Contextual influences on children’s use of vocal affect cues during referential interpretation

Jared M. J. Berman¹, Susan A. Graham¹, and Craig G. Chambers²
¹Department of Psychology, University of Calgary, Calgary, AB, Canada
²University of Toronto, Toronto, ON, Canada

In three experiments, we investigated 5-year-olds’ sensitivity to speaker vocal affect during referential interpretation in cases where the indeterminacy is or is not resolved by speech information. In Experiment 1, analyses of eye gaze patterns and pointing behaviours indicated that 5-year-olds used vocal affect cues at the point where an ambiguous description was encountered. In Experiments 2 and 3, we used unambiguous situations to investigate how the referential context influences the ability to use affect cues earlier in the utterance. Here, we found a differential use of speaker vocal affect whereby 5-year-olds’ referential hypotheses were influenced by negative vocal affect cues in advance of the noun, but not by positive affect cues. Together, our findings reveal how 5-year-olds use a speaker’s vocal affect to identify potential referents in different contextual situations and also suggest that children may be more attuned to negative vocal affect than positive vocal affect, particularly early in an utterance.

Keywords: Vocal affect; Sentence processing; Reference.

The successful interpretation of referential expressions is typically facilitated by the use of contextual information. This is in part because the lexical content of a given expression is rarely sufficient to allow a speaker’s intended referential meaning to be identified. In many cases, helpful cues can be found elsewhere in the broader utterance. These can include cues that are linguistic in character, such as the semantic and presuppositional constraints of predicate terms, (Altmann & Kamide, 1999; Chambers & San Juan, 2008), morphological and syntactic markers (Kamide, Altmann, & Scheepers, 2003), and linguistic prosody (e.g., Dahan, Tanenhaus, & Chambers, 2002; Ito & Speer, 2008; Weber, Braun, & Crocker, 2006). Paralinguistic information in the speech stream can also influence the interpretation of language. This information includes so-called indexical cues.
carried in a speaker’s voice patterns that convey the age, gender, and specific identity of a speaker (e.g., Creel & Bregman, 2011; van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008), as well as more general cues that vary across situations, such as those that reflect the speaker’s emotional disposition at a given moment (e.g., Berman, Chambers, & Graham, 2010; Nygaard & Lunders, 2002).

One important aspect of paralinguistic information stems from the fact that it is often present throughout an entire utterance or string of utterances and is not isolated to particular words or phrases. This creates the potential for these cues to exert an influence well before an indeterminate or ambiguous expression is actually encountered, provided of course, that this information is deemed relevant by language comprehension mechanisms. In the current study, we examined preschoolers’ real-time sensitivity to the emotion-related speech cues conveyed in spoken utterances across different contextual circumstances. Our primary goal was to examine the uptake of these cues in situations when these cues could play a comparatively stronger or weaker role and to explore the time point at which these cues begin to influence interpretation. Throughout the paper we have used the term “vocal affect” to refer to differences in the primary acoustic–phonetic features that speakers use to convey an emotional meaning including variations in pitch level, pitch contours, and speech rate (see Banse & Scherer, 1996; Frick, 1985). These features can signal a speaker’s momentary emotional disposition towards objects and events and do not necessarily reflect a more enduring emotional state experienced by the speaker. This information can therefore provide relevant cues that accompany the linguistic content in speakers’ utterances about objects and events. In Experiment 1, we examined how and when 5-year-old children use speaker vocal affect in contextual situations involving referential ambiguity. In Experiments 2 and 3, we investigated children’s use of these affect cues in referentially unambiguous contexts and also explored potential differences related to the valence of the affect cue (e.g., happy-vs. sad-sounding speech).

As background, it is relevant to consider how adults use vocal affect cues in the course of language comprehension. Studies have demonstrated that listeners can readily identify and use these cues to guide aspects of language processing. For example, Nygaard and Lunders (2002) presented adult participants with homophones (e.g., die/dye) where one of the two forms associated with the sound pattern had an affectively charged meaning (i.e., die), and the other was neutral (e.g., dye). The recorded words were presented in one of three vocal affect conditions: positive (happy-sounding), negative (sad-sounding), or neutral. Participants were asked to transcribe the words they heard. Results suggested that vocal affect had a significant effect on which version of the homophone adults transcribed. More specifically, adults were more likely to transcribe die than dye when the word was spoken using negative vocal affect. Adults’ sensitivity is also apparent from match/mismatch paradigms using unambiguous language. For example, Paulmann, Titone, and Pell (2012) demonstrated that adults are quicker to shift their gaze to a face depicting a particular emotion when the accompanying speech affect was congruent (e.g., “Click on the happy face,” spoken with positive affect) than when it was incongruent.

Given that language interpretation and emotion perception are comparatively distinct processes, it is interesting to consider how the apparently smooth integration observed in adults develops during the course of childhood. In fact, the results of research examining children’s sensitivity to vocal affect have illustrated various ways in which children differ from adults, depending on the age of the children and the specific task used. For example, when the lexical content of a request was incongruent with vocal affect, Friend (2001) found that children as young as 15 months were more likely to regulate their behaviour in response to the vocal affect of the request. However, as children reach preschool age, there is evidence that they become more likely to prioritize the linguistic content of the sentence over the speaker’s vocal affect in certain kinds of tasks. For example, Morton and Trehub (2001) presented 4- to 7-year-old children with linguistic
information that was incongruent with vocal affect information (e.g., “My dog ran away from home” spoken with positive vocal affect) and found that children relied almost exclusively on the content of the sentence to judge a speaker’s emotional state. Interestingly, children’s lack of attention to paralinguistic information in judging the speaker’s emotional state was not simply a failure to understand and correctly categorize the speaker’s vocal affect from the relevant acoustic cues. When sentences were spoken in a foreign language or when speech was low-pass filtered to reduce the perceptibility of the linguistic information, children were more likely to judge a speaker’s emotional state from the affect cues. Further, in contrast to the preschoolers, adults relied exclusively on the speaker’s vocal affect, rather than the linguistic content of the sentence, to judge the speaker’s current emotional state (Morton & Trehub, 2001).

The apparent insensitivity to vocal affect in clear speech observed by Morton and Trehub (2001) in 4- to 7-year-old children is somewhat surprising given that children as young as 15 months have been observed to regulate their behaviour based on this cue (Friend, 2001). One question is whether the task demands involved in judgement paradigms risk underestimating children’s sensitivity to a speaker’s vocal affect. Indeed, more indirect measures from the Morton and Trehub experiments (response latency) suggest that children are, in fact, somewhat sensitive to a speaker’s vocal affect. Further, more recent work using tasks that do not involve incongruity or explicit judgements indicates that children as young as 4 years old are sensitive to a speaker’s vocal affect as a cue to referential intent. Using eye movements as an online measure of comprehension, Berman et al. (2010) examined preschoolers’ use of vocal affect to disambiguate ambiguous utterances. In this study, 3- and 4-year-old children were presented with an array of photos. On critical trials, the array included pictures of two familiar objects of the same kind (e.g., an intact ball and a partially deflated ball) as well as another object of a different kind. Children were then asked to find a particular referent using an ambiguous utterance (e.g., “Look at the ball. . . . Point to the ball”) that varied in speaker affect: positive, negative, or neutral. Eye fixation patterns measured during the ambiguous noun in the first sentence reflected a pattern whereby children considered the “broken” referential candidate (e.g., the deflated ball) most often when it was paired with negative speaker affect, less often with neutral speaker affect, and finally even less often with positive speaker affect. Interestingly, although 4-year-old children’s looking patterns indicated an appreciation for vocal affect, their pointing behaviour did not. Furthermore, when 3-year-old children completed the same task, there was little evidence for an appreciation of vocal affect even using the more implicit measures based on eye gaze. Thus, the ability to use speaker vocal affect to resolve ambiguity is evident yet still at an emergent stage during the later preschool years.

Although the Berman et al. (2010) results clearly demonstrate children’s nascent ability to coordinate vocal affect with linguistic information, aspects of the findings raise additional questions about the core mechanisms underlying this ability in young children. Specifically, the influence of vocal affect information on four-year-olds’ consideration of alternative referents was detected only at the point where the final noun was encountered (“Look at the ball”) and not earlier in the sentence. This is surprising because, as stated earlier, vocal affect cues are distributed across the entire utterance and, in principle, could begin to guide expectations at an earlier point. One possible explanation for this result is that children require a speech sample of a sufficient temporal duration in order to correctly identify vocal affect. Consequently, the use of these cues would be more evident at the end of the sentence, where the critical noun happened to be located. A second explanation is that children’s use of affect cues is essentially strategic and was observed at the noun because of the apparent ambiguity entailed by this expression. A third yet related explanation hinges on the fact that the scenario used to create referential ambiguity was one in which two of the three display objects belonged to the same category. The two same-category objects differed in terms of their properties, thereby entailing that two different ontological
criteria were required to conceptually differentiate all three objects (namely, information about conceptual kinds and situation-specific properties). It is possible that, when presented with this type of context, children experience greater cognitive load and are consequently less able to quickly react to affect cues in vocal paralanguage.

The issues involved in these explanations are important because they highlight specific ways in which children’s information-integration abilities might differ from those of adults. For example, studies of spoken language comprehension in adults have highlighted the apparently automatic and opportunistic uptake of many different kinds of informational cues, sometimes creating a referential “garden-path” situation when these cues point to an interpretation that proves to be incorrect with subsequent utterance information (e.g., Dahan et al., 2002; Heller & Chambers, 2011; Kukona & Tabor, 2011). The idea that paralinguistic cues are treated as strategic resources, used only on demand, would in turn suggest that there is not only a sensitivity issue but also a significant shift in how this information is used at some point in development.

The present set of studies directly addressed these issues by examining how children use vocal affect to interpret utterances in different referential scenarios and whether this information can begin to influence interpretation before a referring expression is heard. We also explore potential differences related to the valence of the affect cues—namely, whether happy- and sad-sounding speech have the same effects on referential processes. Given the evidence for 4-year-olds’ somewhat fragile ability to coordinate speech affect and linguistic information, the current investigation focuses on 5-year-olds. Our guiding assumption was that this age group should show a more robust sensitivity to affect cues, thereby providing a better basis to make experimental comparisons across different contextual situations. This assumption was directly tested in our first experiment by examining whether, unlike 4-year-olds, 5-year-old children show clear evidence of sensitivity to speaker affect in their overt pointing behaviour as well as in their eye gaze patterns. This experiment also provided an important opportunity to further replicate the finding that affect cues have little effect before an utterance-final referential expression was heard and when the context highlights two alternative members of the same category.

Experiments 2 and 3 built on the results from the first experiment and involved contextual situations where there was only one exemplar of each category. This both changed the visual scene and also effectively eliminated the need to rely on affect cues to identify the intended referent (because the referential description would be unambiguous). If children’s referential interpretation continues to show the use of vocal affect information in this situation, we can conclude, for example, that ambiguity is not a necessary trigger for this to occur. Further, a comparison across these two experiments allowed us to examine the possibility that different affect valences might vary in terms of how strongly they can influence aspects of real-time referential interpretation.

**EXPERIMENT 1**

The overarching goal of Experiment 1 was twofold: First, to investigate 5-year-olds’ use of a speaker’s vocal affect to identify a referential candidate when presented with a linguistically indeterminate utterance; and second, to establish whether this sensitivity is reflected in both eye gaze patterns and explicit referential decisions.

On critical trials, children were presented with arrays consisting of two objects of the same kind (i.e., an intact doll and a broken doll) and a third unrelated object (see example in Figure 1). A recorded instruction, using one of three different types of vocal affect, asked children to identify one of the same-kind referents (e.g., “Look at the doll... Point to the doll”).

If 5-year-olds can use vocal affect cues encountered early in the instruction to begin anticipating the intended referent, we should find a difference in looking behaviour before the noun in the first sentence is actually heard, contingent on the speaker’s vocal affect. Given the object array shown in the figure, the broken object provides the most sensitive test case for evaluating this effect because of the difference in its physical status compared to the
other two alternatives. If children use vocal affect to anticipate referential candidates, we should find a pattern whereby children would be more likely to shift their gaze to the broken object in the negative affect condition, less often in the ambiguous neutral condition, and least often in the positive affect condition. In contrast, if the results with 5-year-olds are similar to those observed with 4-year-olds, their fixation patterns should not differ in the prenoun interval based on vocal affect. Rather, sensitivity to the affect cues should only be evident only upon hearing the noun. In addition, if children at this age have a comparatively better ability to integrate vocal affect with language than 4-year-olds, their overt pointing behaviour should mirror the patterns observed in their eye fixations.

Method
Participants
The final sample consisted of fifteen 5-year-olds (9 males; \( M = 5.46 \) years, \( SD = 0.24 \) years), recruited through advertisements within the community. Children were primarily Caucasian, from socioeconomic backgrounds that varied broadly within the more general middle class (although the latter was not formally assessed), and from homes in which English was the primary language spoken (2 children were from homes in which French was also spoken). Five additional children were tested but were removed from the final sample due to insufficient data collected from the eye tracker.

Stimuli
Critical trials. For the critical trials, an array of three images was presented on a large display screen, accompanied by a recorded instruction relating to one of the objects (e.g., “Look at the ball. … Point to the ball”). The images were cropped photographs of real-world objects. Two of the images belonged to the same category but differed in terms of their likelihood to be associated with positive or negative affect (e.g., an intact vs. deflated ball). In addition to objects varying along
the broken/intact dimension, other objects were altered to be either dirty or clean (e.g., a clean vs. dirty stuffed animal). For simplicity, however, we use the terms “intact” and “broken” throughout to differentiate the objects on critical trials that are paired with positive versus negative vocal affect. The third image was always an unrelated distractor object (e.g., toy star; see Figure 1). Distractor objects were included in the display for two reasons. First, they reduced the salience of the physical property differences (e.g., broken vs. intact) between the other two objects present on critical trials. Second, the possibility that the instruction might refer to this object helped draw attention away from the other two objects.

The full set of stimuli was the same as that used in Berman et al. (2010). Three versions of each critical instruction were recorded by a female native speaker of English, differing only in the type of emotional affect conveyed by her voice. The neutral utterances were recorded using neutral-sounding speech, whereas negative and positive utterances were recorded with distinctly sad- or happy-sounding speech, respectively. To ensure that the recorded utterances conveyed the appropriate vocal affect, a pretest was previously conducted in which 12 adults were asked to rate the recorded utterances on a scale ranging from 1 (negative-sounding) to 7 (positive-sounding), with 4 as the midpoint. Raters listened to utterances in a randomized order with a brief sequence of piano tones played between each utterance to reduce carryover or contrast effects. The mean rating scores confirmed that perceived vocal affect was significantly different for the three utterance types (negative: \(M = 2.11\), range: 1.42–2.75; neutral: \(M = 3.29\), range: 2.55–4.36; positive: \(M = 6.13\), range: 5.46–6.58, all \(p < .01\)).

The pairing of object arrays to vocal affect type was cycled across participants such that each array occurred only once in each affect condition. Further, each array was paired with each affect type the same number of times across the experiment (i.e., 3 positive affect trials paired with the doll array, 3 neutral affect trials paired with the doll array, and 3 negative affect trials paired with the doll array).

**Filler trials.** In addition to the critical trials, 12 filler instructions were recorded with neutral vocal affect. The fillers were included to prevent participants from developing specific expectations about the instructions and the objects based on the critical trials. Six filler trials had displays depicting three distinct object types (e.g., a rattle, a unicorn, and a toy car). These trials were used to prevent children from expecting that referring expressions would always be ambiguous and that the photo arrays would always contain a pair of objects from the same category (e.g., two balls). Furthermore, 6 filler trials contained instructions that referred to an object (e.g., an elephant) accompanied by two other unmentioned objects from the same category differing along a dimension that was not used to distinguish objects on critical trials (e.g., a green and yellow bowl). These fillers were included to break any expectation that an utterance would always refer to a member of a same-category pair when such a pair was present and also neutralized any expectation that affective criteria could be used to differentiate same-category objects.

The full set of 24 trials was presented in a computer-controlled random order, and the position of photo objects within each display was also randomized.

**Apparatus**

Children’s eye fixation position was tracked using a Tobii x50 eye tracker placed below a 46-inch computer monitor. The x50 has an accuracy of between 0.5–0.7 degrees of visual angle and allows for some freedom of head movement. Areas of interest were identified for each of the photos in order to establish the screen region the child was fixating at successive time points. Gaze data were logged by the recording software every 20 ms, and a fixation was counted if the child gazed at the same image for more than 95 ms. The auditory stimuli were presented from a set of speakers located directly behind the monitor. In addition to the eye-tracking equipment, a HD camera was positioned behind the child in order to record his or her pointing behaviour.

The experiment was implemented using E-Prime software with Tobii extensions. At the
beginning of each session, the child’s point of gaze was calibrated using Clearview software. Only data from those children who showed accurate calibration on 3 out of 5 test fixation points were included, although for most children calibration was perfect.

Procedure
Testing took place in a quiet room. Children sat on a small chair facing the computer monitor, approximately 1.4 m away. The experiment began with the calibration procedure. Once calibration was complete, the experimenter started the main experiment. Each trial proceeded as follows: First, a photo array was presented on the screen for 3 s without auditory stimuli, followed by a blank black screen for 2 s. Next, the same photo array reappeared on the screen, accompanied by an instruction referring to one of the display objects (e.g., Look at the doll. Point to the doll). After the child pointed, the experimenter advanced the program to the next trial.

Total testing time was approximately 10 min including the calibration procedure and the main experiment. Following the completion of the task, parents were debriefed about the goal and design of the study. All children received a small toy, a T-shirt, and a “Child Scientist” certificate for their participation.

Results

Eye fixation patterns
The eye fixation measures allowed us to assess how children’s sensitivity to a speaker’s vocal affect changed across the first sentence in the recorded instructions. We identified two time intervals for analysis. First, a prenoun interval of 680 ms was defined, beginning 480 ms before the onset of the noun and ending 200 after noun onset. This interval is based on the average duration of the speech in the instruction leading up to the noun (i.e., “Look at the”), plus a 200-ms margin added to each boundary that reflects the typical lag for the eyes to react to auditory information in this experimental paradigm. Within this interval, only paralinguistic information from the vocal affect cues in the utterance was available to assist children in identifying a potential target object. The second interval was a 1,000-ms period beginning 200 ms after noun onset. Although the average noun duration across all three affect conditions was 676 ms, the longest noun was 923 ms in duration. Thus, the 1,000-ms interval allowed us to capture both the integration of the vocal affect cues with the semantic information in the unfolding noun for all noun exemplars. Of course, for some exemplars, this interval included a period after noun offset. However, because the noun was in sentence-final position, there was no concern that adding additional time for some nouns would entail an overlap with the processing of subsequent sentence information.

Prenoun interval. Eye fixation patterns within this interval should reveal whether and how the vocal affect cues that are broadly distributed throughout the speech signal can be used by children in advance of hearing the referring noun. For measurement purposes, we focus on the single “odd-man-out” object whose state contrasts with the other two in terms of its relationship to affect information (in this case the broken item). The potential to individuate this object relative to the other display items should provide the most sensitive and direct index of the influence of affect cues across conditions.

We first calculated the average proportion of fixations to the broken object within a 100-ms time interval centred on the beginning of the prenoun interval and within an interval of the same duration centred on the end of the prenoun interval. These values are depicted in the top panel of Figure 2. A slope score representing the change in fixation proportions from the beginning to the end of the interval was then calculated. To do this, we divided the difference between the endpoint measures by the length of the prenoun region. The values were then rescaled from a change per-millisecond measure to a per-second measure for interpretability. Note that this measure differs from the target advantage score used in the analyses of the noun region that follow because the available linguistic information has not yet provided any reason to focus on the eventually named target.
If children are using affect cues to begin isolating likely referents in advance of the noun, fixations to the odd-man-out (broken object) should show the strongest increases over the time interval in the negative affect condition, fewer increases in the neutral condition, and still fewer in the positive affect condition. These scores did not, however, reflect this pattern and instead showed a comparatively random pattern of slope scores accompanied by high variability (negative: $M = 0.17$, $SD = 0.50$; neutral: $M = 0.06$, $SD = 0.23$; positive: $M = 0.41$, $SD = 0.50$). These data were submitted to an analysis of variance (ANOVA) model with a 1-degree-of-freedom within-participants factor testing for the predicted pattern of linear trend among vocal affect conditions (negative > neutral > positive). Note that this analytic strategy, as opposed to an omnibus analysis, is recommended when specific a priori hypotheses are tested in models with an ordinal factor (Hertzog & Rovine, 1985). The analysis confirmed that the predicted linear trend was not present, $F(1, 14) = 1.17$, $\eta^2_p = .08$, $p = .30$. These

![Figure 2. Change in the mean proportion of fixations to the odd-man-out object in the prenoun interval across vocal affect conditions. To view a colour version of this figure, please see the online issue of the Journal.](image)
results suggest that, in this contextual situation, vocal affect cues encountered early in the utterance have little effect on referential expectations.

**Noun interval.** Next we examined children’s eye fixation patterns as the ambiguous noun was heard. Within this interval, the core question is the extent to which the target and the “competitor” object (i.e., the other linguistically compatible referent for the noun) are distinguished from one another based on the affect cues. Because the eye-tracking methodology can reveal the influence of contextual information on the interpretation of the noun in either the early or late moments of processing, we plot eye fixation proportions every 20 ms across the noun interval.

Figure 3 presents the average proportion of fixations to the three objects across the noun interval for each of the affect conditions. Fixations initiated before the beginning of this interval were excluded from analysis to ensure that observed eye movement behaviours could plausibly be associated with the interpretation of noun information (and its potential integration with paralinguistic information), rather than reflecting a continued bias to fixate a particular object that attracted attention at an earlier time point. For this reason, fixation proportions rise from zero at the beginning of the time interval. The negative affect condition (top panel) shows children’s growing tendency to fixate the broken referent when compared to the intact object at approximately 680 ms after the onset of the noun. The stronger appreciation for the broken referent is reduced in the neutral affect condition (middle panel), and the trend completely reverses in the positive affect condition (bottom panel), where children are more likely to fixate the intact object around 550 ms after noun onset.

To provide a statistical analysis of how children’s looking behaviours were modulated by speaker affect as the noun was processed, we calculated a target advantage score (see Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Heller, Grodner, & Tanenhaus, 2008; Tsang & Chambers, 2011) reflecting the relative tendency to fixate the broken referent over the intact object within the depicted speech interval. This measure captures differences across conditions to differentiate the named target from a meaningful alternative. This was calculated by subtracting the proportion of fixations to the broken referent from the proportion of fixations to the intact referent for each participant and each vocal affect condition, within the depicted time interval (negative: $M = 0.18$, $SD = 0.46$; neutral: $M = 0.07$, $SD = 0.54$; positive: $M = -0.18$, $SD = 0.46$). These scores were then submitted to the repeated measures ANOVA model described earlier, which tests for a linear (decreasing) pattern across the three conditions. This analysis yielded a significant linear effect of vocal affect (negative > neutral > positive), $F(1, 14) = 6.23$, $\eta^2_p = .31$, $p < .05$, confirming that children were consistently more likely to look at the broken referent instead of the intact one as speech became increasingly negative or sad sounding.

To summarize, in contrast to the absence of an effect in the earlier portion of the sentence, eye fixation patterns in the noun interval clearly demonstrate that children’s real-time identification of the referent for the ambiguous noun was guided by accompanying affect cues in the speech stream.

**Pointing behaviours**

Children’s pointing behaviour was coded as a measure of their explicit ability to detect and use vocal affect cues to understand a speaker’s referential intentions after the entire sentence was heard. The experimenter and a trained research assistant coded all of the child’s points from the videotapes while unaware of the specific trial being presented. A second assistant recoded 20% of the data ($n = 3$) to establish interrater reliability. Interrater reliability was excellent (Cohen’s kappa $= .96$, $p < .001$).

Table 1 shows the percentage of points to the two referential alternatives for the ambiguous noun as a function of affect type. Children never pointed to the distractor object, indicating that they understood the instructions. The average number of points to the broken object across conditions was used as the dependent measure, and these data were again submitted to a repeated measures ANOVA testing for the predicted pattern of linear trend (positive-
sounding < neutral < negative-sounding), for which the outcome was significant, \( F(1, 14) = 16.51, \eta_p^2 = .541, p = .001 \). Specifically, 5-year-old children were correspondingly more likely to point to the broken object as the speaker’s vocal affect became increasingly negative sounding.

**Discussion**

The results confirm that 5-year-olds can use vocal affect as a cue to a speaker’s referential intentions both at an implicit level, as evidenced by the eye gaze patterns occurring as the noun was heard,
and at an explicit level, as evidenced by children’s pointing behaviour after the entire utterance was heard. The fact that this sensitivity was found in children’s explicit judgements about the intended referent contrasts with previous research on 4-year-olds (e.g., Berman et al., 2010), where the sensitivity to affect cues was found only in eye fixation measures. The convergence of 5-year-olds’ eye fixation and pointing behaviour confirms that children at this age are more sophisticated in terms of their use of paralinguistic cues and are therefore well suited for exploring the more subtle aspects of how affect cues are integrated with speech information in the experiments that follow. An important similarity with the findings for 4-year-olds, however, is that sensitivity to speech affect was evident only as the noun was encountered. With this in mind, we return to the possible explanations for this effect we described earlier. One possibility is that children might need to hear a certain amount of speech before the affect cue can be reliably recognized and used. Alternatively, the effect could hinge on the referential scenario, which involves the presence of two same-category objects and an ambiguous noun, whose indeterminacy might serve as the trigger to draw on contextual information such as paralinguistic cues.

To explore the potential influence of the contextual scenario, we use unambiguous situations in Experiments 2 and 3 in which there was only one candidate referent of each kind present, and in which an unambiguous description was used. If the comparatively late use of affect cues found in previous work and Experiment 1 is related to the ambiguous scenario, children may be more likely to show earlier sensitivity to vocal affect in this context. This sensitivity can be detected by measuring the extent to which children show evidence of being “garden-pathed” in cases where the speech affect information is misleading in relation to the object eventually denoted by the sentence-final noun. For example, negative speech affect could create an expectation for reference to a broken object, but when the final noun denotes an intact object, the child’s referential hypotheses would be revealed to be incorrect. Because the linguistically unambiguous situation entails that there is a definitive target object, the experimental design is such that the target for a given display can only have physical properties that match one of the endpoints on the affect scale (i.e., intact or broken). In Experiment 2, the critical trials involve an intact target, another intact object, and a contrasting broken object. In Experiment 3, the situation is reversed, with a broken target, another broken object, and a contrasting intact object. By including both of these designs, it is possible to investigate potential differences in how the association of objects with positive versus negative affect can influence children’s real-time referential hypotheses.

EXPERIMENT 2

In Experiment 2, the arrays on critical trials consisted of one broken distractor object (e.g., a broken cell phone) and two intact objects of different kinds (e.g., an intact duck and an intact ball), one of which was the target referent of the sentence. See Figure 4 for a sample array. As in Experiment 1, children heard a recorded instruction to find a referent (e.g., “Look at the ball”), and the instruction varied in terms of three different vocal affects (happy-sounding, neutral, sad-sounding). First, if explicit linguistic ambiguity is necessary to trigger the use of affect cues, we should find no influence of vocal affect information. If, however, the influence of affect cues continues to be observed only at the utterance-final noun (as in Experiment 1), the overall pattern would suggest that children require a certain-sized sample of speech before

| Vocal affect condition | Broken   | Intact   |
|-----------------------|---------|---------|
| Negative              | 65.0 (10.3) | 35.0 (10.3) |
| Neutral               | 41.7 (10.8) | 58.3 (10.8) |
| Positive              | 23.3 (6.0)  | 76.7 (6.0)  |

Note: Standard errors in parentheses.

**Table 1. Experiment 1: Mean percentage of points to each of the referents as function of affect type**
these cues are correctly identified and applied. Finally, if the change in the visual scenario leads children to be more sensitive to affect cues in the unfolding utterance, earlier influences of vocal affect should be detected. Specifically, with the current design, the single “odd man out” (i.e., the broken cell phone in Figure 4) should be incrementally differentiated from the remaining display objects when the paralinguistic cues convey negative affect, less differentiated with neutral affect, and least with the positive affect.

Method
Participants
Fifteen 5-year-olds (7 males; $M = 5.56$ years, $SD = 0.30$ years), recruited through advertisements within the community, were included in the final sample. As in Experiment 1, children were primarily Caucasian, from socioeconomic backgrounds that varied broadly within the more general middle class (although the latter was not formally assessed), and from homes in which English was the primary language spoken (1 child was from a home where French was also spoken). Two additional children were tested but were removed from the final sample due to insufficient data collected from the eye-tracker.

Stimuli
Critical trials. On critical trials, an array of three images was presented on a large display screen, accompanied by a recorded instruction relating to one of the objects (e.g., “Look at the ball. . . . Point to the ball”). The recorded instructions for critical trials were the same as those used in Experiment 1. The images were cropped photographs of real-world objects. Two of the objects (one of which was the target) were intact, and one of the objects was broken/dirty. Critically, unlike Experiment 1, only the target object belonged to the category named in the referring expression (see Figure 4). The pairing of object arrays to vocal affect type was cycled across participants such that an individual participant encountered a given array only once. However, across participants, each array occurred equally often in each of the affect conditions.

Filler trials. In addition to the critical sentences, 12 filler instructions were recorded using neutral vocal affect. The fillers were included to prevent participants from developing specific expectations.
regarding the types of objects shown. There were three types of filler trials: four trials with displays depicting three distinct object types (e.g., a rattle, a unicorn, and a toy car) designed to break a possible expectation that all trials would have one broken object; four trials with displays depicting three distinct object types (e.g., a tree, a sword, and a giraffe), all broken, again creating variety in the extent to which objects could be differentiated from one another by their properties; and four trials with two intact objects and one broken object, with reference made to the broken object, designed to break the expectation that trials with two intact and one broken object would always refer to an intact object.

The full set of 24 trials was presented in a computer-controlled random order, and the position of photo objects within each display was also randomized.

**Apparatus and procedure**
The apparatus and procedure were identical to those used in Experiment 1.

**Results**

**Eye fixation patterns**

**Prenoun interval.** As in Experiment 1, we conducted an analysis comparing the change in the likelihood of fixating the “odd-man-out” in the object arrays (e.g., the broken phone in Figure 4) at the beginning and the end of the prenoun interval. Average fixation proportions at the beginning and end are shown in Figure 2, middle panel. As before, the slope measures from each affect condition reflecting the change over the prenoun interval were submitted to a within-participants ANOVA testing for the predicted pattern of linear trend among vocal affect conditions. The effect was significant, $F(1, 14) = 11.18$, $\eta_p^2 = .44$, $p = .005$, indicating that the predicted linear trend was present. Specifically, increases in the likelihood of fixating the broken object over the span of the prenoun interval were greatest in the negative vocal affect condition, less in the neutral condition, and least in the positive condition (negative: $M = 0.27$, $SD = 0.23$; neutral: $M = 0.09$, $SD = 0.35$; positive: $M = -0.36$, $SD = 0.43$). This result suggests that, when presented with situations that do not involve two referents from the same category, children were able to rapidly use a speaker’s vocal affect to anticipate a potential referent.

**Noun interval.** We used the same noun interval as that used in Experiment 1, beginning 200 ms after noun onset and ending at 1,200 ms after noun onset. As before, fixations initiated before the beginning of this interval were excluded from analysis to ensure that the fixation proportions reflect the integration of noun information. Figure 5 presents the average proportion of fixations to the three objects across the noun interval for each of the positive, neutral, and negative affect conditions. Critically, all three panels reflect a similar pattern of fixations. Specifically, beginning between 450–650 ms after the onset of the noun, there was a clear preference for the linguistically named target referent.

As in Experiment 1, we used a target advantage score for the statistical analyses. In the current experiment, however, there is obviously no linguistically defined “competitor” belonging to the same category as the target. However, the display nonetheless contains a single display object that contrasts with the target in terms of its match to affect cues (i.e., the broken object), analogous to Experiment 1. Using this object in the calculation of the target advantage score therefore allows us to continue to evaluate the use of affect cues to differentiate the target from a relevant alternative, even though it is obvious that less competition is expected overall.

As before, the target advantage scores were submitted to a repeated measures ANOVA testing for the predicted linear effect. Consistent with the pattern apparent in Figure 5, the analysis showed no effect of vocal affect, $F(1, 14) = 2.349$, $\eta_p^2 = .144$, $p = .148$. This outcome, and in particular the absence of slowed target fixations in the negative affect condition, indicates that children had little tendency to perseverate on potentially incorrect referential predictions they generated based on vocal affect in the earlier part of the sentence. Rather, their fixation patterns seemed to reflect only the linguistic information carried by the noun.

In summary, when ambiguity was removed from the array and the speaker’s utterance, 5-year-old
children rapidly used vocal affect cues present early in the utterance to guide their referential hypotheses. However, as noun information was heard, children were able to rapidly recover from any incorrect referential predictions and locate the referent of the expression.

**Pointing behaviours**

Interrater reliability for 20% of the data \((n=3)\) was perfect (Cohen’s kappa=1.00, \(p < .001\)). Children’s pointing behaviour for this experiment indicated that children understood the task well. As expected, given the unambiguous description, children pointed to the correct target 100% of the time, regardless of the affect condition.

**Discussion**

These results demonstrate that 5-year-olds can use vocal affect cues early in an utterance to guide their referential expectations. Specifically, the eye fixation data indicated that children (incorrectly) anticipated

---

**Figure 5.** Proportion of fixations to display objects within the noun interval for Experiment 2. To view a colour version of this figure, please see the online issue of the Journal.
the broken object as the candidate referent when the instruction was spoken with negative vocal affect, but did so less with neutral affect and even less with positive affect. The altered referential situation used in Experiment 2 apparently allowed the effects of affect to be used more rapidly as the utterance unfolded in time. This outcome rules out two explanations that were consistent with the data up to this point. First, it cannot be the case that the indeterminacy of the sentence-final ambiguous noun in Experiment 1 (and earlier studies) was responsible for triggering the use of affect cues. If this were the case, we should have found no influence of vocal affect in the current experiment involving unambiguous reference. Second, the late effects of affect cues observed in past studies cannot reflect a need to hear a sufficiently long sample of speech before affect cues can be reliably identified. In the current experiment, affect cues influence referential hypotheses in advance of the noun. This outcome is broadly consistent with the finding that slightly older children (6 years of age) can accurately label a single-word utterance in terms of corresponding vocal affect (Waxer & Morton, 2011). The remaining explanation is that changes to the referential scenario are somehow responsible for children’s earlier ability to draw on affect cues—a topic that we return to later. One important point here is that the recorded auditory stimuli used in this experiment were identical to those used in Experiment 1. This means that unintended differences in presence or salience of the vocal affect cues across materials cannot account for the earlier sensitivity to affect cues observed here.

In the final experiment, we asked whether children would perform in a similar manner when positive affect, rather than negative affect, would provide misleading cues in the given referential scenario. This manipulation can reveal the extent to which the valence of affective information is important in developing referential expectations. Given research indicating that valence factors can influence both adults’ and children’s ability to correctly recognize vocal affect (e.g., Nelson & Russell, 2011; Pell & Kotz, 2011), it is possible that we may see different effects when positive affect versus negative affect provides the misleading cues, due to either the strength of these cues or the ease of relating the affect cues to objects of a given state.

EXPERIMENT 3

In Experiment 3, we once again provided children with a referential scenario consisting of three distinct object types, accompanied by an unambiguous utterance; however, in this study we altered the task in two important ways. First, we selected a broken object as the target object. (Recall that in Experiment 2, the target of the critical trials was always an intact object.) To maintain consistency with the overall design of Experiment 2, the arrays in the current experiment were therefore altered to include two broken objects and one intact object. Second, we changed the visual features of the intact distractor object such that it would be more likely to be associated with positive affect. For example, we used an object like a party hat or a decorated balloon rather than an object that was simply intact. These more appealing or exciting objects also helped to counteract the potential for attention to be drawn to the more visually complex nature of the two broken objects in the arrays (see Figure 6 for an example array).

Our core question was whether we would find a pattern analogous to the one found in Experiment 2 but with the affective valences reversed. Specifically, our question was whether positive affect cues in the early part of the utterance would lead children toward a referential expectation that is ultimately incorrect. If so, this would suggest that the potential to identify and link positive and negative affect to referential candidates is more or less equivalent and entails similar kinds of effects.

Method

Participants

Fifteen 5-year-olds were included in the final sample (7 males; M = 5.50 years, SD = 0.28 years), recruited through advertisements within the community. Participants were all from homes where English was the predominant language (1 child also spoke Spanish in the home).
Five additional children were tested but were removed from the final sample due to insufficient data collected from the eye tracker.

Stimuli

Critical trials. On critical trials, an array of three images was presented on a large display screen, accompanied by a recorded instruction relating to one of the objects (e.g., “Look at the ball. Point to the ball”). The recorded instructions were to the same as those used in the first two experiments, and, as before, the images were cropped photographs of real-world objects. Two of the images were of the broken type discussed earlier, and one of the images was of the intact type but was enhanced in a way to make it more likely to be associated with positive vocal affect. One of the two broken objects was the target of the referring expression (see Figure 6 for a sample array). The pairing of object arrays to vocal affect type was cycled across participants such that each array occurred only once in each affect condition.

Filler trials. These were the same as those in Experiment 2.

The full set of 24 trials was presented in a computer-controlled random order, and the position of photo objects within each display was also randomized.

Apparatus and procedure

These were identical to those in Experiments 1 and 2.

Results

Eye fixation patterns

Fixations were analysed using the same two analysis windows as those described in Experiments 1 and 2.

Prenoun interval. As before, we calculated the change (from the beginning to the end of the prenoun interval) in the proportion of fixations to the odd-man-out object (in this case, the intact object, which contrasted with the two broken/dirty objects). Mean fixation proportions at the beginning and end of the prenoun interval are...
shown in the lower panel of Figure 2. Of particular interest was whether vocal affect in the early part of the utterance influenced consideration of the intact object across conditions, with increases in positively-sounding affect across conditions leading to correspondingly greater consideration of this candidate. However, the mean slopes across conditions did not reflect any systematic pattern and were accompanied by high variability (negative: $M = 0.10, SD = 0.58$; neutral: $M = 0.08, SD = 0.50$; positive: $M = 0.17, SD = 0.46$). A within-participants ANOVA further confirmed that the predicted pattern of linear trend across vocal affect conditions was not present, $F(1, 14) = 0.246, \eta^2_p = .017, p = .628$. This pattern stands in contrast to the one observed in Experiment 2 where increasingly negative-sounding vocal affect led to increased consideration of the single broken object.

Noun interval. We identified the same window of analysis as that used in Experiments 1 and 2, beginning 200 ms after noun onset and ending 1,200 ms after noun onset. As before, fixations initiated before the beginning of this interval were excluded from analysis. Figure 7 presents the proportion of fixations to the three display objects across the noun interval for each condition. First, the negative affect condition (top panel) reflects children’s increasing consideration of the (broken) target object when compared to the intact competitor object, beginning at approximately 620 ms after the onset of the noun. Consideration of the target is comparatively reduced in the neutral affect condition (middle panel), and the trend completely reverses in the positive affect condition (bottom panel), where children initially show stronger referential consideration of the distractor in the noun region, before eventually fixating the target referent.

We again used a target advantage score to quantify the relative tendency to fixate the target over the distractor object associated with the opposite affective valence within the depicted speech interval. These scores were submitted to the same ANOVA model as that used for the prenoun measures. This analysis yielded a significant linear effect of vocal affect, $F(1, 14) = 4.863, \eta^2_p = .258, p = .045$. This outcome indicates that, as vocal affect moved from happy sounding to sad sounding across conditions, children were correspondingly better at differentiating the (broken) target object from the (intact) distractor object upon encountering the noun.

Pointing behaviours
Interrater reliability for 20% of the data ($n=3$) was perfect (Cohen’s kappa = 1.00, $p < .001$). As expected, children pointed to the correct target 100% of the time, indicating that they understood the task and could correctly identify the intended referent.

Discussion
In the current experiment, children did not use vocal affect to guide their consideration of the most directly isolable object in advance of hearing the noun. However, when the noun was encountered, we found an interference-like effect of positive affect that reduced fixations to the (broken) target object when a competitor object’s properties were more likely to be associated with positive affect. That is, children were more and more drawn to the intact competitor object when the vocal affect cues were increasingly happy sounding across conditions, even though the linguistic information available at that point in the speech stream was compatible with only the target referent (i.e., the target object could be differentiated from the alternatives on the basis of the initial sound in its corresponding noun). The influence of vocal affect was therefore delayed compared to the results of Experiment 2, where the effects were observed before the noun was heard. However, as indicated by children’s pointing data, children nonetheless selected the correct referent for the referring expression.

GENERAL DISCUSSION
Three experiments explored 5-year-olds’ ability to use a speaker’s vocal affect to incrementally guide referential hypotheses during real-time comprehension. Together, the results of all three
experiments help to clarify the core mechanisms underlying preschoolers’ ability to integrate vocal affect cues with linguistic information to understand a speaker’s referential intent.

First, results from Experiment 1 demonstrated that 5-year-olds used vocal affect to identify a referent for an ambiguous description. This sensitivity was reflected both in children’s explicit behavioural decisions and in their eye gaze behaviour. Consistent with previous research using the same experimental scenario with younger children (i.e., Berman et al., 2010), eye gaze patterns reflected the use of vocal affect only at the point where the noun was encountered. In other words, children did not seem to begin isolating a compatible referent from affect cues available earlier in the utterance. The 5-year-olds in the current study did, however, show a more robust ability to use affect cues than the younger children tested in past studies. This was reflected in the fact that children’s overt decisions about the target referent (as captured by their points to display objects) reflected the same sensitivity to affect cues as that found in their fixation patterns.

In partial contrast to Experiment 1, the results of Experiment 2 (using unambiguous referential

Figure 7. Proportion of fixations to display objects within the noun interval for Experiment 3. To view a colour version of this figure, please see the online issue of the Journal.
descriptions) demonstrated that children can use vocal affect information earlier in an utterance to develop hypotheses about possible referents. Specifically, when the utterance was spoken with increasingly more negative-sounding affect, children were more likely to anticipate reference to a single broken candidate in the visual scene. Although this object proved to be the incorrect referent when the noun was heard (i.e., children were temporarily garden-pathed), recovery from this prediction was comparatively swift, with no lingering penalty for an incongruent affect cue. The anticipatory effect rules out the idea that the “late” use of vocal affect information found in earlier studies resulted from the need to hear a longer sample of speech before vocal affect could be correctly identified. Further, the use of the affect cues in a situation with unambiguous reference demonstrates that the use of this information is not a type of strategic response that is triggered only when an ambiguous expression is encountered. However, Experiment 3 showed that the converse case involving a single intact object did not lead to the same result. Specifically, this object was not more likely to be anticipated when the utterance was accompanied by increasingly positive-sounding vocal affect. No affect-related differences were found in the prenoun interval in this experiment, despite an otherwise similar design. The relevant influence of vocal affect was, however, observed after the onset of the noun. Specifically, more positive-sounding vocal affect increased the tendency to fixate the intact distractor object, slowing identification of the (broken) target object.

What might account for the timing differences across Experiments 2 and 3? One possible explanation comes from research indicating that adults are quicker to identify sad vocal affect than positive vocal affect. In one such study, Pell and Kotz (2011) found that on average it took adults 576 ms to identify sadness in vocal affect versus 977 ms to recognize happiness. Thus, our result may reflect a timing difference for the relevant recognition processes, which in turn allow the more promptly detected negative-sounding vocal affect cues to be quickly linked to a single referent bearing relevant characteristics (i.e., a broken object, presented among two intact ones). A related possibility is that children are simply more successful at recognizing negative vocal affect than positive vocal affect. For example, Nelson and Russell (2011) found that children from 3 to 5 years old were significantly better at recognizing sadness (72%) when hearing a speaker’s voice than happiness, anger, or fear (34%). This finding is also consistent with research demonstrating that adults are more successful in recognizing sadness from a speaker’s voice than happiness (Paulmann & Pell, 2011). These differences in the recognition of different types of vocal affect extend cross-culturally (Pell, Monetta, Paulmann, & Kotz, 2009) and are also found when adults are listening to foreign languages (Pell, Paulmann, Dara, Alasseri, & Kotz, 2009; Scherer, Banse, & Wallbott, 2001). Another possible explanation for the timing differences is that it is easier to associate a relevant emotional disposition with broken or dirty objects than with intact or embellished objects. For example, perhaps a broken cell phone is more easily related to an upset or sad emotional reaction than a decorated rubber ducky is to a happy reaction.

When considered in conjunction with previous research, our findings highlight the developmental emergence of children’s use of vocal affect as a marker of referential intent in language. Past work indicates that 3-year-olds do not show either an implicit or an explicit sensitivity to vocal affect cues, whereas 4-year-olds do reflect the appropriate sensitivity, but only in their eye gaze, patterns and not in their overt decisions about the identity of the intended referent (Berman et al., 2010). The present experiments suggest that by 5 years of age, children show an appreciation for vocal affect at both the implicit and the explicit level. What might account for this age difference, given that 4-year-olds showed sensitivity to vocal affect in their eye gaze, but not in their pointing? Perhaps younger children were overwhelmed by the cognitive demands of the task and with age have more available cognitive resources to successfully cope. In fact, a number of recent studies have suggested that children’s inability to
combine linguistic and affective speech cues may result from the overextension of cognitive capacities (e.g., Friend, 2000; Morton & Munakata, 2002; Morton & Trehub, 2001; Morton, Trehub, & Zelazo, 2003; Waxer & Morton, 2011).

If correct, this capacity explanation may provide a clue to explain another facet of the current findings—namely, why a potentially ambiguous referential situation seems to delay the point at which affect information begins to play a role. Recall that Experiments 1 and 2 used identical linguistic stimuli and visual displays except for the fact that two objects of the same category were included in the object array in Experiment 1, leading to referential ambiguity when the description was heard. Importantly, then, the use of affect cues in advance of the noun in Experiment 2 cannot be due to the acoustic features of the recorded materials on critical trials, nor were there differences in the potential to use affect information to isolate a single “odd-man-out” at this point in time. Further, the filler trials ensured that participants would not adopt a particular global processing strategy based on repeated exposure to either ambiguous or unambiguous descriptions. For example, the critical trials in Experiment 1 (with ambiguous descriptions) were intermixed with fillers in which descriptions were unambiguous. The relevant differences therefore seem to be related to the presence of two same-category exemplars in the Experiment 1 displays (which in turn makes an ambiguous description possible). Assuming 5-year-olds still reflect some capacity limitations in the coordination of linguistic and nonlinguistic information, it is plausible that the need to differentiate the array of visual objects using both category membership information and information about objects’ idiosyncratic properties may tax children’s attentional and representational systems. This would then lead to difficulties in the rapid use of vocal affect information during referential interpretation. Although speculative, this explanation bears some interesting relationships with other documented cases in which contexts with multiple same-category exemplars have an impact on children’s real-time comprehension. For example, a number of studies have shown that children exhibit difficulties in recognizing how these kinds of contexts can help disambiguate the syntactic structure of unfolding utterances (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2001; Trueswell, Sekerina, Hill, & Logrip, 1999). Future studies using older children or adults could clarify the extent to which young children’s capacity limitations might delay the integration of vocal affect cues in certain contextual situations.

Finally, the present results add to the growing literature documenting how paralinguistic properties of speech can rapidly aid in real-time interpretation. For example, studies of speaker disfluencies (e.g., Arnold, Altmann, Fagnano, & Tanenhaus, 2004; Arnold, Fagnano, & Tanenhaus, 2003; Barr & Seyfeddinipur, 2009; Kidd, White, & Aslin, 2011) have found that both children and adults expect a speaker to be referring to new information rather than information that has already been established in the discourse when the speaker produces a disfluency (i.e., “Umm”; Arnold et al., 2004; Kidd et al., 2011). Other types of paralinguistic information also play a role, such as cues to the identity of a speaker (e.g., Creel, 2012; Creel & Bregman, 2011). Our results extend these findings, suggesting that information in the speech stream that conveys an emotional disposition can be used by children to narrow their referential hypotheses under various circumstances.

Original manuscript received 30 January 2012
Accepted revision received 8 July 2012
First published online 4 September 2012

REFERENCES

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition, 73, 247–264.
Arnold, J. E., Altmann, R., Fagnano, M., & Tanenhaus, M. K. (2004). The old and thee, uh, new. Psychological Science, 9, 578–582.
Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: Eyetracking evidence of the
time-course of pronoun resolution. *Cognition*, 76, B13–B26.
Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal thee, um, new information. *Journal of Psycholinguistic Research*, 32, 25–36.
Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614–636.
Barr, D. J., & Seyfeddinipur, M. (2009). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, 25, 441–455.
Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2010). Preschoolers’ appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, 107, 87–99.
Chambers, C. G., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition*, 108, 26–50.
Creel, S. C. (2012). Preschoolers’ use of talker information in on-line comprehension. *Child Development*. doi: 10.1111/j.1467-8624.2012.01816.x
Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass*, 5(5), 190–204.
Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language*, 47, 292–314.
Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, 97, 412–429.
Friend, M. (2000). Developmental changes in sensitivity to vocal paralanguage. *Developmental Science*, 3, 148–162.
Friend, M. (2001). The transition from affective to linguistic meaning. *First Language*, 31, 219–243.
Heller, D., & Chambers, C. G. (2011, June). Effects of adjective type on referential expectations in discourse. Poster presented at the 2011 Conference on Experimental Pragmatics (XPRAG), Barcelona, Spain.
Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, 108, 831–836.
Hertzog, C., & Rovine, M. (1985). Repeated-measures analysis of variance in developmental research: Selected issues. *Child Development*, 56, 787–809.
Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L., & Trueswell, J. (2001). One frog, two frog, red frog, blue frog: The development of parsing strategies in children. *Journal of Psycholinguistic Research*, 29, 597–626.
Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58, 541–573.
Kamide, Y., Altmann, G., & Scheepers, C. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32, 37–55.
Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers’ referential intentions. *Developmental Science*, 14, 925–934.
Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, 35, 1009–1051.
Morton, J. B., & Munakata, Y. (2002). Are you listening? Exploring a developmental knowledge-action dissociation in a speech interpretation task. *Developmental Science*, 5, 435–440.
Morton, J. B., & Trehub, S. E. (2001). Children’s understanding of emotion in speech. *Child Development*, 72, 834–843.
Morton, J. B., Trehub, S. E., & Zelazo, P. D. (2003). Sources of inflexibility in 6-year-olds’ understanding of emotion in speech. *Child Development*, 74, 1857–1868.
Nelson, N. L., & Russell, J. A. (2011). Preschoolers’ use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, 110, 52–61.
Nygaard, L. C., & Lunders, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition*, 30, 583–593.
Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, 35, 192–201.
Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: Evidence from eye movements. *Speech Communication*, 54, 92–107.
Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, 6, e27256.
Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, 33, 107–120.
Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37, 417–435.
Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*, 76–92.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*, 89–134.

Tsang, C., & Chambers, C. G. (2011). Appearances aren’t everything: Shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology: Learning, Memory & Cognition, 37*, 1065–1080.

van Berkum, J., van den Brink, D., Tesink, C., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*, 580–591.

Waxer, M., & Morton, J. B. (2011). Children’s judgments of emotion from conflicting cues in speech: Why 6-year-olds are so inflexible. *Child Development, 82*, 1648–1660.

Weber, A., Braun, B., & Crocker, M. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49*, 367–392.