .130$. The low standard deviations and the lack of differences between match and mismatch trials suggested that cues were consistent across participants.

Fifteen adults heard vocal cues (two random selections each from the sample of 15 clearly recorded audio tracks) and judged whether each cue expressed humor or sweetness. The order of the cues was approximately counterbalanced across participants. All participants judged the sweet cues to be sweet, and 14 out of 15 (93%) judged the humorous cues to be humorous, confirming that the vocal cues signaled what they were intended to signal.

Actions. Immediately after the vocal cue, each infant saw one of four actions. The experimenter wore a hat that concealed her eyes throughout the experiment to prevent the infant from attempting to make eye contact with her. The absence of gaze exchange between the infant and experimenter reduced the likelihood for distraction to occur. Additionally, the experimenter kept a neutral facial expression when performing actions.

The following four actions are described from the infant's viewpoint. In the humorous action involving the toy cat, the experimenter reached for the cat with her right hand and placed it on her head (2 seconds). Next, she stroked her head with the cat from the right to the left side (4 seconds; Figure 2A). In the sweet action involving the cat, the experimenter reached for the cat with her right hand and placed it in the crook of her bent left arm, centered in front of her torso (2 seconds). Next, she stroked the cat on its back from right to left (4 seconds; Figure 2B). In the humorous action with the doll, the experimenter held the doll in her bent left arm such that the doll was on its side facing the infant, with the head facing right. The experimenter first reached for the shoe with her right hand and put it on the doll's hand (2 seconds). Next, she removed the shoe from the doll's hand and put the shoe back on the doll's hand (4 seconds; Figure 2C). The sweet action with the doll was similar to the humorous one, except that the shoe was placed on the doll's foot (Figure 2D).

Twenty adults judged whether each action was humorous or sweet from randomly selected recordings. Half of the participants saw the actions in the

FIGURE 2 (A) Humorous action with a cat: The actor stroked her head with the toy cat. (B) Sweet action with a cat: The actor stroked the cat. (C) Humorous action with a doll: The actor placed a shoe on the doll's hand. (D) Sweet action with a doll: The actor placed a shoe on the doll's foot. (Color figure available online).

order of humorous/cat, sweet/cat, humorous/doll, and sweet/doll. The other half saw the actions in the reverse order. All participants judged that the sweet actions were sweet, and 19 out of 20 (95%) judged that the humorous actions were humorous, confirming that the actions were humorous and sweet as intended.

Each test trial consisted of a pretrial and a main trial. The main trial immediately followed the pretrial, and the same action was demonstrated during the main trial and pretrial. During the pretrial, infants first heard the vocal cue (5 seconds); after a pause (1 second), they watched the action once (6 seconds). After another pause (1 second), infants heard the vocal cue again (5 seconds). After a third pause (1 second), they watched the last 4-second segment of the action. Thus, the pretrial lasted 23 seconds. During the main trial, the last 4-second segment of the action was repeated until the trial ended (see "Procedure").

Design

Between-subject variables included condition (same agent, different agent), cue (humorous, sweet), and toy (cat, doll), whereas within-subject variables included action type (humorous, sweet) and trial type (match, mismatch). It is important to note that we compared looking times to the exact same actions in the main trials, with only the vocal cue uttered during the pretrial differentiating whether the trial type was match or mismatch.

Procedure

Prior to the experiment, both experimenters briefly spoke to the infant. The infant sat on a parent's lap about 60 cm in front of the lowered curtain. Parents were instructed to remain quiet and neutral during the experiment. Each infant watched two test events—a mismatch event on one trial and a match event on the other—and was randomly assigned to one of the following four combinations of vocal cues and actions: 1) the humorous cue paired with the humorous action involving the cat (i.e., match event) on one trial and the same cue paired with the sweet action involving the cat (i.e., mismatch event) on the other trial; 2) the humorous cue paired with the humorous and sweet actions involving the doll; 3) the sweet cue paired with the humorous and sweet actions involving the cat; and 4) the sweet cue paired with the humorous and sweet actions involving the doll. The order of the events (i.e., whether the match event was shown first) was counterbalanced across infants. In the different-agent condition, the speaker stood to the left of the actor but out of the infant's view and made the visual aspects of the test events as similar as possible across conditions. In each condition, half of the infants heard sweet cues and half heard humorous cues; half of the infants saw actions involving the toy cat and half involving the doll.

Two observers, blind to the research hypothesis and unable to see the events, watched the infants through peepholes in the fabric-covered frames. Although the observers could still hear the vocal cues, they could not detect whether the trial consisted of a match or mismatch event as they could not see the action that followed the vocal cue. In addition, the same speaker uttered the vocal cues across conditions. Therefore, observers could not tell from vocal cues alone which condition the infant was assigned to.

The observers pressed a button connected to a computer when the infants looked at the event. The looking times recorded by the primary (typically more experienced) observer were used to determine the termination of the trials. Looking times during the pretrial and main trial portions were computed separately. The infants' looking times during the 23-second pretrial ranged from 19.0 seconds to 23.0 seconds and averaged 22.0 seconds,

indicating that they tended to look continuously at the event during the pretrial. The main trial ended when the infants: 1) looked away from the event for 1 consecutive second after having looked at it for at least 4 cumulative seconds, or 2) looked at it for 30 cumulative seconds. The 4-second minimum value was chosen to ensure that the infants had the opportunity to watch at least one cycle of the action during the main trial. Interobserver agreement during the main trial portions was calculated for 29 infants (90.6%; only one observer was present for 3 infants). Cohen's kappa confirmed that agreement was high, $k = .74$.

RESULTS

Preliminary analyses revealed no significant effects involving sex, order, or toy (cat or doll) interacting with action type (humorous, sweet) and cue (humorous, sweet), $Fs(1, 24) < 1.32$, $ps > .1$. Therefore, these factors (i.e., sex, order, and toy) were excluded in the subsequent analyses.

The infants' looking times during the main trial (Figure 3) were compared by a $2 \times 2 \times 2$ ANOVA, with condition (same or different agent) and cue (humorous, sweet) as between-subject factors and with action type (humorous, sweet) as a within-subject factor. The analysis yielded a significant interaction of condition $\times$ cue $\times$ action type, $F(1, 28) = 5.15$, $p = .031$, $\eta_p^2 = .155$. Planned comparisons indicated an interaction between cue and action type in the same-agent condition, $F(1, 14) = 7.88$, $p = .014$, $\eta_p^2 = .360$, whereas

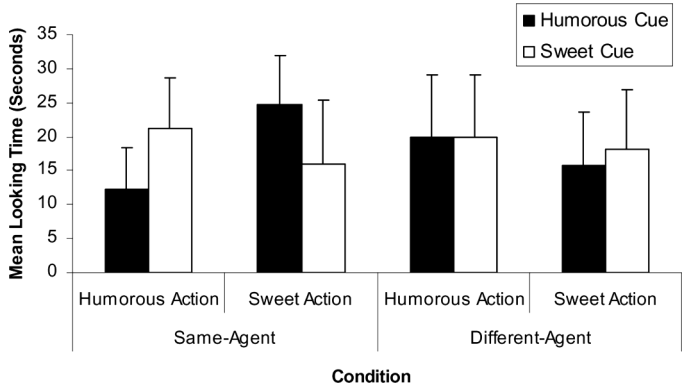

FIGURE 3 The infants' mean looking time at the humorous and sweet actions when paired with humorous and sweet cues in the same-agent and different-agent conditions. Error bars represent standard deviations.

no significant interaction was found in the different-agent condition, $F(1, 14) = 0.16$, $p = .696$. The infants in the same-agent condition looked significantly longer when cues and actions did not match ($M = 23.0$ seconds, $SD = 7.3$ seconds) than when they matched ($M = 14.0$ seconds, $SD = 8.0$ seconds), whereas those in the different-agent condition tended to look equally at the mismatch ($M = 17.7$ seconds, $SD = 8.7$ seconds) and match events ($M = 19.0$ seconds, $SD = 8.8$ seconds). Nonparametric Wilcoxon signed-rank tests confirmed these results (same agent: $T = 24$, $p < .01$; different agent: $T = 58.5$, $p > .10$).

Note that there was no main effect of vocal cue (humorous: $M = 18.11$, $SD = 8.76$; sweet: $M = 18.78$, $SD = 8.63$) or of action type (humorous: $M = 18.27$, $SD = 8.58$; sweet: $M = 18.62$, $SD = 8.81$). This suggests that: a) the infants who heard the humorous cue responded similarly to those who heard the sweet cue, and b) infants responded similarly when seeing the humorous or sweet action.

The results indicated that the infants in the same- but not the different-agent condition responded with heightened attention when the vocal cues did not match the subsequent actions. These findings support that the 15-month-olds appeared to expect the agent's goal-directed action to be congruent with the vocal cue she had just expressed, but they were not sensitive to the inconsistency between actions and vocal cues provided by different agents.

DISCUSSION

Matching Vocal Cues to Goal-Directed Actions

In the same-agent condition, when the experimenter expressed a sweet cue, the 15-month-olds looked reliably longer when the same experimenter performed a humorous action than when she performed a sweet action. In contrast, when the experimenter expressed a humorous cue, the infants exhibited the opposite looking pattern: They looked reliably longer at the sweet than at the humorous action. In the different-agent condition, however, the infants looked about equally at the match and mismatch events. These results suggest that the infants seemed to relate vocal cues to the appropriate goal-directed actions, within, but not between, agents.

The present study adds to research on infants' ability to perform cross-modal matching and demonstrates that infants not only match faces to voices and voices to actions in general (Bahrick, Hernandez-Reif, & Flom, 2005; Klin, Lin, Gorrindo, Ramsay, & Jones, 2009) but also match specific types of vocal cues to specific types of goal-directed actions. This is especially

impressive as infants responded differently to the exact same actions; the only component to differentiate between match and mismatch events was the vocal cues briefly uttered prior to the actions. Additionally, our task served as a conservative measure of infants' ability to match cues and actions by keeping facial cues constant at all times. In a naturalistic setting, facial expressions of their social partners tend to vary depending on the cues uttered, which would make it easier for infants to perform cue–action matching.

The present study also informs us about infants' sensitivity to vocal acoustic cues in general. The vocal cues used in the present experiment included: a) a brief vocalization of laughter or "Awww," and b) the sentence, "I'm going to play with the kitty/baby," spoken with different intonations. Because the content of the sentence was identical in humorous and sweet utterances, the infants could only rely on the vocalization and/or speech intonation to make the distinction between humor and sweetness. Past research found that 6-month-olds differentiated among positive, neutral, and negative speech (e.g., Singh et al., 2002), and that 4-month-olds distinguished among intonation contours indicating approval and disapproval (e.g., Fernald, 1993; Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990). More recently, Sakkalou and Gattis (2011) found that intonation alone was sufficient for 14- to 18-month-olds to distinguish intentional actions from mistakes. Here, we showed that 15-month-olds were sensitive to the differences between acoustic cues that signaled two types of communicative intentions, humor versus sweetness, and were capable of mapping these cues to corresponding actions. Future research can extend this finding to examine the roles of vocalization and speech intonation separately. For example, one could examine whether infants still make the distinction if they hear a brief vocalization only (i.e., laughter vs. "Awww").

The present study also adds to existing research showing that infants detect goal-directed actions based on previous actions (e.g., Bíró & Leslie, 2007; Brandone & Wellman, 2009; Sommerville & Crane, 2009; Woodward, 1998), actor attributes (e.g., Guajardo & Woodward, 2004), emotional cues (e.g., Barna & Legerstee, 2005; Behne et al., 2005; Legerstee, Barna, & DiAdamo, 2000; Phillips et al., 2002), and object affordances (e.g., Huang, Heyes, & Charman, 2002; Meltzoff, 1995). Whereas past research has found that infants detected goal-directed actions by all of these means and used emotional cues to differentiate intentional from unintentional actions (e.g., Carpenter et al., 1998; Hoicka & Gattis, 2008), the present study showed that infants appeared to relate two different types of positively emoted vocal cues appropriately to two different types of goal-directed actions. Thus, infants did not simply respond to the affective valence of vocal cues but were also sensitive to differences between positively emoted vocal cues and were able to use them to predict others' goal-directed actions.

Intentional Understanding Versus Associative Learning

Gergely, Nádasdy, Csibra, and Bíró (1995) and Tomasello and Rakoczy (2003) suggested that infants see others as intentional agents from the end of the first year. Based on this view, the infants in the present study may have understood that humorous actions fulfill goals of humorous intentions and that sweet actions fulfill goals of sweet intentions. The infants in the same-agent condition looked longer at the mismatch than at the match events because they were intrigued or surprised that the actor did not fulfill her intentions. Additionally, the infants in the different-agent condition did not make such a distinction because one person's actions may or may not fulfill another person's goals. Thus, according to this intentional framework, our 15-month-olds used vocal cues to interpret the intention of the speaker as being humorous or sweet and expected her subsequent actions to be consistent with this intention.

Another way to interpret the present findings is through the framework of associative learning (e.g., Elsner, 2007). Infants may have observed that certain vocal cues (e.g., humorous cues) tend to co-occur with certain actions (e.g., humorous actions). Based on this view, infants' success in matching vocal cues to subsequent actions does not require their interpretation of underlying goals or intentions. Instead, it requires past exposure to events that illustrate the statistical relations (i.e., which vocal cues tend to co-occur with which actions). If associative or statistical learning underlies the present findings, an interesting question to be addressed in future research would be why infants apply the association within, but not across, individuals.

Although it may be the case that infants did not match cues to actions during the different-agent condition because one person's cue need not impact on another person's actions, other reasons may explain the null result. For example, infants might have been distracted or confused by the disembodied voice or simply had too much information to process. However, hearing someone without seeing that person is not an uncommon experience, and infants are very likely to experience similar situations in their everyday lives by the age of 15 months. Indeed, if infants had been surprised by the different-agent condition, we would expect them to look longer overall during the different-agent than during the same-agent condition. But they did not. Regardless of the condition assigned, both experimenters spoke to the infants beforehand and brought their attention to the presence of two experimenters. Research showed that infants as young as 6 months match individual faces and corresponding voices (e.g., Bahrick et al., 2005); thus, we believe that the brief interaction prior to the experiment should have prevented our 15-month-olds from being confused when a

person different from the actor voiced the vocal cue in the different-agent condition. Additionally, hearing words from voices alone without seeing a corresponding face has been shown to help 10-month-olds categorize novel visual stimuli (Plunkett, Hu, & Cohen, 2007), suggesting that hearing a voice without seeing the corresponding face should not be confusing to 15-month-olds.

Conclusions

Although further research is still needed to specify the developmental mechanism underlying the present findings, our results clearly demonstrate a developmental advancement in the understanding of humor. Whereas 10-month-olds spontaneously respond to humorous actions (Sroufe & Wunsch, 1972), 15-month-olds are capable of distinguishing between humor and sweetness schemes manifested in vocal cues and actions and are sensitive to events that violate either scheme. The findings also support that infants at this age do not rely on affective valence alone when reacting to others' actions. Instead, they possess a fairly sophisticated understanding of vocal cues and actions even when they are along the same emotional spectrum.

REFERENCES

Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face–voice relations. *Developmental Psychology*, *41*, 541–552.

Barna, J., & Legerstee, M. (2005). Nine- and 12-month-old infants relate emotions to people's actions. *Cognition & Emotion*, *19*, 53–67.

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology*, *41*, 328–337.

Bíró, S., & Leslie, A. M. (2007). Infants' perception of goal-directed actions: Development through cue-based bootstrapping. *Developmental Science*, *10*, 379–398.

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science*, *20*, 85–91.

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*, 287–314.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*, 315–330.

Elsner, B. (2007). Infants' imitation of goal-directed actions: The role of movements and action effects. *Acta Psychologica*, *124*, 44–59.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, *64*, 657–674.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Guajardo, J. J., & Woodward, A. L. (2004). Is agency skin-deep? Surface attributes influence infants' sensitivity to goal-directed action. *Infancy*, *6*, 361–384.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, United Kingdom: Cambridge University Press.

Hoicka, E., & Gattis, M. (2008). Do the wrong thing: How toddlers tell a joke from a mistake. *Cognitive Development*, *23*, 180–190.

Hoicka, E., & Gattis, M. (2011). *Acoustic differences between humorous and sincere communicative intentions*. Manuscript submitted for publication.

Hoicka, E., Jutsum, S., & Gattis, M. (2008). Humor, abstraction, and disbelief. *Cognitive Science*, *32*, 985–1002.

Huang, C. T., Heyes, C., & Charman, T. (2002). Infants' behavioral reenactment of 'failed attempts': Exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology*, *38*, 840–858.

Katz, G. S., Cohn, J. F., & Moore, C. A. (1996). A combination of vocal F0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*, *67*, 205–217.

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to nonsocial contingencies rather than biological motion. *Nature*, *459*, 257–263.

Ladd, D. R., Silverman, K., Tolkmitt, F., Bergmann, G., & Scherer, K. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, *78*, 435–444.

Legerstee, M., Barna, J., & DiAdamo, C. (2000). Precursors to the development of intention at 6 months: Understanding of people and their actions. *Developmental Psychology*, *36*, 627–634.

McGhee, P. E. (1979). *Humor: Its origin and development*. San Francisco, CA: W. H. Freeman and Company.

Meltzoff, A. (1995). Understanding the intentions of others: Reenactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*, 838–850.

Newman, R. (2003). Prosodic differences in mothers' speech to toddlers in quiet and noisy environments. *Applied Psycholinguistics*, *24*, 539–560.

Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). Fifteen-month-old infants detect violations in pretend scenarios. *Acta Psychologica*, *124*, 106–128.

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, *13*, 539–545.

Phillips, A., Wellman, H., & Spelke, E. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*, 53–78.

Plunkett, K., Hu, J., & Cohen, L. B. (2007). Labels can override perceptual categories in early infancy. *Cognition*, *106*, 665–681.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- to 18-month-olds. *Developmental Psychology*, *22*, 12–21.

Sakkalou, E. C., & Gattis, M. (2011). *Fourteen- to 18-month-old infants infer others' intentions from intonation*. Manuscript submitted for publication.

Shultz, T. R. (1976). A cognitive–developmental analysis of humor. In A. J. Chapman & H. C. Foot (Eds.), *Humor and laughter: Theory, research, and applications* (pp. 11–36). London, United Kingdom: John Wiley & Sons.

Singh, L., Morgan, J., & Best, C. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, *3*, 365–394.

Sommerville, J. A., & Crane, C. C. (2009). Ten-month-old infants use prior information to identify an actor's goal. *Developmental Science*, *12*, 314–325.
Sroufe, L. A., & Wunsch, J. P. (1972). The development of laughter in the first year of life. *Child Development*, *43*, 1326–1344.
Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language*, *18*, 121–147.
Trainor, L. J. (1996). Infant preferences for infant-directed versus non-infant-directed play songs and lullabies. *Infant Behavior and Development*, *19*, 83–92.
Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis for preferences of infant-directed singing. *Infant Behavior and Development*, *20*, 383–396.
Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.