Infants infer intentions from prosody

Elena Sakkalou*, Merideth Gattis

School of Psychology, Cardiff University, Cardiff, United Kingdom

ARTICLE INFO

Keywords:
Intentions
Prosody
Imitation
Mental state understanding

ABSTRACT

Two studies were conducted to examine infants’ ability to discern intentions from lexical and prosodic cues. Two groups of 14–18-month-olds participated in these studies. In both studies, infants watched an adult perform a sequence of two-step actions on novel toys that produced an end-result. In the first study actions were marked intentionally with both lexical and prosodic cues. In the second study, the lexical markers of intention were presented in Greek, thus providing infants with prosodic but not lexical cues. In both studies, infants reproduced more intentional than accidental actions, suggesting that infants can infer intentions from prosodic cues.

© 2011 Elsevier Inc. All rights reserved.

Ever since Premack and Woodruff coined the term theory of mind in 1978, a wealth of developmental studies have investigated the trajectory of children’s coming to attribute mental states to themselves and others. Recent advances in nonverbal methods have yielded evidence for some aspects of mental state understanding in infants as young as 9 months (Behne, Carpenter, Call, & Tomasello, 2005; Carpenter, Call, & Tomasello, 2005). However, we still understand little about how infants begin to attribute mental states, such as intentions, desires, and beliefs, to others. In two experiments, we aimed to investigate the contribution of prosodic cues to infants’ understanding of mental states.

Previous research on mental state understanding has focused mainly on how infants use visual cues to infer mental states. These studies have shown that infants use visual information about head turns, eye gaze, and actions to distinguish an actor’s attention to one object versus another, and facial expressions to distinguish encouragement and prohibition as well as intentionality (Behne, Carpenter, & Tomasello, 2005; Meltzoff, 1995; Moll & Tomasello, 2004; Moore & Cockum, 1998; Nielsen, 2006;
Olineck & Poulin-Dubois, 2005; Sorce, Emde, Campos, & Klinnert, 1985; Woodward & Guajardo, 2002). For example, Behne, Carpenter, Call, et al. (2005) demonstrated that 9–18-month-olds became more impatient when an experimenter was unwilling to give them a toy (indicated by a partial offer of the toy, accompanied by a teasing smile) than when the experimenter was unable to give them a toy (indicated by dropping the toy, accompanied by a facial expression of surprise and frustration). Similarly, in another study, Behne, Carpenter, and Tomasello (2005) demonstrated that 14–24-month-olds can differentiate communicative gestures and facial expressions from non-communicative gestures and facial expressions. Infants observed an adult either gesturing (pointing or gazing) in a communicative manner toward a box to indicate that there was a toy inside the box, or gesturing in an absently minded and thus non-communicative manner toward the box. In the communicative conditions, children of all ages were able to use communicative cues to identify the location of the toy. In the non-communicative conditions, children searched equally at both locations (Behne, Carpenter, & Tomasello, 2005). These studies and a growing number of others indicate that infants are able to identify and use facial and action cues signaling the intentions of social partners.

Although the relevance of visual cues is established, we still know little about the role of auditory cues in mental state understanding. An important acoustic cue that conveys mental states is prosody. Prosodic information offers meaning beyond that which is conveyed by lexical information and encompasses several acoustic features of speech, such as pitch, intensity, duration, and pitch contour. For example, prosody helps in deducing others’ mental states by marking changes of topic in discourse, signaling the moment of turn-taking between speaker and listener, and conveying complex attitudes such as condescension (Cruttenden, 1997). Furthermore, prosody signals the emotional state and the communicative intentions of the speaker, such as when someone is being ironic or sarcastic (Cruttenden, 1997; Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985). Here we present two experiments investigating whether infants use prosodic features of human speech to distinguish complex mental states.

These experiments were motivated by research indicating that audition plays an important role in introducing infants to the social world (Fernald, 2004). Infants perceive prosody from an early age. As they get older, they use prosodic cues as a guide to how to act in the world.

Prosody is a notable feature of infant-directed speech (Katz, Cohn, & Moore, 1996). Mothers use prosody to signal communicative intent to infants. Prosodic patterns in maternal speech convey messages such as approval versus disapproval and encourage the infant to attend to an object or to what is being said (Fernald, 1989, 1993; Papousek, Bornstein, Nuzzo, Papousek, & Symmes, 1990; Spence & Moore, 2003; Stern, Spieker, & MacKain, 1982). Mothers use rising contours to engage or encourage the child’s attention to an object or other visual stimuli, falling contours to soothe or comfort their infants, and rise–fall contours to communicate approval. In addition, mothers adapt the acoustic profiles in their vocalizations as the intent and context of their communications change. Kitamura and Burnham found that mothers used more expressive vocalizations when communicating with their infants at 6 and 12 months and varied their pitch range to communicate more directive vocalizations at 9 months (Kitamura & Burnham, 2003).

Infants are not only surrounded by prosodic cues; they are also sensitive to prosody from an early age. Very young infants have a preference for the wider spectral properties of infant-directed speech (Cooper & Aslin, 1990; Fernald & Kuhl, 1987). By 5 months, infants smile in response to approval vocalizations and display negative affect in response to disapproval vocalizations, regardless of whether the approval and disapproval messages are expressed in their native language (Fernald, 1993).

Infant preferences for specific prosodic patterns change with development just as maternal vocalizations change with infant age. Kitamura and Lam (2009) reported that infants preferred comforting vocalizations at 3 months, approving vocalizations at 6 months, and directive vocalizations at 9 months, a pattern that parallels changes in maternal vocalizations reported by Kitamura and Burnham (2003). This developmental trend in infants’ preferences for different vocalizations suggests that infants’ sensitivity to prosody changes with development, from first appreciating prosodic messages of simple reassurance to later appreciating more complex comments on the world surrounding the infant.

By the beginning of the second year, infants are not only able to discriminate between different kinds of vocalizations; they are also beginning to consider prosody as a guide to how to act on the world.
At least two studies investigating social referencing in 12-month-olds have compared the effects of visual and vocal maternal cues when infants were presented with novel toys or placed on a visual cliff. The results of both studies indicate that vocal cues are more effective than facial cues in regulating infant behavior (Mumme, Fernald, & Herrera, 1996; Vaish & Striano, 2004). For example, Vaish and Striano (2004) reported that infants were quicker to cross a visual cliff when mothers made positive vocalizations than positive facial expressions, even though mothers had their backs to infants during vocalization trials.

Based on the evidence regarding maternal use of prosody, infant sensitivity to prosody, and the effect of prosody on infant behavior in the second year, we reasoned that prosody might also support infants’ inferences about intentional versus accidental actions. To test whether infants infer intentions from prosody, we used a paradigm developed by Carpenter, Akhtar, and Tomasello (1998) to investigate toddlers’ imitative behavior. Carpenter et al. examined whether infants between the ages of 14 and 18 months imitate intentional actions more than accidental actions. An experimenter performed two actions on a toy, indicating one as accidental by saying “Whoops” and the other as intentional by saying “There.” When given the opportunity to play with the toy, infants reproduced more of the actions that had been marked as intentional than those that had been marked as accidental.

The Carpenter et al. paradigm allowed us to examine the contribution of prosodic cues to infants’ understanding of others’ intentions. We hypothesized that prosodic cues allow infants to distinguish intentional and accidental actions and on that basis predicted that infants would reproduce more intentional than accidental actions, even when prosody was the only indicator of mental state. Experiment 1 replicated the methodology used by Carpenter et al. By combining acoustic analyses with behavioral analyses, we sought to clarify the prosodic patterns in the original paradigm and at the same time to assess infants’ ability to distinguish intentional and accidental actions in an imitation paradigm. We hypothesized that the utterances accompanying intentional and accidental actions would have distinct prosodic contours and that infants would utilize the utterance cues to infer intentionality, leading them to reproduce more intentional than accidental actions.

Experiment 2 investigated the effect of prosody in the absence of lexical cues. To do so, the experimenter used the same prosodic contours but spoke in an unfamiliar language to mark actions as intentional or accidental. We again combined prosodic and behavioral analyses and were able to address the question of whether infants can infer intention from prosody alone. We hypothesized that prosodic contours, rather than lexical information, would influence infants’ understanding of intentions, again leading them to reproduce more intentional than accidental actions, but this time solely on the basis of prosody.

1. Experiment 1

The goals of Experiment 1 were: (1) to replicate the methodology used in the Carpenter et al. (1998) study with lexical and prosodic cues, (2) to identify the prosodic contours signaling intentional and accidental actions, and (3) to evaluate the effect of lexical and prosodic cues on understanding intention. We hypothesized that the accidental (Whoops) and intentional (There) utterances would have distinct prosodic contours and that infants would utilize the words and the prosodic cues to reproduce more intentional than accidental actions.

2. Method

2.1. Participants

Participants were 28 infants (11 girls; mean age 16 months, 11 days; range 14 months, 14 days to 18 months, 14 days). Another two infants began the procedure but were excluded due to restlessness. Five more infants were excluded due to experimenter error, parental interference or toy malfunction. Participants were tested in a child friendly lab at Cardiff University. Infants received a t-shirt for participation in the study.
2.2. Materials

Five novel toys (one for the warm-up and four for the experimental trials) were constructed from wooden boxes and from plastic and iron items. As in the Carpenter et al. (1998) study, two objects were mounted on the toys, and a distinct action was performed on each of these two objects. An interesting result, such as a puppet appearing in a window (the end result) occurred after the experimenter performed the second of the two actions. The toys were constructed so that a single experimenter was able to produce the end result surreptitiously. The experimenter manipulated the action items mounted on the front of the toy with the left hand and discreetly operated the end result in the back of the toy with the right hand.

2.3. Design

In the Carpenter et al. (1998) study, two actions were modeled on each toy. Accidental actions were immediately followed by the word “Whoops” and intentional actions by the word “There,” each with the appropriate prosody. The prosodic contours are shown in Figs. 1 and 2.

There were two orders of action sequence: the I–A sequence, where the first action was indicated as intentional and the second as accidental, and the A–I sequence, where the first action was accidental and the second intentional. All infants participated in each sequence twice. The sequences were presented alternately, and there were four possible counterbalancing orders. In addition, for each of these orders, any given toy was presented in a different ordinal position. For example, the “puppet” toy was presented second in order 1 and third in order 2. The order in which the actions were performed was also counterbalanced. For example, for the “puppet” toy, in one order the rolling action was first and indicated as intentional, and the lifting action was second and indicated as accidental. Alternatively, in a different order, the lifting action was first and indicated as intentional, and the rolling action was second and indicated with accidental prosody.

2.4. Procedure

During the experiment, the infant sat on the parent’s lap in front of a table. The experimenter sat at an adjacent side of the table near the infant. Two cameras recorded the events. One focused on a profile view of the infant from the waist up, and the second captured both experimenter and infant.

In the warm-up phase, the experimenter modeled an action on a toy, and as an end result, music played through two speakers that were fixed on the left- and right-hand side of the toy. There were no verbal cues during this demonstration. Some infants were reluctant to touch the toy, in which case the experimenter encouraged them by saying, “Can you make it work?” If infants still did not reproduce the action, the experimenter used verbal encouragement until they were able to do it alone or with the experimenter’s help. We encouraged infants to touch the toy to ensure that they were comfortable with the toys and would be willing to perform actions on the experimental ones on their own.

During the test phase, the experimenter modeled two actions on each toy. Before modeling the actions, the experimenter monitored the infant’s attentional states and directed the infant’s attention to the toy by saying, “Look, this is how it works,” or “Ready, this is how it works.” The actions had a short delay between them and were enacted rapidly in order to appear credible. For example, for the I–A sequence, with the “puppet” toy, the experimenter first performed the rolling action and said “There” with a purposeful and satisfied tone of voice. Then, she performed the pushing action and said “Whoops” with a dissatisfied and surprised tone of voice. A puppet appeared (the end-result), about 0.5–1 s after the second action. Following the demonstration, infants were told, “It’s your turn now. Can you make it work?”

The same actions were modeled a second time on the same toy, and the infant again had a chance to produce a response. Thus, in total, infants saw two demonstrations and had two opportunities to respond for each of four toys, yielding a total of eight demonstrations and eight response trials. As in the Carpenter et al. study, during the infants’ response phase, the end result was activated only if the infant had reproduced the intentional action, but regardless of whether the infant had also reproduced
an accidental action. The end result followed between 1 and 2 s after performing the intentional action. Throughout testing, the experimenter maintained a neutral facial expression.

2.5. **Scoring**

Infants’ responses were scored from video recordings. For each of the eight trials, infants were scored for reproducing actions in the following categories: intentional action only, accidental action only, both actions in the same order as demonstrated, both actions in the opposite order from that demonstrated, and neither action. To account for one instance where one infant produced neither action, the actions infants had performed in each category were divided by the number of trials on which infants did respond (usually eight) and then converted to percentages.
Reliability analyses were carried out by a coder blind to experimental condition who scored videos of seven infants. Comparison between primary and secondary coding yielded a Cohen's kappa of .92, indicating good reliability.

3. Results

3.1. Prosody manipulation checks

Analyses of prosodic features of the experimenter’s demonstration were carried out on a sample of “Whoops” and “There” utterances. Four utterances were taken from four random trials from 14 participants. Two intentional utterances and two accidental utterances were taken from each of the 14
participants for analysis (50% of the sample). Analyses compared the mean pitch (Hz), the pitch range (Hz), the duration in seconds (s), and the amplitude (dB) of “Whoops” and “There” utterances. Pairwise comparisons demonstrated significant differences between intentional and accidental prosody. Intentional prosody had higher amplitude, t(57) = 20.86, p < .001, and longer duration, t(57) = 11.25, p < .001, than accidental prosody. The mean pitch, t(57) = 17.66, p < .001, was higher and the pitch range was wider for accidental prosody, t(57) = -4.21, p < .01. This high pitch and wide pitch range was mainly due to lexical reasons, in particular due to the “ps” sound in the word “Whoops,” which is a high frequency sound with a wide frequency range. The means and standard deviations appear in Table 1.

Pitch contour analyses revealed that the intentional marker, “There,” was characterized by a falling pitch, and the accidental marker, “Whoops,” was characterized by a rising pitch. (See Figs. 1 and 2.)

3.2. Behavioral analyses

We conducted a Kruskal–Wallis test to test for an effect of toy order. Results showed no differences among the four toy orders in both the A–I and the I–A sequences, so toy order was excluded from further analyses.

We then compared the overall percentage of intentional actions to the overall percentage of accidental actions. This analysis included all types of responses including correct intentional responses and accidental responses. Assumptions for parametric analyses were not met; we therefore used non-parametric statistics. A Wilcoxon signed-rank test showed that infants produced more intentional actions (M = 72.62, SD = 17.91) than accidental actions (M = 49.40, SD = 21.92), z = -2.90, p < .001. The same pattern was found in both the A–I (z = -2.20, p < .05) and the I–A (z = -2.12, p < .05) sequence. For means and standard deviations, see Table 2.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Accidental–Intentional (A–I)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intentional</td>
<td>70.24</td>
<td>25.50</td>
</tr>
<tr>
<td>Accidental</td>
<td>50.60</td>
<td>25.65</td>
</tr>
<tr>
<td>Intentional–Accidental (I–A)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intentional</td>
<td>75.00</td>
<td>28.05</td>
</tr>
<tr>
<td>Accidental</td>
<td>48.21</td>
<td>32.58</td>
</tr>
</tbody>
</table>
Table 3
Infants’ percent production of the four possible action types by sequence of presentation.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Experiment 1</th>
<th></th>
<th></th>
<th>Experiment 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Accidental–Intentional (A–I)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intentional only</td>
<td>49.41</td>
<td>25.65</td>
<td></td>
<td>49.26</td>
<td>31.15</td>
<td></td>
</tr>
<tr>
<td>Accidental only</td>
<td>29.76</td>
<td>25.50</td>
<td></td>
<td>22.92</td>
<td>24.48</td>
<td></td>
</tr>
<tr>
<td>Same order</td>
<td>12.50</td>
<td>22.05</td>
<td></td>
<td>20.68</td>
<td>22.97</td>
<td></td>
</tr>
<tr>
<td>Incorrect order</td>
<td>8.33</td>
<td>15.71</td>
<td></td>
<td>7.14</td>
<td>15.60</td>
<td></td>
</tr>
<tr>
<td>Intentional–Accidental (I–A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intentional only</td>
<td>51.79</td>
<td>32.58</td>
<td></td>
<td>50.00</td>
<td>33.52</td>
<td></td>
</tr>
<tr>
<td>Accidental only</td>
<td>25.00</td>
<td>28.05</td>
<td></td>
<td>24.40</td>
<td>29.04</td>
<td></td>
</tr>
<tr>
<td>Same order</td>
<td>8.93</td>
<td>15.54</td>
<td></td>
<td>8.63</td>
<td>13.94</td>
<td></td>
</tr>
<tr>
<td>Incorrect order</td>
<td>14.29</td>
<td>18.54</td>
<td></td>
<td>16.97</td>
<td>21.08</td>
<td></td>
</tr>
</tbody>
</table>

We next examined whether this effect was due to infants selectively imitating intentional actions only by comparing responses of the intentional action only to the other three response patterns: accidental action only, both actions in the same order as demonstrated, and both actions in the opposite order from that demonstrated (see Table 2). Infants performed significantly more intentional actions than accidental actions (A–I sequence, $z = 2.20$, $p < .05$; I–A sequence, $z = 2.11$, $p < .05$), and more intentional actions than actions performed in the same order (A–I sequence, $z = 3.63$, $p < .001$; I–A sequence, $z = 3.95$, $p < .001$) and incorrect order (A–I sequence, $z = 4.13$, $p < .001$; I–A sequence, $z = 3.39$, $p < .001$). For means and standard deviations, see Table 3.

3.3. Additional manipulation checks

Although the experimenter was instructed to maintain a neutral facial expression and neutral actions, we carried out an analysis to ascertain that no facial or action cues signaled intentionality. An independent coder viewed seven videos with the volume turned off and rated whether the experimenter’s actions and facial expressions were intentional or accidental. The coder identified accidental and intentional actions of the experimenter at chance levels (50%). A chi-square analysis showed no significant association between correct “intentional only” actions produced by the infants and actions identified correctly and incorrectly by the coder, $\chi^2(1) = .30$, $p = .59$, confirming that intention was not identifiable on the basis of actions and/or facial expressions. This result strengthens our conclusion that imitation performance was determined by vocal markers to intention.

As in the Carpenter et al. (1998) study, in order to ascertain that the results were attributable to discrimination between accidental and intentional messages and not to training or practice, we carried out two additional analyses. We conducted a Friedman test to examine any differences between first and second trials, comparing performance on the first trials (trials 1, 3, 5, and 7) with performance on the second trials (trials 2, 4, 6, and 8) for all the toys. Infants responded correctly in 48% (SD = 25) of the first trials and in 51% (29) of the second trials. This difference was nonsignificant, $\chi^2(1) = .80$, $p = .37$. We conducted a Cochran’s Q-test to examine potential differences in performance between the first and eighth trials. Infants performed correctly 52% of the time (SD = 51) in the first trial and 68% of the time in the eighth trial (SD = 48). This difference was nonsignificant, $Q(1) = .09$, $p = .76$.

3.4. Summary of results

Experiment 1 demonstrates that vocal utterances marking intentional and accidental actions had distinct prosodic contours and confirms that 14–18-month-olds copied more intentional than accidental actions. Additional analyses demonstrate that this effect was due neither to facial nor action cues, nor to learning during the study.
4. Discussion

Analyses of infant behavior in the first experiment replicate those of Carpenter et al. (1998). After watching a demonstration of both intentional and accidental actions on a novel toy, infants reproduced more intentional than accidental actions. Prosodic analyses indicated that in addition to the lexical marker, “There,” intentional actions were marked by higher amplitude, longer duration, and a falling pitch contour. Accidental actions, in addition to the lexical marker, “Whoops,” were marked by higher mean pitch, higher pitch range, and a rising pitch contour. These differences in prosodic features indicate that intentional and accidental actions are marked by a complex pattern of prosodic cues. Although not all studies of maternal vocal cues have provided prosodic analyses (Mumme et al., 1996; Vaish & Striano, 2004), those that did so indicate that when speaking to younger infants, mothers use falling contours to convey either disapproval or comfort (depending on the rate of fall), and rising contours to draw attention (Fernald, 1992, 1993; Sullivan & Horowitz, 1983). Prosodic markers of intention versus accident do not therefore correspond to early markers of positive versus negative affect or approval versus disapproval, but are distinct.

Kitamura and Burnham reported that maternal use of prosodic cues changes with development, becoming increasingly complex for older infants (Kitamura & Burnham, 2003). Furthermore, infant sensitivity to prosodic cues also changes (Kitamura & Lam, 2009). On this basis, we reasoned that in the second year, infants might use complex prosodic features to identify complex mental states such as purposeful and accidental behavior. In our second experiment, we examined whether infants are able to discriminate intentional from accidental actions based on prosodic cues after lexical information is removed.

5. Experiment 2

In Experiment 1, 14–18-month-olds demonstrated an understanding of the intentions of others by responding differently to accidental actions, marked by an accompanying exclamation, “Whoops,” and intentional actions, marked by the exclamation, “There.” Analyses demonstrated that these two exclamations had distinct prosodic features, raising the question of whether infants were inferring intentions from the lexical markers “Whoops” and “There,” or from the prosody in which these words were expressed. Research suggests that prosody is a highly salient cue that provides various kinds of information to infants, supporting the claim that infants can use prosodic cues to make inferences about others’ intentions.

Consequently, the main goal of our second experiment was to examine whether infants can make attributions about mental states based on prosodic features such as those used in Experiment 1. We therefore used novel words to mark the two actions on the toys – the Greek words “Ochi” and “Nato” – and identified them as accidental or intentional with appropriate accidental and intentional prosody, based on the prosodic information identified in Experiment 1. To control for the possibility that lexical characteristics of the Greek words might affect intention understanding, half of the infants heard “Ochi” as intentional and “Nato” as accidental, whereas the other half heard “Ochi” as accidental and “Nato” as intentional. This experiment therefore consisted of two novel word conditions (Nato–Intentional and Ochi–Accidental; Ochi–Intentional and Nato–Accidental).

Because we added the variable of lexical–prosodic pairing to the design, we doubled the size of our sample in Experiment 2. We were therefore also able to address a secondary research question regarding age and intention understanding. Carpenter et al. (1998) grouped all infants aged 14–18 months together for their analyses and did not investigate whether intention understanding differs for younger and older infants. However, more recent studies on intention understanding have reported developmental differences in mental state understanding around this age (Bellagamba, Camaioni, & Colonnese, 2006; Olineck & Poulin-Dubois, 2005). In a replication of the Carpenter et al. study that included gaze, gestures, and facial and intonation as well as lexical cues, Olineck and Poulin-Dubois (2005) found that 18-month-olds were better able to discriminate between intentional and accidental actions than 14-month-olds, leading them to propose that infants’ intention understanding abilities increase between 14 and 18 months. To address this question, we included age-related analyses in our design.
We hypothesized that (1) despite the absence of familiar lexical cues such as “Whoops” and “There,” infants would still produce more intentional than accidental actions by making attributions about mental states, (2) the distinction between intentional and accidental actions would correspond to prosodic rather than phonetic features of the accompanying utterances, and (3) older infants would perform better than younger infants.

6. Method

6.1. Participants

Fifty-six infants (26 girls; mean age 16 months, 7 days, range 13 months, 27 days to 18 months, 15 days) participated. Another four infants were tested but grew restless and were thus excluded. Eight more infants were excluded due to experimenter error, parental interference, or toy malfunction. Infants came from English-speaking homes, although a small proportion came from bilingual homes (18%). No infants were from Greek-speaking homes.

6.2. Materials and design

The materials and design were the same as for Experiment 1.

6.3. Procedure and scoring

The procedure for Experiment 2 was the same as for Experiment 1 except that infants heard the lexical items “Ochi” and “Nato” instead of “Whoops” and “There.” Twenty-eight infants participated in the Nato–Intentional and Ochi–Accidental condition (first novel word condition) where “Nato” was expressed as intentional and “Ochi” as accidental. Another 28 infants participated in the Ochi–Intentional and Nato–Accidental condition (second novel word condition), where “Ochi” was expressed as intentional and “Nato” as accidental. Infant responses were scored as in Experiment 1. Also as in Experiment 1, the words were uttered so that intentional prosody sounded like the experimenter meant to perform the action and was satisfied (falling pitch pattern), whereas accidental prosody expressed dissatisfaction and surprise (rising pitch pattern). These patterns are shown in Figs. 1 and 2. Infant responses were coded from videotape by the first author. A coder blind to condition scored the videos of 14 infants, seven from each condition. Cohen's kappa was .93, indicating good reliability.

7. Results

7.1. Prosody manipulation checks

We conducted analyses of prosodic features of the experimenter’s demonstrations on a sample of utterances from each condition. Four utterances (two intentional and two accidental) were taken from four randomly selected trials from 28 participants (50% of the sample). Analyses compared mean pitch (Hz), pitch range (Hz), duration (s) and amplitude (dB). Pairwise comparisons demonstrated significant differences between intentional and accidental prosody. Intentional prosody had higher mean pitch, $t(57) = 17.66, p < .001$, higher amplitude $t(57) = 20.86, p < .001$, and longer duration $t(57) = 11.25, p < .001$. Pitch range, however, was wider for accidental prosody $t(57) = -4.21, p < .001$. Means and standard deviations can be seen in Table 1. Differences were consistent for accidental and intentional intonations across the two conditions.

We also carried out a judgment experiment with two adults to establish internal validity for intentional and accidental utterances expressed in the experiment. One male and one female adult heard 56 pairs of intentional and accidental utterances, half from the Nato–Intentional and half from the Ochi–Intentional condition. Four utterances were chosen at random from 14 sessions; number of utterances from the $A$–$I$ and $I$–$A$ orders were counterbalanced. The adults were told they would hear pairs of utterances from an experiment carried out with infants, in which an adult expressed accidental
and intentional prosody while performing actions on toys. They were told that there were two conditions in the experiment, so they would need to focus on prosody when making their judgments. Their instructions were to press one button if they thought the first utterance from the pair was intentional and to press a different button if they thought the second utterance was intentional. The utterances were cut and presented in pairs through the program SuperLab. The participants wore headphones throughout the experiment. The adults judged the intentional and accidental prosody correctly on 89% of utterances.

7.2. Behavioral analyses

To adjust for 18 instances in which eight infants produced neither action, raw scores were converted to percentages. Assumptions of parametric data were not met; therefore, we analyzed the data using nonparametric statistics. A Kruskal–Wallis test revealed no differences among the four toy orders in either the A–I or the I–A sequences. We therefore excluded toy order from further analyses.

To ensure that the observed effect was due to prosody and not lexical cues, we treated the pairing of lexical and prosodic cues in each condition as a significant counterbalancing variable. No infants spoke Greek and the words were therefore novel; however, we wanted to ensure that the lexical items did not convey semantic information through phonemic features. Therefore, we compared infants’ intentional actions in the first (Nato–Intentional; Ochi–Accidental) and second (Ochi–Intentional; Nato–Accidental) conditions. A Kruskal–Wallis test showed no significant difference between intentional actions in the first condition ($M = 75.59, SD = 25.95$) and second condition ($M = 78.57, SD = 21.21$), $\chi^2(1, N = 56) = .11, p = .74$, in the A–I order. Similarly, no differences were found between first condition ($M = 77.38, SD = 29.21$) and second condition ($M = 72.92, SD = 28.83$), $\chi^2(1, N = 56) = .48, p = .49$, in the I–A order. We conducted the same comparison for the accidental actions and found no significant difference between first condition ($M = 47.92, SD = 31.64$) and second condition ($M = 50.89, SD = 30.03$), $\chi^2(1, N = 56) = .08, p = .78$, in the A–I order, nor between first condition ($M = 48.51, SD = 34.47$) and second condition ($M = 51.48, SD = 33.80$), $\chi^2(1, N = 56) = .09, p = .77$, in the I–A order. We therefore collapsed the two conditions in further analyses.

A Wilcoxon signed-rank test (including all responses) was used to compare overall percentages of intentional actions to accidental actions. Infants produced more intentional actions ($M = 76.12, SD = 18.35$) than accidental actions ($M = 49.70, SD = 23.03$), $z = −4.42, p < .001$, in both the A–I ($z = −3.73, p < .001$) and the I–A ($z = −3.10, p < .002$) sequences. For means and standard deviations, see Table 2. To examine whether the effect was due to selective imitation of intentional actions only, further analyses compared intentional only responses to the other three response patterns (accidental action only, both actions in the same order as demonstrated, and both actions in the opposite order from that demonstrated). Infants produced more intentional actions than accidental actions in both the A–I ($z = −3.54, p < .001$) and the I–A ($z = −3.10, p < .002$) sequences. For means and standard deviations, see Table 3.

Further analyses examined age differences. Collapsing across conditions enabled us to conduct separate analyses for the two age groups (divided by a median split, with median age 16 months, 6 days). Older infants performed significantly more intentional only actions than accidental only actions (A–I sequence, $z = −3.53, p < .001$; I–A sequence, $z = −2.45, p < .05$), and more intentional actions than actions performed in the same order (A–I sequence, $z = −3.48, p < .001$; I–A sequence, $z = −4.23, p < .001$) and incorrect order (A–I sequence, $z = −4.03, p < .001$; I–A sequence, $z = −3.93, p < .001$). However, the younger age group showed no significant differences between intentional only and accidental only actions (A–I sequence, $z = −1.40, p = .16$; I–A sequence, $z = −1.91, p = .06$). There was however, a significant difference between intentional actions and actions performed in the same order (A–I sequence, $z = −2.24, p < .05$; I–A sequence, $z = −3.55, p < .001$) and in the incorrect order (A–I sequence, $z = −3.50, p < .001$; I–A sequence, $z = −2.28, p < .05$). See Table 4 for means and standard deviations. In addition to comparing performance for each age group separately, we compared performance between the two age groups. A Mann–Whitney test with intentional only and accidental only actions as the dependent variables and age as the independent variable showed no significant differences between intentional only and accidental only actions: A–I intentional only sequence, $U(54) = 291.00, z = −1.70, p = .09$; intentional only I–A sequence, $U(54) = 342.50, z = −.83, p = .41$; accidental only A–I sequence, $U(54) = 314.00, z = −1.36, p = .18$; accidental only I–A sequence, $U(54) = 385.50, z = −.12, p = .91$. Similar
nonsignificant differences were found between actions performed in order and in incorrect order: correct order, A–I sequence, $U(54) = 390.00, z = -.35, p = .74$; correct order, I–A sequence, $U(54) = 375.00, z = -.35, p = .73$; incorrect order, A–I sequence, $U(54) = 337.00, z = -1.26, p = .21$; incorrect order, I–A sequence, $U(54) = 303.00, z = -1.61, p = .11$.

7.3. Additional manipulation checks

As in Experiment 1, the experimenter maintained a neutral facial expression and neutral actions, and we again analyzed whether facial or action cues signaled intentionality. An independent coder viewed 14 videos with the volume turned off and rated whether the experimenter’s actions and facial expressions were intentional or accidental. The coder identified accidental and intentional actions at chance levels (53% correct). A chi-square analysis to reveal any associations between the correct intentional actions produced by infants and the actions identified correctly or incorrectly by the coder was nonsignificant, $\chi^2(1) = .94, p = .33$, suggesting there was no relation between the adult’s correct identifications and infants’ tendency to produce the intentional action. This result confirms it was not possible to identify intention on the basis of actions and/or facial expressions and, together with the effects of experimental condition, supports the conclusion that infants’ performance was determined by prosody.

We performed additional analyses to control for any practice effects throughout the eight trials. We first examined if there was a difference between infants’ performance on the first trials (trials 1, 3, 5, and 7) and on the second trials (trials 2, 4, 6 and 8) using a Friedman test. Older infants performed the intentional action 50% (SD = 23) of the time in the first trials and 59% (SD = 26) of the time in the second trials. This difference was not statistically significant for the older age group, $\chi^2(1) = 2.58, p = .11$. Younger infants performed the intentional action 41% (SD = 29) of the time in the first trials and 48% (SD = 29) of the time in the second trials. This difference was not statistically significant, $\chi^2(1) = .53, p = .47$. We used a Cochran’s Q-test to examine differences between the first and eighth trials. Older infants performed the intentional action 75% (SD = 44) of the time in the first trial and 67% (SD = 48) of the time in the eighth trial. This difference was not statistically significant, $Q(1) = 2.00, p = .74$, for the older age group. Younger infants performed the intentional action 44% (SD = 51) of the time in the first trial and 67% (SD = 48) of the time in the eighth trial. This difference was not statistically significant, $Q(1) = 2.00, p = .16$.

Because more infants participated in the second experiment, we were also able to conduct an analysis on the first and therefore uncontaminated trial. Using a Cochran’s Q-test, we compared infants’ performance of intentional actions, accidental actions, or both actions together on just the first of the eight trials. The test revealed no significant differences for the younger age group but significantly more intentional actions for the older age group, $Q(2) = 7.14, p < .05$. Response proportions for each
action type can be seen in Fig. 3. Wilcoxon tests for the older group showed a significant difference between intentional only and accidental only actions, $z = -2.13, p < .05$, and between intentional only and both actions, $z = -2.13, p < .05$.

7.4. Comparisons between Experiment 1 and Experiment 2

A Kruskal–Wallis test showed no significant difference between the percentage of intentional actions performed in Experiment 1 ($M = 72.62, SD = 17.91$) and Experiment 2 ($M = 76.12, SD = 18.35$), $\chi^2(1) = .64, p = .42$. Nor was there a difference for accidental actions in Experiment 1 ($M = 49.40, SD = 21.92$) and Experiment 2 ($M = 49.70, SD = 23.03$), $\chi^2(1) = .001, p = .97$.

7.5. Summary of results

To summarize, we found no difference in Experiment 2 between the two novel word conditions, and 14–18-month-olds produced, overall, more intentional than accidental actions when these two conditions were collapsed. Additionally, infants produced more intentional actions when comparisons were made between accidental actions only and both actions together. However, when the two age groups were analyzed separately, we found that older infants performed significantly more intentional than accidental actions but that the younger age group did not. Finally, no difference appeared between production of intentional and accidental actions in the two experiments.

8. Discussion

We compared infant imitation of actions marked prosodically as accidental or intentional. Infants were able to use prosodic cues to discriminate between the two, demonstrated by their selective imitation of intentional actions. This effect was more pronounced in older infants; older but not younger infants copied significantly more intentional than accidental actions, indicating that prosody is a valid cue to intentional actions from 16 months. This result is consistent with findings of other relevant studies; however, conclusions drawn from age comparisons should be interpreted with some caution.

Prosodic analyses indicated that intentional actions were marked by higher amplitude, higher mean pitch, longer duration, and a falling pitch contour. Accidental actions were marked by wider pitch range and a rising pitch contour. Thus, overall, the results of prosodic analyses in Experiment 2 were similar to the findings of Experiment 1.

9. General discussion

We conducted two experiments investigating the role of prosody in infants’ understanding of intention. In both experiments, infants saw demonstrations where an experimenter performed two actions on toys and marked those actions as intentional or accidental. In the first experiment, actions were
marked by the word “There” expressed with intentional prosody, or “Whoops” expressed with accidental prosody. In the second experiment, actions were marked with Greek words produced with intentional or accidental prosody. In Experiment 1, we aimed to replicate the methodology of the Carpenter et al. paradigm, and by combining acoustic analyses with behavioral analyses, to clarify the prosodic cues provided to infants. Results demonstrated that intentional and accidental exclamations had distinct prosodic contours and that infants utilized vocal cues to infer intentionality, indicated by the fact that they produced more intentional than accidental actions. Experiment 2 made use of this paradigm to investigate the effect of prosody on intention understanding more directly, using exclamations with the same prosodic contours, but controlling for lexical cues by using an unfamiliar language to mark actions as intentional or accidental. Infants were able to infer intentions from prosody even in the absence of lexical and facial cues; therefore, we conclude that infants infer intentions from prosody as early as 16 months.

A comparison between Experiment 1 and Experiment 2 further supports this claim by confirming that infants were equally likely to copy intentional actions in both studies, suggesting that infants do not need lexical information to infer intentions in this scenario. Previous studies had examined prosody as a secondary marker and had given primary emphasis to other cues, such as lexical cues (Carpenter et al., 1998) or a combination of lexical, facial and gestural cues (Olineck & Poulin-Dubois, 2005). We controlled for facial and action cues in both experiments, and in Experiment 2 we also controlled for lexical cues, allowing us to examine the effect of prosodic cues more specifically. The absence of differences between the two experiments supports the conclusion that infants do not require lexical and facial cues to make attributions about intentions.

Both experiments contribute to the expanding literature on the importance of prosody in infancy. Prosody is an important aspect of communication that infants rely on in their interactions with other people. From an early age, infants seem sensitive to the tone of voice in communication. In the first months of life, infants prefer comforting vocalizations to approving or directive vocalizations, but those preferences change with development (Kitamura & Lam, 2009). By 12 months, they use vocalizations as a guide to acting on the world, showing a greater willingness to cross a visual cliff in response to encouraging vocalizations (Vaish & Striano, 2004). Furthermore, studies have suggested that prosody may aid language acquisition (Johnson & Seidl, 2008; Thiessen, Hill, & Saffran, 2005). The infants we tested are at the stage in development where lexical items are becoming increasingly important, and they may pay particular attention to prosody as a significant cue to the meaning of lexical items.

The present study suggests that by 16 months, infants’ sensitivity to prosodic cues develops into a more sophisticated ability – the ability to gather information about what others intend and do not intend in their own interactions with inanimate objects. Infants are able to use prosodic cues as a guide to how to act on the world, demonstrated by their tendency to copy intentional actions more than accidental actions. Results of the second experiment demonstrate that acoustic features of speech accompanying actions allow infants to identify intentionality in perceptually similar actions. It is also important to note that infants did not make this distinction based on learning through the study, as our analyses show that performance was no better in the last trials. The significance of these two experiments lies in the fact that prosody by itself can provide infants with the necessary information about another person’s mental state.

Our age-related results are consistent with recent findings demonstrating a developmental progression in infants’ ability to differentiate between purposeful and non-purposeful mental states. Infants first begin to show an understanding of mental states, such as simple goals and intentions, and imitate according to this understanding from approximately 12 months of age, but it is not until later in the second year that they demonstrate an understanding of the distinction between accidental and purposeful behavior, as measured by imitation (Bellagamba et al., 2006; Carpenter et al., 2005; Nielsen, 2009; Olineck & Poulin-Dubois, 2005).

Our findings may thus reflect developments in intention understanding generally, in the use of intentions to guide imitative behavior, in infants’ ability to use a new set of prosodic cues to distinguish meaning, and in this case to distinguish intentions from accidents. Recent studies using looking time measures indicate that infants can infer intentions from prosodic cues earlier in infancy (Sakkalou & Gattis, n.d.), suggesting that the changes we report may be tied specifically to the use of intentions to guide imitative behavior.
A recent study by Olineck and Poulin-Dubois (2009) examined relations between two mental state understanding tasks, a behavioral re-enactment task (Meltzoff, 1995) and a selective imitation task (similar to the one conducted in this study). No significant relation between these two tasks was found. The authors proposed that the behavioral re-enactment task may have measured intention understanding at a behavioral level, whereas the selective imitation task may measure intention at a more abstract level (Olineck & Poulin-Dubois, 2009). Interestingly, when autistic children were tested on these two tasks, they were able to complete the behavioral re-enactment task (Aldridge, Stone, Sweeney, & Bower, 2000; Carpenter, Pennington, & Rogers, 2001) but tended to copy a similar number of intentional and accidental actions on the selective imitation task (D’Entremont & Yazbek, 2007). Autistic children may have performed poorly on the selective imitation task because they failed to take into account socio-pragmatic cues associated with the actions, such as facial expressions and mannerisms of the experimenter. The results of these studies, together with the results reported here, support the claim that distinguishing intentions from accidents requires attribution of mental states from socio-pragmatic cues, such as prosody.

An alternative account is that, rather than using complex prosodic patterns to infer intentions, infants may attend to a subset of cues, in particular the contrast between positive and negative vocalizations, or the contrast between rising and falling prosodic contours. Prosodic analyses of vocalizations in our study indicate, however, that the prosodic features do not map neatly onto earlier contrasts such as encouragement and prohibition: Previous research has reported that mothers use rising contours to engage or encourage attention to an object and falling contours to prohibit. In both experiments reported here, however, accidental vocalizations had rising pitch contours and intentional vocalizations had falling pitch contours. In addition, multiple studies have found that very young infants are sensitive to positive–negative contrasts (Fernald & Kuhl, 1987; Kitamura & Lam, 2009). If infants were simply demonstrating a preference for positive over negative vocalizations, we would expect younger infants to perform as well as older infants. Instead, age-related analyses for Experiment 2 showed that the tendency to copy intentional actions was stronger for older infants, indicating that response performance relied on a more complex set of cues. Further studies are needed to examine the role of contrast in prosodic processing throughout infancy more generally and to learn whether infants infer other mental states beyond intention from prosodic cues.

The question of what specific cues infants use to discriminate mental states relates to a larger one about the universality of the information available from prosody. Fernald (1993) reported that the prosodic signature of different communicative messages such as prohibition and approval is similar in different European languages, although attenuated in Japanese. More recently, Bryant and Barrett (2007) presented similar communicative messages spoken in English to a non-English–speaking population from a South American indigenous culture. They reported that participants could use prosodic cues to distinguish between prohibition, approval, attention, and comfort, especially when presented in infant-directed speech. Future research might use similar methods to investigate the universality of prosodic cues to intention.

Together with past research, our results suggest that the sensitivity to prosodic information evident in early infancy continues to develop over a longer period. By 16 months, prosodic sensitivity combines with intention understanding to yield new insights into the mental states of others. Based on our results and previous studies showing that prosody is important in highlighting crucial words (Fernald & Mazzie, 1991; Grassmann & Tomasello, 2007), we propose that the developmental role of prosodic sensitivity, combined with infants’ developing understanding of intentions, may extend far beyond imitative behavior, introducing infants to the world of communication.

References


