University of Calgary

PRISM Repository | https://prism.ucalgary.ca

The Vault | Open Theses and Dissertations

2013-06-26

Preschoolers' Use of Vocal Affect as a Cue to Referential Intent

Berman, Jared

Berman, J. (2013). Preschoolers' Use of Vocal Affect as a Cue to Referential Intent (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/27418

http://hdl.handle.net/11023/766

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Preschoolers' Use of Vocal Affect as a Cue to Referential Intent

by

Jared Berman

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

CALGARY, ALBERTA

JUNE, 2013

© Jared Berman 2013

Abstract

In this dissertation, I examined the development of preschoolers' use of a speaker's vocal affect to guide their understanding of referential situations. Across all experiments, children's eye gaze was measured as an index of their underlying interpretive processes.

The first set of experiments examined 5-year-olds' use of vocal affect to understand the referential intentions of a speaker. Across three experiments, differences in 5-year-olds' use of vocal affect across multiple contexts (i.e., ambiguous: two same category referents vs. unambiguous: only one linguistically defined referent) and valence types (i.e., positive vs. negative) were investigated to determine whether vocal affect might help children identify a linguistically ambiguous referent or speed the detection of a linguistically unambiguous referent. Results indicated that 5-year-olds use both positive and negative vocal affect to resolve referential ambiguity. Further, when 5-year-olds were provided with unambiguous contexts, negative, but not positive vocal affect altered children's eye gaze early in the utterance (i.e., prenoun).

Second, I examined whether 4- and 5-year-olds could use vocal affect to learn new words. Results indicated that both 4- and 5-year-olds' eye gaze reflected the use of both positive and negative vocal affect to map novel labels to novel objects. When 5-year-olds completed extension and generalization tasks, negative but not positive vocal affect played a role in children's ability to learn new words.

Finally, I examined whether 3- and 5-year-olds can use a speaker's vocal affect to match the face of the speaker. Eye gaze results indicated that both 3- and 5-year-olds could use a speaker's vocal affect to find her matching emotional face. However, there

were significant differences across valence types in terms of timing. Specifically, children's eye gaze reflected an earlier use of negative compared to positive vocal affect. However, when children were asked to explicitly locate the matching face (i.e., through pointing) only 5-year-olds displayed this ability.

The findings from this dissertation demonstrate that preschoolers can use a speaker's vocal affect as a cue to referential intent. Furthermore, when provided with negative vocal affect children were not only more accurate, but were able to use this information earlier in the utterance.

Acknowledgements

I would like to extend my heartfelt gratitude to my supervisors, Dr. Susan Graham and Dr. Craig Chambers, for their guidance, patience, feedback, and support over the past six years. I would also like to express my appreciation to my committee members, Dr. Suzanne Curtin and Dr. Penny Pexman, as well as my external examiners Dr. Marc Pell and Dr. Mary Grantham-O'Brien for their insightful comments and their participation at my oral examination. I would also like to thank Heather MacKenzie for her assistance in preparing the stimuli for this project as well as all of her support during our time together in Calgary. I would especially like to thank all of the children and parents who kindly volunteered their time to participate in this research. I am also extremely grateful to the members of the Language and Cognitive Development Lab (both past and present) who always provided a fun and stimulating academic environment.

To my parents and extended family: Thank you. Your support, confidence and love were unconditional and helped me make it through my graduate education. To my wife: thank you for always being there for me. I have appreciated your love and patience over the past six years and I look forward to many more years of happiness.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

The ability to accurately perceive and use a speaker's emotional cues is an essential part of becoming an effective communicator. Consider, for example, the utterance "I'll be there soon!". The message that the speaker conveys varies considerably depending whether the speaker's voice conveys excitement (perhaps a parent coming to his or her child's play) or worry (perhaps a parent going to pick up a sick child). This emotional information, which I will refer to as *vocal affect*, is conveyed alongside the content of the statement and arises from variations in pitch level, pitch contours, and speech rate (see Banse & Scherer, 1996; Frick, 1985).

In this dissertation, I explore children's sensitivity to vocal affect from multiple perspectives. In Chapters 2 and 3, I investigate preschoolers' use of vocal affect to gauge a speaker's referential intent. More specifically, I examine 4- and 5-year-olds' use of vocal affect to help determine which object is the speaker's intended referent in situations where the linguistic content of a noun phrase does or does not provide sufficient information for reference resolution on its own (e.g., "Look at the ball" vs. "Look at the blicket"). In Chapter 4, I examine preschoolers' sensitivity to vocal affect cues in situations that do not require the referent of a noun phrase to be identified. The main question here was children's ability to "match" the same spoken sentences used in Chapters 2 and 3 to a face depicting a corresponding emotion. In all cases, a primary consideration is the time course of interpretation.

In this first chapter, I review studies that address the emergence of the different component abilities involved in the perception of vocal affect, as well as research that describes how these components interact in the mature (adult) comprehension system. I begin by describing infants' detection and discrimination of emotional signals across modalities. Next, I review the literature addressing both adults' and children's use of vocal affect in different experimental contexts. Finally, I discuss recent work investigating preschoolers' use of vocal affect as a marker of referential intent, and consider the outstanding questions arising from this work.

Infants' Use of Emotional Cues

Emotion detection and discrimination. The ability to detect and discriminate a speaker's emotional state is a critical first step children must navigate before they can begin to make use of this information in communicative exchanges. Research has demonstrated that infants' ability to discriminate emotional categories is most accurate when they are provided with information from multiple channels, namely the face and the voice. However, when the emotional information is provided through only a single channel (i.e., modality specific information, typically face *or* voice), it is not until later in infancy where children can use this information to discriminate emotional categories (Flom & Bahrick, 2007). For example, infants at 4 months of age can reliably discriminate between happy, sad, and angry emotional expressions as long as the information is conveyed through both the face and the voice. However, it is not until 5 months of age that infants can discriminate the same emotions based on vocal affect alone and not until 7 months of age that the same feat can be accomplished with only facial affect (i.e., a facially-signalled emotional state; Flom & Bahrick, 2007). Research by Walker-Andrews and colleagues suggests that infants may use vocal affect alone before 7 months of age, but only when an infant's mother provides the cue. For example, Kahana-Kalman and Walker-Andrews (2001) found that 4-month-old infants would look longer at films of their mother smiling versus a film of their mother frowning if her accompanying recorded speech was spoken with positive affect. When their mothers' accompanying speech was produced with negative affect, infants looked longer at films of her frowning. When, however, the infants watched an unfamiliar woman smiling or frowning and speaking in either positive or negative affect, infants looked equally at both films, regardless of the type of speech. By 7 months of age, however, infants can use unfamiliar faces and voices to discriminate across emotional categories (Walker-Andrews, 1986). Finally, more recent ERP studies have found patterns supporting the view that there is more consistent discrimination and recognition for multimodal versus unimodal emotional stimuli prior to 7-months of age (Grossmann, Striano, & Friederici, 2006; Grossman, Striano, & Friederici, 2007).

The emergence of infants' use of emotion. As infants reach the end of their first year of life, researchers have investigated how they move beyond detecting and discriminating emotions to how they begin to make use of emotional information to make behavioural decisions. One important line of research has used social-referencing paradigms to examine this question. For example, Klinnert (1984) had mothers of 12- and 18-month-old infants pose happy or fearful expressions towards novel objects. As expected, infants were most likely to leave their mother's side to explore the novel objects in the happy expression condition and least likely in the fearful expression condition. Similarly, Mumme and Fernald (2003) found that 12-month-olds could use an experimenter's negative emotional cues (i.e., frowning and negative vocal affect) to ignore a newly discovered object (see also Mumme, Fernald & Herrera, 1996). Further, using a visual cliff paradigm, Vaish and Striano (2004) demonstrated that infants would cross the cliff more quickly when presented with multi-modal encouragement (i.e., positive vocal and facial affect) compared to facial or vocal affect alone. Together, these results suggest that infants as young as 12 months of age can use information provided by their parents' emotional signals to guide their behaviour. Further, in conjunction with the early infancy results described above, these results also show that infants are more successful across the first year of life both in discriminating emotions as well as using emotional information when they are provided with multi-modal rather than modality-specific emotional information.

Speaker Vocal Affect

During early childhood, infants are frequently within arm's reach of their primary caregiver. As such, their exposure to emotional information is most often conveyed through multiple channels. However, as children develop physically, their ability for self-locomotion increases, as does their capacity to interact with objects in the world. These changes arguably heighten the relevance of cues that can convey emotional information even when the speaker is not the explicit focus of the child's visual attention. Consider, the following question from a mother to her daughter standing at the bottom of the stairs: "What did you do?". The meaning of her utterance varies considerably depending on whether the mother is excited (perhaps the child left a present on the kitchen table) or upset (perhaps the child made a sibling cry). In this example, the underlying emotion conveyed through the mother's vocal affect could help the child infer the mother's intent: either to thank the child or to admonish her. That is, if the child is going to understand her mother's statement, she will need to quickly integrate the emotional information with the specific content of the sentence to understand her mother's intent. These issues of when and how vocal affect is integrated into language processing have been explored in the adult processing literature.

Adults' sensitivity to vocal affect. Research has investigated adults' sensitivity to vocal affect from several perspectives. One line of research has investigated the online influence of vocal affect on adults' classification of emotions (Pell, 2005a, 2005b; Pell, Jaywant, Monetta, & Kotz, 2011; Paulmann & Pell, 2010; Pell & Skorup, 2008;). For example Pell (2005a) developed a task to measure the online nature of emotional processing (i.e., Facial Affect Decision Task: FADT). In the first iteration of the FADT, Pell considered how vocal affect might prime decisions about the representativeness of emotional faces. More specifically, adults were primed with nonsense speech (e.g., "Someone migged the pazing") spoken in one of four different vocal affect conditions (i.e., pleasant surprise, happiness, sadness and anger) and then needed to make a yes/no judgment regarding whether a face represented a "true" emotion or a simply a facial expression (i.e., a grimace). Results from this set of studies suggested that adults were influenced by the vocal affect of the nonsense sentence prime when making decisions regarding the representativeness of an emotional face. That is, when the vocal affect prime was congruent with the target face, adults were faster to judge the appropriateness of the target face. Building on this work, Pell and colleagues (2011) used the FADT to demonstrate that, when adults were provided with multiple congruent cues to emotion (i.e., vocal affect, semantics, and facial cues), their accuracy was increased compared to conditions where only one cue was provided. Overall, research conducted using the FADT had suggested that when adults were primed with a speaker's vocal affect they could use this information extremely quickly to guide their decisions regarding the representativeness of an emotional face. However, this task does not directly test the time course of adults' recognition of emotional categories. In order to assess this, Pell and Kotz (2011) conducted an auditory gating study to investigate the amount of emotional speech adult listeners needed to hear before they could recognize the underlying emotion of the speaker. Here, the authors found that negative emotions (i.e., anger, fear, and sadness) were recognized more quickly than a positive emotion (i.e., happiness). The one exception to this trend was the recognition of disgust, which took significantly longer than all other emotions tested. Overall, this line of research has found that adults can use a speaker's vocal affect online to help make decisions regarding whether or not an emotional face reflects a "true" emotion. Further, these results suggest that adults are able to recognize negative emotions more quickly than positive emotions.

Another line of research has suggested that adults are influenced by vocal affect in tasks involving the resolution of lexical ambiguity (e.g., Nygaard & Lunders, 2002; Nygaard & Queen, 2008). For example, Nygaard and Lunders (2002) examined adults' perception of homophones that differed in terms of the strength of their association with a particular emotion (e.g., bridal/bridle or die/dye). The test words were recorded with different types of vocal affect, and participants in the experiment were asked to transcribe the words they heard. Results indicated that listeners were more likely to transcribe the homophone with the spelling that reflected the affect-congruent meaning (i.e., adults will transcribe "die" rather then "dye" when the word is spoken with negative vocal affect). These results suggest that vocal affect affects linguistic processing even at the level of individual words.

Taken together, the findings described above suggest that adults use vocal affect rapidly to facilitate the understanding of language. However, to date, we know less about children's use of vocal affect, particularly in experiments measuring online behaviour. The research that has been completed has investigated children's use of vocal affect to identify a speaker's underlying emotional disposition. The following section describes the developmental research related to this topic.

Children's use of vocal affect to recognize a speaker's emotional state. Research on preschoolers' use of vocal affect to recognize a speaker's emotional state has been a topic of interest over the past fifteen years. Studies have shown that, by four years of age, children can accurately judge a speaker's emotional state based on vocal affect cues if the emotional content of the speech sample is either neutral or filtered out (Friend, 2000, Morton & Trehub, 2001; Morton, Trehub, & Zelazo, 2003; Nelson & Russell, 2011; Quam & Swingley, 2011). For example, Nelson and Russell (2011) compared children's use of vocal affect to their use of facial or postural affect to determine which cues were most reliably used to judge an individual's underlying emotional state. Here, children between the ages of three and five were presented with short video clips of a female actor displaying one of four different emotions (i.e., happiness, sadness, anger, or fear) in four different conditions (i.e., face-only, posture-only, voice-only, and multi-cue). Not surprisingly, preschoolers were more successful when provided with multiple cues to emotional state, but were able to use all three cue types individually to identify the experimenter's emotional state. However, there were important differences across cue types. In particular, in the voice-only condition, compared to the face- or posture-only conditions, children were least successful at correctly judging the speaker's emotional state. In this task, children listened to a neutral sentence (i.e., "I felt this feeling before; it was just a few days ago") presented in one of the four different vocal affect conditions and were asked to label the speaker's emotional state. The finding that children were least successful in the voice-only condition likely stems from significant recognition differences between sad (72%) and the other 3 emotions (happiness, anger, and fear; 34%). That is, when the different emotional categories are individually inspected, the children's responses suggest they can recognize negative, but not positive emotions when presented through the vocal channel. Together, these results suggest that not only might it be comparatively difficult for preschoolers to use a speaker's vocal affect to recognize her associated emotion, but that there may be important differences across emotional categories.

What about situations where the information expressed in an utterance is *not* neutral, and the vocal affect of the speaker might not be congruent with the content of her speech? In order to study the relative importance of these cues, researchers have used experimental paradigms where these two cues to a speaker's underlying emotion offer conflicting information. In one such study, Morton and Trehub (2001) had preschoolers listen to sentences that described either happy or sad events for the speaker (e.g., "My dog ran away from home", for a sad event). Each sentence was presented with both positive and negative vocal affect. The pairing of sentences with both happy and sad vocal affect allowed for the creation of conflict trials wherein a sentence describing a happy event (e.g., "I got an ice cream for being good") would be paired with sad vocal affect and cases where a sentence describing a sad event (e.g., "All the kids at camp tease me") was paired with a happy vocal affect. In these conflict situations, 4-year-olds relied almost exclusively on the content of the sentences to judge the emotional state of the speaker (for a similar result see Friend, 2000). This same study demonstrated that by age 9 children begin to show attenuation in their reliance on the content of the sentences and move towards more adult-like behaviour. More specifically, in situations where the content of the sentence was incongruent with the vocal affect of the speaker, 9-year-olds and adults used the speaker's vocal affect to judge the speaker's emotional state.

There are a number of possible explanations for why preschool children between the ages of 4 and 9 privilege the sentence content over the vocal affect of the speaker in this kind of experimental paradigm. On one account, young children's reliance on linguistic content over vocal affect stems from their difficulty adjudicating between conflicting sources of information, rather than from a disregard for the information carried by vocal affect (Morton, Trehub, & Zelazo, 2003). That is, children's failure to use vocal affect could be explained as a by-product of preschoolers' emerging ability to coordinate multiple cues during language comprehension, rather than as a failure to correctly use affective information or recognize its relevance for interpretation. A related possibility is that the prioritization of linguistic content over vocal affect stems from a more specific difficulty involving the coordination of multiple emotional cues in particular (Waxer & Morton, 2011). More specifically, Waxer and Morton propose that children's difficulty in tasks using conflict to assess the use of vocal affect stems from the fact that happy and sad are mutually inhibitory emotions and therefore this task may have taxed children's limited cognitive resources and led to their apparent disregard of vocal affect.

Although as mentioned above, 4-year-olds' explicit judgments privileged sentence content over vocal affect, Morton and Trehub (2001) did note that 4-year-olds had significantly longer response latencies for the conflict trials compared to trials where the vocal affect of the speaker and the content of the sentence matched. This result provides some indirect evidence that children were processing both the content and the vocal affect of the utterance. It is therefore possible that a task that is more naturalistic or less cognitively demanding may allow younger children to display an appreciation for vocal affect. The latency data also raise the possibility that more automatic and time-sensitive measures may be needed to accurately track developmental patterns in the use of vocal affect. One possible way to test children's coordination of linguistic information and vocal affect would therefore be to use a gaze tracking methodology. Here, an eye tracker may provide information regarding the ongoing decision processes children engage in when attempting to adjudicate multiple cues. Further, this methodology would be sensitive to the moment-by-moment processing that is happening as speech is being encountered.

Children's use of vocal affect to resolve linguistic ambiguity. One recent study used an eye tracking methodology to investigate children's use of vocal affect to gauge a speaker's referential intentions (Berman, Chambers, Graham, 2010). In this study, 3- and 4-year-olds were presented with an array of three photos including two members of the same category that differed in their properties (e.g., an intact doll and a broken doll). On critical trials, children were then asked to find a particular referent using an ambiguous phrase (e.g., "Look at the doll…Point to the doll"). Across conditions, vocal affect was varied (happy-sounding, sad-sounding, or neutral), and eye gaze was recorded as the sentences were played. For consistency, throughout the remainder of this dissertation "happy-sounding" speech will be referred to as positive vocal affect and "sad-sounding" speech will be referred to as negative vocal affect. Findings indicated that 4-year-olds' eye gaze, but not their pointing behaviour, reflected sensitivity to the speaker's vocal affect. For example, as the ambiguous noun (e.g., "*doll"*) unfolded in time, 4-year-old children considered the "broken" referential candidate 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. These results suggest that the information gained from the speaker's vocal affect was combined with the semantic content of the noun phrase once ambiguity was encountered. Interestingly, during the initial part of the carrier sentence (i.e., "Look at the"), there was no effect of the speaker's vocal affect on children's eye gaze behaviour. These results suggest that children's eye gaze is not immediately influenced by the speaker's vocal affect during sentence processing.

Although 4-year-olds' eye gaze indicated an appreciation for vocal affect, their associated pointing behaviour did not. Importantly, pointing behaviour in this task is used as an index of children's final referential decision after all information from the utterance has been understood. When combined with the eye gaze data, the pointing data suggest that children's ability to integrate vocal affect information with speech information is still at an emergent and somewhat fragile stage. Furthermore, neither 3-year-olds' pointing behaviour nor their eye gaze reflected the use of the speaker's vocal affect. These results suggest that 4-year-olds, but not 3-year-olds can use vocal affect to help resolve referential ambiguity and understand a speaker's intent, however, even at 4 years of age this skill is still developing.

Summary and Overview of Dissertation Chapters.

To this point, research has highlighted a development trend whereby infants' use of emotional information is quite fragile and they often require multimodal emotional cues in order to detect and discriminate between emotions. The recognition of emotions is then observed to guide simple behavioural decisions in infants. As children reach the preschool years, we begin to see more direct evidence of the use of vocal affect. For example, 3- to 5-year-old children can judge the emotional state of a speaker as long as the content of her speech is neutral (Nelson & Russell, 2011). Further, recent evidence has suggested that 4-year-olds can use a speaker's vocal affect to infer referential intent (Berman et al., 2010). However, when the task requires preschoolers to engage in more complex forms of combinatory processing, their use of vocal affect appears to be significantly delayed. More specifically, when preschoolers are faced with conflicting vocal affect and content cues, they appear insensitive to vocal affect when judging the emotional state of the speaker. Taken together, these results and their associated explanations point to a number of areas for future study regarding children's use of vocal affect across multiple perspectives.

In this dissertation, I examine preschoolers' use of vocal affect across a number of different contexts with a specific focus on the following questions: (1) are there differences in valence and timing in children's use of vocal affect?; (2) does children's use of vocal affect depend on the presence of ambiguity?; and (3) can children use vocal affect to learn new words? All three chapters use an eye tracking methodology in order to assess the moment-by-moment processes occurring during these tasks. This methodology allows for insights into children's communicative abilities by capturing the ongoing interpretive process, and not merely the outcome in terms of the effects on children's behaviour. Further, these measures can in fact capture transitory interpretations that differ from children's final behavioural decisions.

First, with regards to valence and timing, Berman and colleagues (2010) suggested that children can use both positive and negative vocal affect to resolve ambiguity but did not directly test differences across valence types. Recent work with adults has found valence differences across emotional categories. Specifically, Paulmann and Pell (2011) found that adults were better able to recognize sadness compared to happiness in a speaker's voice. Further, the evidence to date suggests that adults are not only better at recognizing sadness, but are also significantly faster in their recognition when presented with negative rather than positive vocal affect (Pell & Kotz, 2011). More specifically, Pell and Kotz (2011) found that it took adults approximately 400 ms longer, on average, to recognize positive versus negative vocal affect. Recent developmental research has also suggested that children are more successful at labeling sad versus happy vocal affect in a free labelling task (Nelson & Russell, 2011). Finally, results from Berman and colleagues (2010) were suggestive of a stronger influence of negative vocal affect. Although these differences were not directly tested, 4-year-olds began to differentiate between the broken and intact objects approximately 350 ms after the onset of the noun in the negative affect condition whereas it was not until approximately 500 ms after the onset of the noun where positive vocal affect aided children in differentiating between the two referential possibilities. These results point to a stronger and more rapid influence of negative vocal affect, however, research to date has not found a differential impact of negative vocal affect or an early use of this cue in sentence processing. Across all chapters in this dissertation, experiments were set up to directly test differences across valence types (i.e., positive versus negative vocal affect).

With regards to context, we know that 4-year-olds can use vocal affect (at least implicitly) to help resolve ambiguity in referential situations, however, we know little about their use of vocal affect in unambiguous referential situations. Recall, that Berman et al. (2010) presented children with utterances that were referentially indeterminate (i.e., "Look at the doll", when there were two potential doll referents in the array). Here, results suggested that even though vocal affect cues were available during the carrier phrase (i.e., "Look at the…") children did not use vocal affect during this period to help begin to make referential decisions. Instead, preschoolers' eye gaze reflected a relatively late use of this cue (i.e., during the occurrence of the ambiguous noun: "doll"). There are a number of possible explanations for this result. First, it is possible that children simply needed a sufficiently long speech sample before they could recognize or apply vocal affect information, thereby making the relevant effects apparent only for the sentence-final word. Another possibility is that children's use of vocal affect was more strategic and became evident during the noun because this is where children recognized ambiguity in the unfolding sentence. A third possibility is that the multi-referent display (i.e., two dolls) might separately impose representational demands on children's cognitive systems and lead to a delayed use of vocal affect. These issues, if addressed, would provide insights into children's information-integration systems and how these may differ from adults' more automatic use of vocal affect.

To address the above possibilities in Chapter 2, I examine how context influences 5-year-olds' use of vocal affect, using the same basic experimental paradigm as Berman and colleagues (2010). More specifically, I use different visual contexts to examine both the underlying mechanisms responsible for children's use of vocal affect as well as important differences in valence across vocal affect types (i.e., positive, neutral and negative). Across three experiments, I compare 5-year-olds' use of vocal affect in situations where the linguistic information provided does or does not resolve ambiguity (e.g., "Look at the doll", when the visual array contains either one or multiple dolls). This manipulation allows for closer inspection of the effect of altering the visual and linguistic context of the task. Further, manipulating the specific object of the referring expression (i.e., a broken vs. decorated cell phone when an individual child hears "Look at the phone") allows for the direct comparison of valence effects. Results from these experiments indicate that by five years of age children can explicitly use vocal affect to resolve referential ambiguity. Further, the results suggest a differential use across valence types. More specifically, during the carrier phrase (i.e., "Look at the…") children's eye gaze suggested that 5-year-olds used negative, but not positive vocal affect in unambiguous contexts to anticipate reference to a particular object in the display.

In addition to resolving referential ambiguity, there are other contexts where vocal affect may help children resolve other kinds of indeterminacies. For example, word learning situations share a number of features with the task used by Berman and colleagues (2010). When learning new words in an ostensive context, children are faced with inherent ambiguity. Both the objects used and the specific labels for these objects are novel to children. More specifically, in these tasks children are provided with known carrier phrases (e.g., "Look at", "Find the", or "Where is the"), which end in a novel word (e.g., "wugen", "daxel", or "blicket") and are required to map the label to one of two novel objects presented. Researchers have investigated a variety of social cues that children rely on to resolve the inherent ambiguity in the word learning process. However, the use of vocal affect as a cue to learn new words has not been directly tested. This is somewhat surprising given the pervasiveness of vocal affect in everyday communicative exchanges. In addition, even young infants have begun to use this information (in conjunction with other emotional cues) in social situations, so it seems quite possible that preschoolers are using this information to learn new words. Directly testing children's ability to use vocal affect to learn new words may provide additional information regarding children's use of vocal affect and whether ambiguity is necessary (at this age) to trigger its use. Further, if preschoolers can use vocal affect to learn new words this result would add to the growing literature suggesting that word learning is a social process that is driven by a processing of the referential intent of others (Bloom, 1997).

In Chapter 3, I report two experiments with 4- and 5-year-olds designed to test the hypothesis that children can use a speaker's vocal affect to learn new words. More specifically, I examine whether children can use a speaker's positive and negative vocal affect to map novel labels to unfamiliar objects. The first experiment compares 4 and 5-year-olds' ability to map novel words to novel objects on the basis of vocal affect cues. In the second experiment, 5-year-olds were tested to see whether they could also extend and generalize the newly learned labels to new category exemplars. Across both experiments, children's eye gaze was measured to provide an index of their ongoing interpretive processes. Children's conscious and final referential decisions were measured based on pointing behaviour. Results from this chapter indicate that both 4-

and 5-year-olds can implicitly use vocal affect to help map new words to objects. However, it is not until five years of age that children demonstrate their learning through pointing. Five-year-olds can also extend and generalize their learning to new exemplars. Lastly, the results suggest that children were more sensitive to negative vocal affect.

Further, in Chapter 4, I examine an additional task that may provide insight into preschoolers' use of vocal affect. I use the vocal stimuli from chapters 2 and 3 and replace the known objects from Chapter 2 with faces to investigate whether using these faces, which are possibly more socially relevant than objects, may simplify the task and allow younger children to display an appreciation for the use of this cue. More specifically, children's eye gaze was measured while they were asked to point to the face that matched how the speaker was feeling when she made a specific utterance. For example, children heard the phrase "Look at the doll" in one of three vocal affect conditions (i.e., positive, neutral, or negative) and were asked to point to one of the three emotional faces (i.e., happy, neutral, or sad) that matched how the speaker was feeling. Results from this chapter indicate that both 3- and 5-year-olds' eye gaze reflected an ability to link vocal affect to the associated facial affect, however, only 5-year-olds could display their learning through their pointing behaviour. In addition, the results suggested that negative vocal affect played a role earlier in the time course of the utterance.

In the final chapter, I review the full set of findings and draw on the results to describe the developmental progression of preschoolers' use of vocal affect, the mechanisms that may underlie the use of vocal affect and how the use of this cue fits into children's broader social-emotional development. In addition, I describe limitations of the research conducted for this dissertation as well as future directions for research in this area.

The following chapter is a reproduction of a published work, for which permission has been granted by the publishers, Taylor and Francis, for the use in this dissertation. I was the primary investigator and main contributor for the publication titled, Contextual influences on children's use of vocal affect cues during referential interpretation, in the journal *Quarterly Journal of Experimental Psychology*, and was responsible for the conception, design, data collection and analysis with Dr. Craig Chambers and Dr. Susan A. Graham acting as my supervisors.

Berman, J. M. J., Chambers, C., & Graham, S. A. (*2013*). Contextual influences on children's use of vocal affect cues. *Quarterly Journal of Experimental Psychology, iFirst, 1-22.* doi: 10.1080/17470218.2012.713367

Chapter 2: Contextual influences on children's use of vocal affect cues during referential interpretation

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 preschooler's 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 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 behavior 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 behavior based on this cue (Friend, 2001). One question is whether the task demands involved in judgment 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 judgments 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 on-line 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 behavior 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 behavior 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 two-fold: 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 behavior 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 pre-noun 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 compared to 4-year-olds, their overt pointing behavior should mirror the patterns observed in their eye fixations.

Figure 1: Sample array for Experiment 1.

Method

Participants

The final sample consisted of fifteen 5-year-olds (9 males; M = 5.46 years, SD = .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 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. broken 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 will use the terms "intact" and "broken" throughout to differentiate the objects on critical trials that are paired with positive vs. negative vocal affect. The third image was always an unrelated distracter object (e.g., toy star; see Figure 1). Distracter 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 those 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 ps $< .01$).

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 behavior.

The experiment was implemented using E-Prime software with Tobii extensions. At the beginning of each session, the child participant'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 three seconds without auditory stimuli, followed by a blank black screen for two seconds. 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 ten minutes 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 pre-noun 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 1000 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 1000 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.

Pre-noun 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. Note that this measure differs from the target advantage score used in the analyses of the noun interval that follow because the available linguistic information has not yet provided any reason to focus on the eventually-named target.

We first calculated the average proportion of fixations to the broken object within a 100 ms time interval centered on the beginning of the pre-noun interval, and within an interval of the same duration centered on the end of the pre-noun 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 pre-noun 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 less 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 df 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, $F\,(1, 14) = 1.17$, $\eta_p^2 = .08$, $p = .30$. These results suggest that, in this contextual situation, vocal affect cues encountered early in the utterance have little effect on referential expectations.

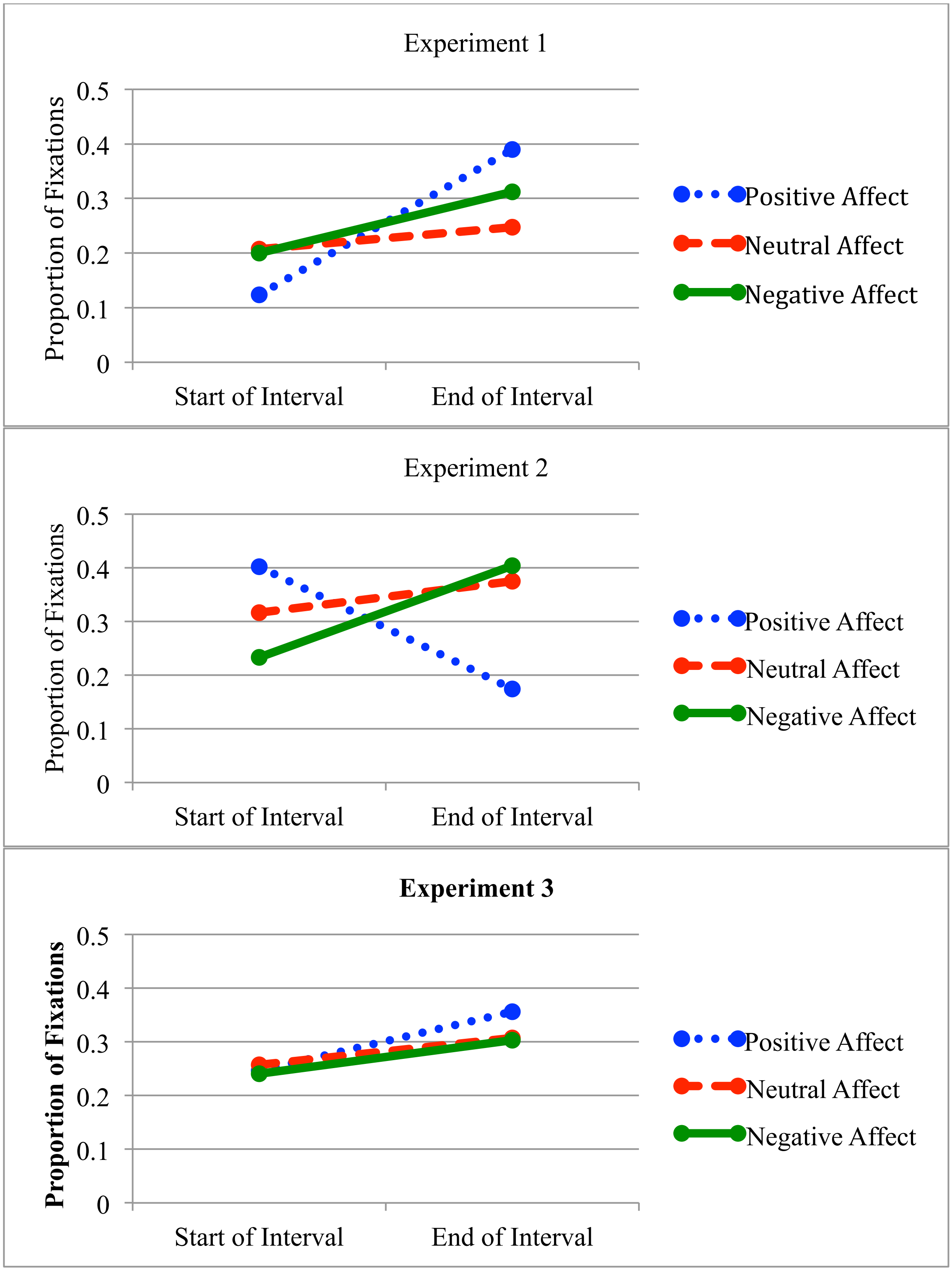


Figure 2. Change in the mean proportion of fixations to the odd-man-out object in the pre-noun interval across vocal affect conditions.

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 behaviors 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 680ms 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 550ms after noun onset.

To provide a statistical analysis of how children's looking behaviors 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, &

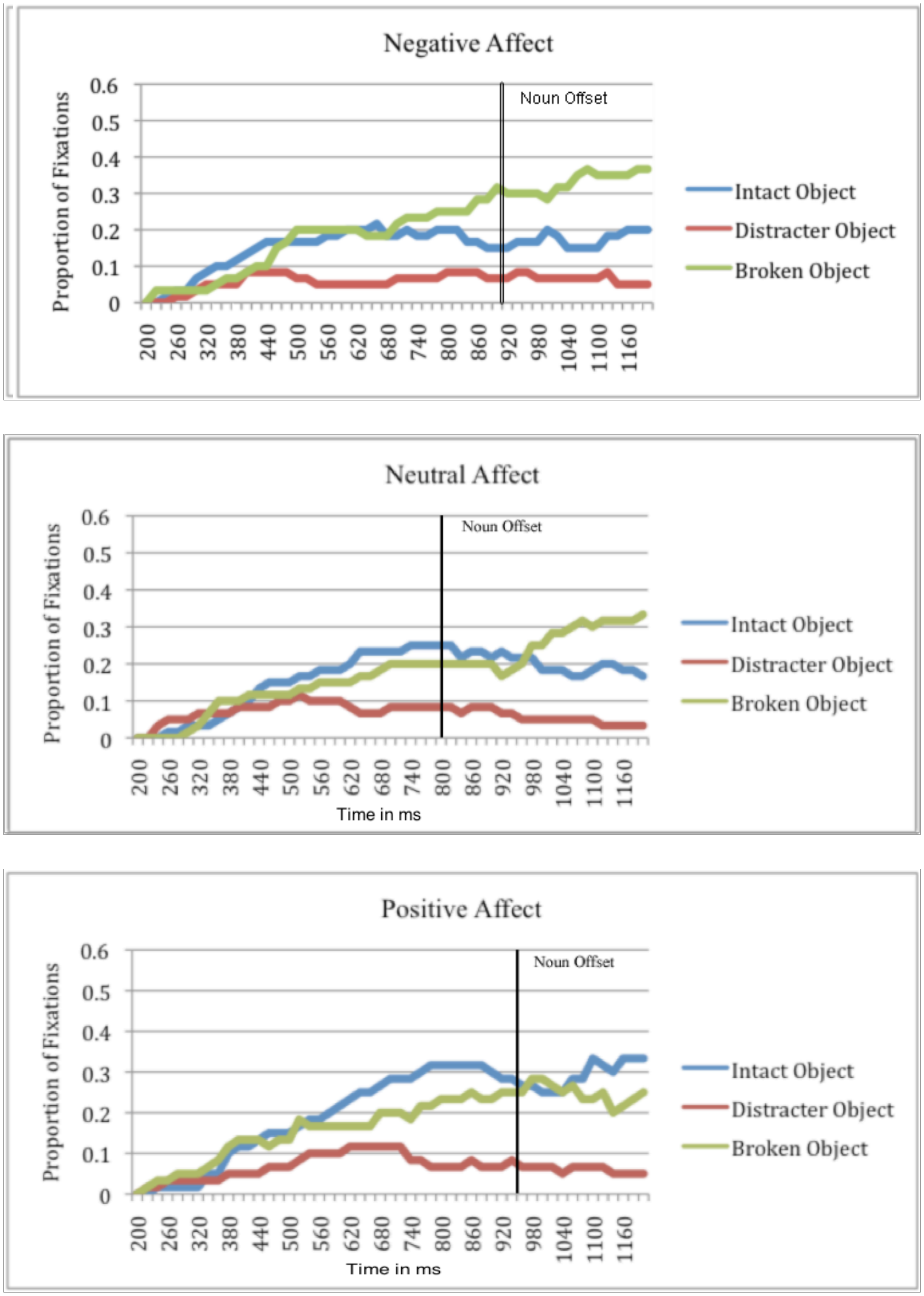


Figure 3. Proportion of fixations to display objects within the noun interval for Experiment 1.

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: $F\ (1, 14) = 6.23$, $\eta_p^2 = .31$, $p < .05$), 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 Behaviors

Children's pointing behavior 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 inter-rater reliability. Inter-rater reliability was excellent (Cohen's Kappa = 0.96; 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 distracter 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.

Table 1

Experiment 1: Mean percentage of points (and standard errors) to each of the referents as function of affect type

Vocal Affect Condition	Object	
	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)

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, as well as at an explicit level, as evidenced by children's pointing behavior 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 behavior 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 earlier work and Experiment 1 is related to the ambiguous scenario, children may be more likely to show earlier sensitivity to vocal affect earlier 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 vs. negative affect can influence children's real-time referential hypotheses.

Experiment 2

In Experiment 2, the arrays on critical trials consisted of one broken distracter 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 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.

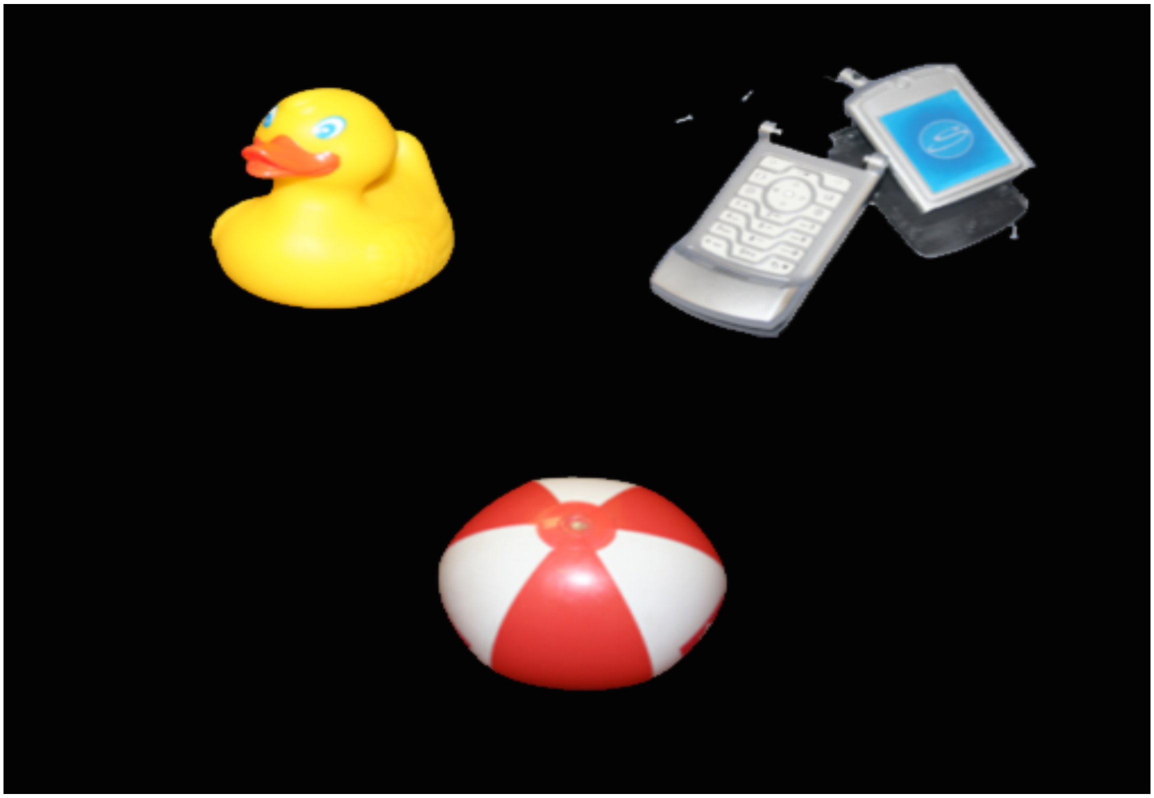


Figure 4. Sample array for Experiment 2.

Method

Participants

Fifteen 5-year-olds (7 males; *M* = 5.56 years, *SD* = .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 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 & Procedure

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

Results

Eye Fixation Patterns

Pre-noun 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 pre-noun 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 pre-noun 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 pre-noun 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 used in Experiment 1, beginning 200ms after noun onset and ending at 1200 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, η_p^2 = .144, p = .148). This outcome, and in particular that 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.

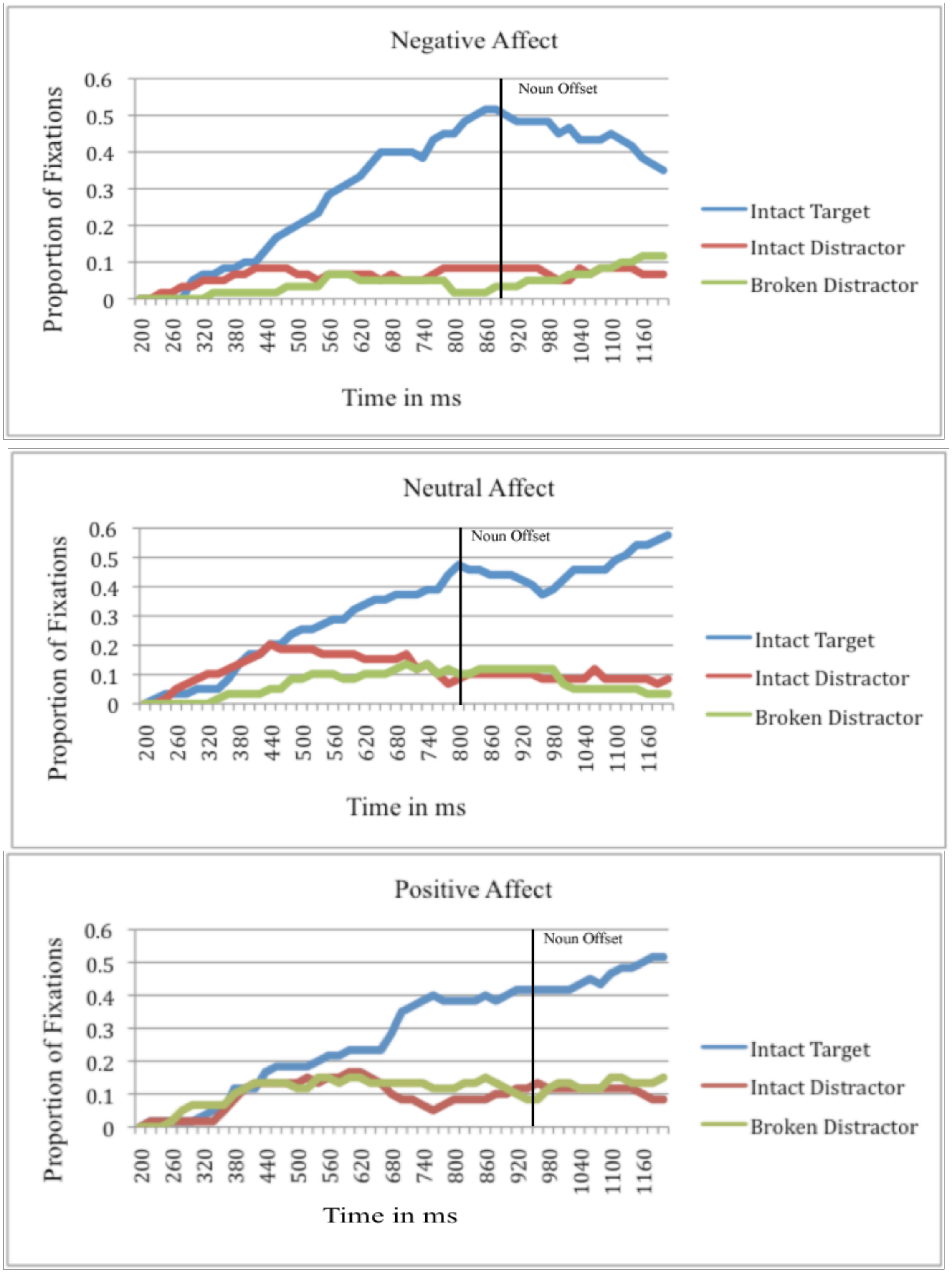


Figure 5. Proportion of fixations to display objects within the noun interval for Experiment 2.

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 Behaviors

Inter-rater reliability for 20% of the data ($n = 3$) was perfect (Cohen's Kappa = 1.00; $p < .001$). Children's pointing behavior 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 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 distracter object such that it would be more likely to be associated with positive affect. For example, we used 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* = .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 removed from the final sample due to insufficient data collected from the eye tracker.

Figure 6. Sample array for Experiment 3.

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. Same as 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 & Procedure

Identical to Experiments 1 & 2.

Results

Eye Fixation Patterns

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

Pre-noun interval. As before, we calculated the change (from the beginning to the end of the pre-noun 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 pre-noun 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 positive-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) = .246, η_p^2 = .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 used in Experiments 1 and 2, beginning 200 ms after noun onset and ending 1200 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 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 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 distracter 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 distracter object associated with the opposite affective valence within the depicted speech interval. These scores were submitted to the same ANOVA model used for the pre-noun measures. This analysis yielded a significant linear effect of vocal affect, F (1, 14) = 4.863, η_p^2 = .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.

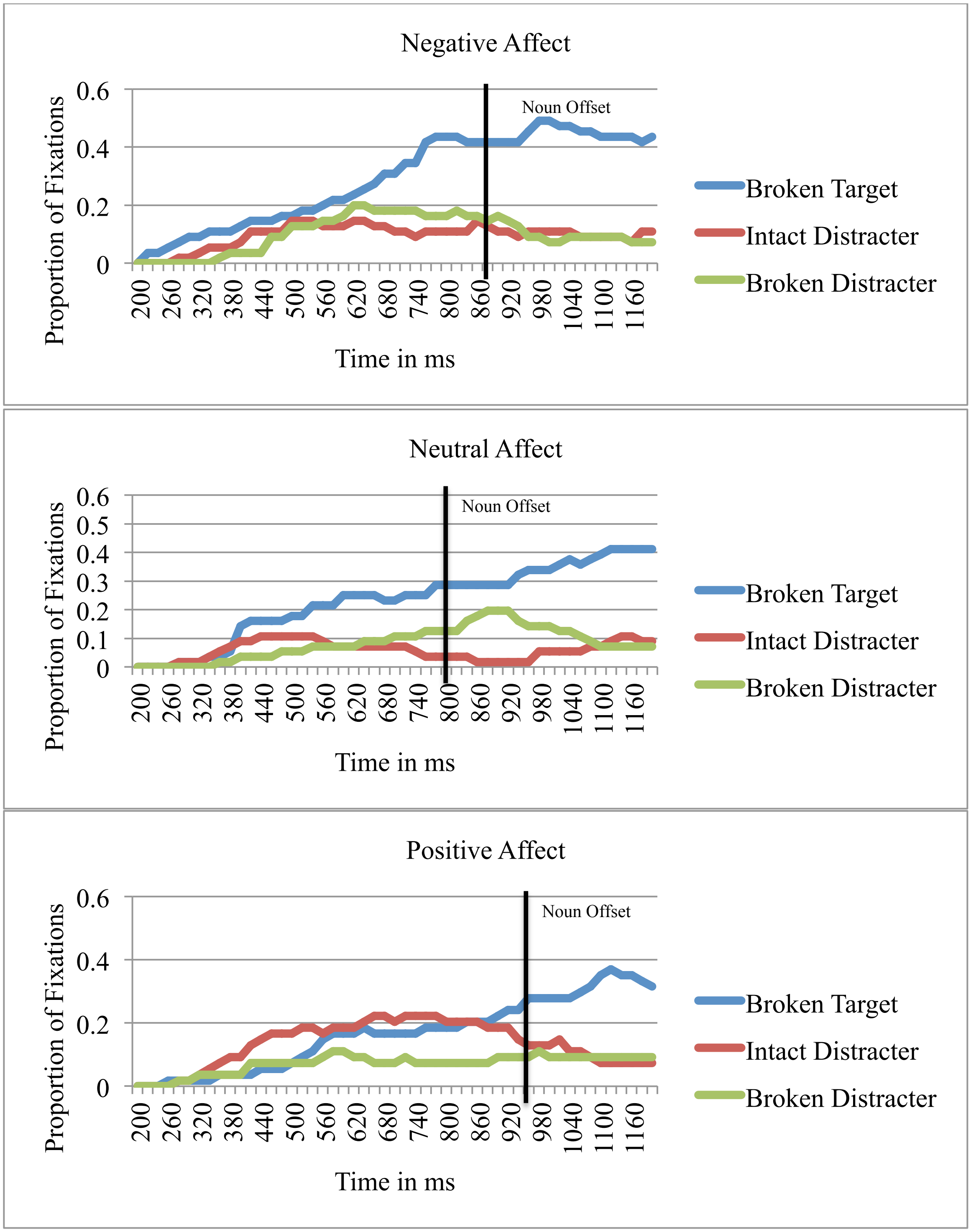


Figure 7. Proportion of fixations to display objects within the noun interval for Experiment 3.

Pointing Behaviors

Inter-rater 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 behavioral decisions and in their eye gaze behavior. Consistent with previous research using the same experimental scenario with younger children (i.e., Berman et al., 2011), 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 compared to 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 found in their fixation patterns.

In partial contrast to Experiment 1, the results of Experiment 2 (using unambiguous referential 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 pre-noun 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 distracter 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 compared to 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 compared to 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 compared to 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 then 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 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 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 & Trehub, 2001; Morton & Munakata, 2002; 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, in press; 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.

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 theee, 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. (in press). Preschoolers' use of talker information in on-line comprehension. *Child Development*.

Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass*, *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). 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.

The following chapter is a reproduction of a published work, for which permission has been granted by the publishers, John Wiley and Sons, (Licence #3150360971663) for the use in this dissertation. I was the primary investigator and main contributor for the publication titled, Preschoolers use emotion in speech to learn new words, in the journal *Child Development*, and was responsible for the conception, design, data collection and analysis with Dr. Craig Chambers and Dr. Susan A. Graham acting as my supervisors. Mr. Dallas Callaway was also responsible for some of the data collection and material development for Experiment 1.

Berman, J. M. J., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to learn new words. *Child Development.* *doi: 10.1111/cdev.12074*

Chapter 3: Preschoolers use emotion in speech to learn new words

"*It's not what you said, but how you said it.*" As illustrated by this familiar statement, the interpretation of an utterance is shaped by many factors beyond the linguistic content of the utterance itself. The very same words can convey different meanings depending on the intentions of the speaker, which are often partially conveyed through the *paralinguistic* aspects of a spoken utterance (i.e., intensity, voice quality, and nonlinguistic aspects of intonation and rhythm). One potentially powerful means by which speakers can signal communicative intent is through the emotional cues conveyed in speech. In two experiments, we examined whether 4- and 5-year-olds use emotional information available in the speech stream to learn novel words.

Young word learners are often faced with indirect learning contexts such as when a novel noun is not accompanied by an explicit demonstration of the word's referent (e.g., Akhtar, 2004; Akhtar, Jipson, & Callanan, 2001; Akhtar & Tomasello, 1996; Lieven, 1994). Yet, even in these situations, children are remarkably adept at isolating the intended referent from amongst many possible candidates, and establishing an appropriate word-concept mapping. To explain children's apparent success at this task, researchers have examined the diverse cues that are available to word learners (see Akhtar & Tomasello, 2000; Bloom, 2000; Diesendruck, 2007 for reviews). One powerful set of nonlinguistic cues involves children's ability to infer referential intent from speaker-produced behaviors (e.g., Moore, Angelopoulos, & Bennett, 1999). For example, Baldwin (1991, 1993) demonstrated that 18- and 19-month-olds, when presented with a novel word by a speaker, follow the direction of the speaker's gaze to an object and subsequently link the novel word to that referent. Other research has shown that children use facial gestures as behavioral markers of referential intentions (Graham, Stock, & Henderson, 2006; Henderson & Graham, 2005; Tomasello & Barton, 1994; Tomasello, Strosberg, and Akhtar, 1996).

One potential cue that has received comparatively little empirical attention in the word learning literature is the way in which a speaker's emotional state or disposition is conveyed via paralinguistic cues such as pitch level, pitch contours, and speech rate (see Banse & Scherer, 1996; Bachorowski & Owren, 1995; Frick, 1985). This informational cue, which we will refer to as vocal affect, is communicated in parallel with the linguistic content of an utterance and often reinforces linguistic information. However, when linguistic information alone does not allow the listener to unambiguously infer the intent of the speaker, vocal affect can serve as a primary cue to clarify a speaker's intended meaning. For example, the statement *"My mother is coming to visit"* can communicate something very different if spoken in an exhausted-sounding voice versus a happy-sounding voice. Studies of on-line processing in adults have shown these vocal affect cues are rapidly combined with linguistic meaning as language is encountered (e.g., Nygaard & Lunders, 2002; Nygaard & Queen, 2008; Paulmann, Titone & Pell, 2012).

Given that vocal affect is pervasive in communicative exchanges, it could function as a highly effective referential cue in the word learning situations faced by young children. Indeed, both attention to and responsiveness to vocal affect emerge early in development. Young infants show a strong preference to listen to infant-directed speech, whose prosodic features typically reflect positive affect, over adult directed speech (Cooper & Aslin, 1990; Fernald, 1985; Singh & Morgan, 2002). Infants also respond to a speaker's vocal affect in appropriate ways. That is, prelinguistic infants respond more positively to messages produced with positive vocal affect than those produced with negative vocal affect (Fernald, 1993). Similarly, 12-month-olds will decrease their exploration of a toy after hearing fearful-sounding speech versus neutral sounding speech (Mumme & Fernald, 2003).

When combining linguistic and paralinguistic information during spoken language processing, children exhibit changing sensitivities to vocal affect throughout the preschool and school-age years. By 4-years of age, children can accurately judge a speaker's emotional state based on vocal affect cues if the linguistic content is either neutral or filtered out (Friend, 2000, Morton & Trehub, 2001; Morton, Trehub, & Zelazo, 2003; Nelson & Russell, 2011; Quam & Swingley, 2011). However, young children are less sensitive to vocal affect cues when these cues conflict with propositional or linguistic content (Friend, 2000; Morton & Trehub, 2001). For example, when children are presented with a sentence that describes a sad event (e.g., "I dropped my ice cream cone") spoken with happy affect, they will rely on the linguistic content, judging the speaker to be sad. This focus on linguistic content over paralinguistic cues persists even when children are given explicit instructions to attend to vocal affect (Morton & Munakata, 2002). However, a recent study suggests that the prioritization of linguistic content over paralanguage stems from a difficulty in managing conflicting emotional cues, rather than a bias to prioritize linguistic content per se (Waxer & Morton, 2011). More specifically, Waxer and Morton propose that children's difficulty in these types of tasks stems from the assertion that happy and sad emotions are mutually inhibitory and therefore overburden children's cognitive abilities and cause their apparent inflexibility.

Although research on children's resolution of conflicting linguistic and paralinguistic cues has been informative, recent work has also examined the more naturalistic situation where linguistic content is simply indeterminate, rather than at odds with vocal affect cues. Berman, Chambers, and Graham (2010) presented 3- and 4-year-olds with arrays of three objects that included two objects of the same kind whose physical state differed (e.g., a broken and an intact doll). On critical trials, children were directed to look at one of the objects using a referentially ambiguous utterance (e.g., "Look at the doll") that varied in speaker vocal affect (negative, neutral, positive). Of particular interest was children's ability to draw on vocal affect cues in a dynamic manner as the utterance unfolded in time. Results indicated that four-year-olds' eye gaze patterns demonstrated appropriate sensitivities to vocal affect information. For example, they were likely to immediately shift their gaze toward the broken object upon hearing the ambiguous description when it was paired with negative vocal affect, less likely with neutral vocal affect, and finally even less likely with positive vocal affect. This pattern was not observed with 3-year-olds, suggesting that a developmental progression in the use of affect cues for speech understanding occurs within this interval. Even 4-year-olds, however, did not show clear evidence of fully adult-like sensitivity to vocal affect in their overt and final selection of a referent (as assessed by pointing), suggesting that children at this age are still at a transitional stage in the ability to coordinate speech affect cues with linguistic information.

To summarize then, from 4 years of age, children show increasing sophistication in the ability to use vocal affect cues to contour or augment the meanings conveyed by linguistic information. Although the findings to date suggest this is still only a nascent ability in 4-year-olds, sensitivity to vocal affect might be sufficiently robust to help guide the process we focus on here, namely, the acquisition of novel words. In Experiment 1, we examined 4- and 5-year-olds' sensitivity to a speaker's vocal affect as a cue to identifying the referent of a novel word. We focused particularly on whether children would detect the congruence between the emotion conveyed in a speaker's voice and characteristics of novel objects whose state is observed to change in a particular way. If children correctly identify the link between vocal affect and a particular object, this linkage could assist them in identifying that object as the referent of a novel noun. In Experiment 2, we examined whether 5-year-olds would extend and generalize labels that were newly acquired via vocal affect.

Experiment 1

The goal of Experiment 1 was to establish whether 4- and 5-year-olds would use a speaker's vocal affect to establish an initial mapping of a word. On critical trials, preschoolers were first presented with a pretest scene depicting two novel objects in their original states (e.g., a dog toy and a drink lid; see Figure 1, Phase 1), accompanied by an utterance directing them to look at the objects (e.g., "Look, look at these!", spoken with neutral vocal affect). Next, the same two objects were each presented in an altered state, where the changes in question were expected to increase the objects' association with either negative or positive vocal affect (i.e., one object was now broken and the other was decorated through a change in pattern). This method helps address the fact that, for familiar objects presented in a static manner, concept-specific knowledge can influence whether positive or negative valence should be associated with a given object. For example, although in many cases an object observed to be "in pieces" is considered

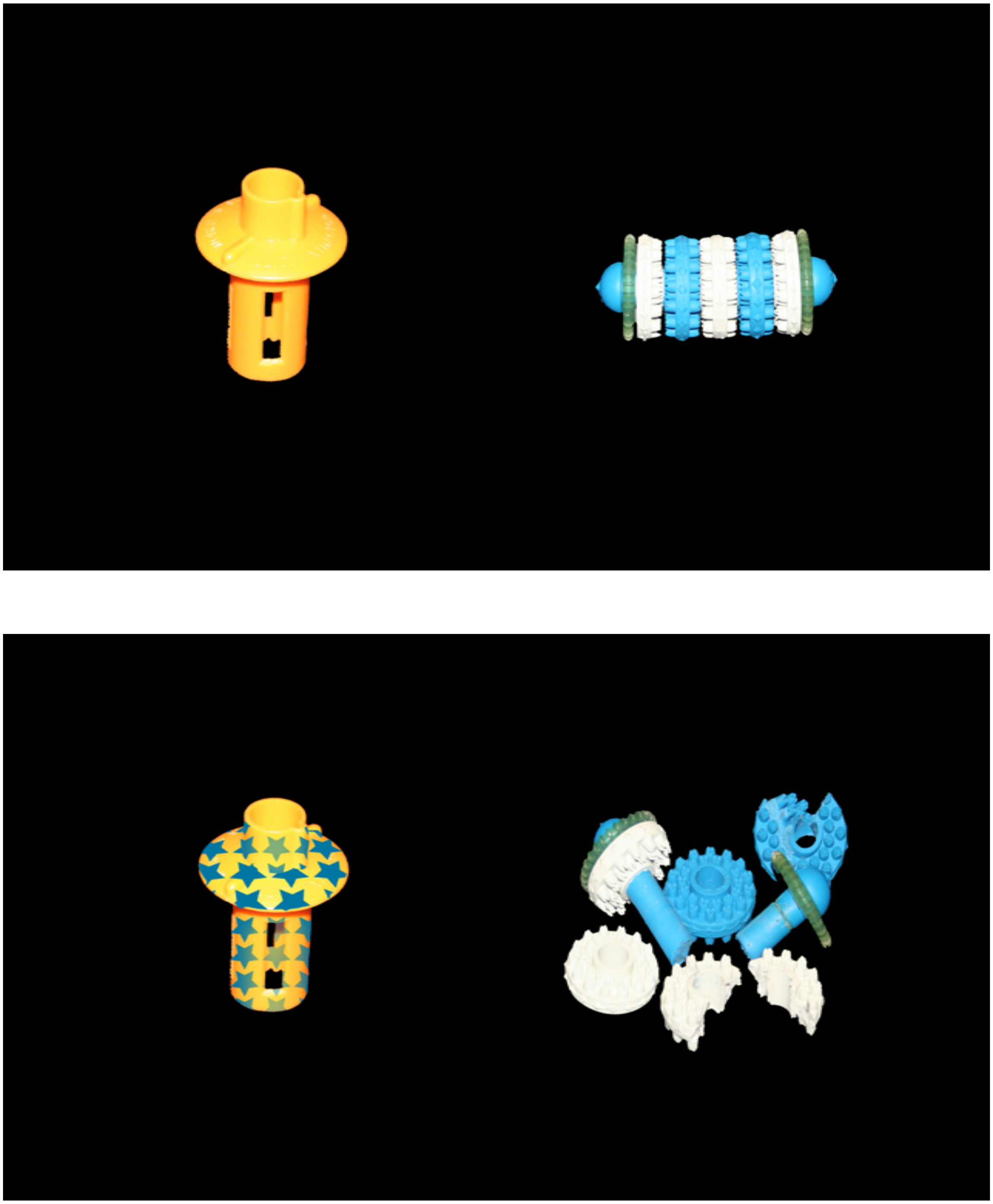


Figure 1. Sample of Phase 1 (top panel; familiarization) and Phase 2 (bottom panel; mapping) object displays for a critical trial in Experiment 1.

broken and therefore has negative valence, this would not be the case for certain toys like Lego blocks. By presenting unfamiliar objects in "before" and "after" states, it is possible to maintain more control over the intended interpretation.

Preschoolers were then directed to look at a referent using a novel word (e.g., “Look at the *gonlet.* Point to the *gonlet.*”), spoken using one of three different vocal affects (negative, neutral, positive). In order to precisely establish how preschoolers might rely on vocal affect as a cue to learn new words, we evaluated preschoolers’ responses using two different measures. First, children’s points toward a selected object were recorded to capture the final product of the referential decision process. Second, eye gaze measures were used to quantify the unconscious and more graded consideration of referential alternatives as the speech stream unfolded in time.

If preschoolers can use vocal affect to infer the correct referent for the unfamiliar word, they should look more often at the broken referent than the enhanced referent in the negative vocal affect condition, and the opposite should occur in the positive affect condition. The neutral affect condition provided a baseline for measuring response behaviors when potentially helpful affect cues were absent. Comparisons across the 4- and 5-year-old age groups provide a means to examine developmental changes in children's interpretive patterns.

Method

Participants

Forty-eight English-speaking children were included in the final sample: 24 4-year-olds (12 males, 12 females; *M* = 4.69 years, *SD* = .24 years) and 24 5-year-olds (12 males, 12 females; *M* = 5.82 years, *SD* = .22 years). Participants were predominately

Caucasian and spoke English as their first language. An additional 36 children were tested but were excluded due to insufficient eye gaze data caused by excessive head or body movements (n = 16), experimenter error (n = 5), or failure to complete (n = 15).

Design and Stimuli

***Critical Trials*: ** Twelve critical trials were constructed, each involving a display showing two unfamiliar objects side-by-side (e.g., a dog toy and a juicer top; see Figure 1). On each trial, participants first viewed a pretest scene showing a pair of objects in their 'original' states (Phase 1, see top panel of Figure 1), accompanied by the utterance "Look, look at these!", spoken with neutral affect. This pretest was followed by a display in which the objects were now shown in a changed state (Phase 2, see bottom panel of Figure 1). The depicted changes were such that the objects would differ in their likelihood to be associated with either negative or positive vocal affect (i.e., an object might now appear to be broken, or might be enhanced by a change in pattern). To control for any differences in individual objects' potential to capture or maintain visual attention, the selection of which object was broken and which was enhanced in Phase 2 was counterbalanced across participants.

When the object properties were presented in their changed form in Phase 2, the child participant heard a recorded instruction referring to one of the novel objects. These utterances took the form "Look at the X…Point to the X" where 'X' was one of 12 novel bisyllabic words (e.g., *wugen*, *blicket*). Three versions of each critical instruction were recorded by a female native speaker of English, differing in terms of the emotional affect conveyed by the speaker's voice, and using adult-directed speech. *Neutral* sentences were recorded with correspondingly "neutral" intonational cues, whereas *positive* and *negative* versions were recorded with a distinctly happy or sad-sounding voice, respectively. See Table 1 for the average fundamental frequency range and standard deviation, duration, and amplitude for each type of utterance. Note that measurements are provided for the first instruction only ("Look at the X") as our analyses focus on eye gaze patterns during this instruction. In order to ensure that utterances were representative of the different affect conditions, we asked 25 adults (M = 21.56 years, SD = 4.90 years) to rate the critical utterances using a Likert-type scale (1 = negative, 4 = neutral, 7 = positive). Raters heard the utterances in randomized orders with the sound of piano keys inserted between the stimuli to control for any carryover or contrast effects. Adults rated the negative (M = 1.64, SD = .51), neutral (M = 3.77, SD = .30), and positive (M = 6.31, SD = .43) utterances as significantly different from one another in the expected direction (ps < .01 for all pairwise comparisons).

Table 1

Average acoustic measurements for utterances used in each condition.

Condition	Average Duration	Average Amplitude	F_0 Average Range[a]
Negative	1.69 s (0.13)	64.84 (1.57)	124.54
Neutral	1.32 s (0.13)	66.73 (1.19)	133.67
Positive	1.62 s (0.13)	73.71 (1.15)	327.91

Note: SD in parentheses

[a]Reflects highest F_0 in the utterance – lowest F_0 in the utterance

Four object pairs were assigned to each of the three affect conditions, yielding 12 critical trials. A given novel word was always paired with a particular object set. The pairing of object sets to affect conditions was cycled across groups of participants to ensure that each pair of objects was presented with a positive, negative, and neutral utterance. Note that a given child saw a specific object set only once. Thus, for some children, the enhanced object within a pair would be a candidate referent for the novel label in the positive affect condition while for other children, the broken object in that same pair would be the candidate referent in the negative affect condition.

Filler Trials: Twelve filler trials using novel and familiar objects were also included. The filler trials were included to prevent participants from developing implicit or conscious expectations about the overall structure of critical trials, which might in turn exaggerate or deplete the effect of interest. For example, the utterances accompanying the object displays were all recorded using neutral affect (see Appendix for on the full list of filler trials), ensuring that children could not simply attend to paralinguistic cues and ignore the linguistic content of the utterances. Linguistic meaning was also reinforced by including familiar objects with known names, paired with either another familiar object or an unfamiliar object. When familiar and unfamiliar objects were presented together, the accompanying instructions could refer to either the known object, using its conventional name, or the unfamiliar object, whose identity could be inferred via mutual exclusivity. In addition, any potential expectation that object pairs would contrast on the enhanced-broken dimension was addressed by including fillers in which the two members of the pair changed in the same way during Phase 2 of the trial (either broken or enhanced, counterbalanced across trials).

Overall, children identified the correct target 95.14% (*SD* = 15.97%) of the time on the filler trials. Filler trials were intermixed with critical trials and were presented in a computer-controlled random order for each participant.

Procedure

Testing sessions began with the participants seated in front of a 46-inch computer monitor. The auditory stimuli were presented via speakers that were located directly behind the monitor. Children's eye movements were recorded using a Tobii x50 eye-tracker. After a short calibration procedure, E-prime software (with Tobii Extensions) was used to initiate the first of 24 experimental trials. Each trial proceeded as follows: a pair of object photographs was presented on screen for 10 seconds to allow sufficient time for the participants to encode the visual stimuli. During this time, participants were instructed to look at the stimuli using an utterance spoken in neutral affect (i.e., "Look, look at these!"). Next, a blank black screen replaced the object display for two seconds, followed by the presentation of the objects in their changed form. The second display was accompanied by an utterance spoken in one of three vocal affect types for the critical trials (e.g., "Look at the *X*. Point to the *X*.") or in neutral affect for the filler trials. The second display remained on screen until participants pointed to one of the objects, at which point the experimenter advanced the program to the next trial.

Results

Our analyses focus on children's performance on the critical trials. Of interest was the extent to which preschoolers considered the broken versus enhanced object across the three affect conditions. We first describe children's pointing behaviors, and then move to a consideration of their eye gaze patterns.

Pointing Behaviors

Children's pointing behavior was used as a measure of their explicit ability to detect and use a speaker's vocal affect cues to identify and map a referent to the novel word. Children's points to display objects were coded from videotape by an assistant who was unaware of the experimental hypotheses and who had no information about the vocal affect condition corresponding to any given trial. A second assistant recoded 20% of the data (n = 10) to establish inter-rater reliability. Inter-rater reliability was excellent (Cohen's Kappa = .93; $p < .001$).

Table 2 shows the percentage of points to each of the referential alternatives as a function of affect type and age group. To provide a single measure capturing whether children's pointing reflected attention to vocal affect, we compared points to the broken object across vocal affect conditions and age groups. We summed the number of points to this object over the four trials included in each affect condition and calculated percentage scores. These data were submitted to an analysis of variance (ANOVA) model with a 1 df within-participants factor testing for the predicted pattern of linear trend among vocal affect conditions (negative > neutral > positive) and a between-participants factor for age group (4-year olds vs. 5-year-olds). 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). This analysis revealed a main effect of Affect Type, $F\,(1, 46) = 11.89$, $\eta_p^2 = .21$, $p < .001$, as well as a significant Affect Type x Age interaction, $F\,(1, 46) = 4.47$, $\eta_p^2 = .09$, $p < .05$, indicating that the two age groups differed in whether the predicted linear pattern was present. To explore the significant interaction, we carried out separate analyses for each age group testing for the expected pattern of linear change. These analyses indicated the effect was not significant for the 4-year-olds (p = .30) but was significant for the 5-year-olds, F (1, 23) = 12.75, η_p^2 = .36, p < .01. Specifically, 5-year-old children, but not 4-year-olds, were more likely to point to the broken object as the speaker's vocal affect became increasingly negative-sounding (see Table 2).

In the final analyses of the pointing data, we examined potential differences stemming from the valence of vocal affect (i.e., happy- vs. sad-sounding vocal affect) within the 5-year-old age group. Inspection of the means in Table 2 suggests that children were more successful in linking the negative vocal affect with the broken object than in linking the positive vocal affect with the enhanced object. To explore this possibility, we again used the percentage of broken-referent selections as the measure of reference, and compared this against chance level responding (50%). These analyses indicated that 5-year-olds pointed to the broken referent more often than would be expected by chance in the negative affect trials t (23) = 4.29, p < .001, and in the neutral affect trials, t (23) = 2.33, p < .03. In the positive affect condition, children's selection of the broken referent over the alternative referent did not differ from chance level performance (p > .87). The findings from the neutral affect condition suggest that children had a bias to point towards the broken object, perhaps due to its greater visual complexity. In addition, a review of the acoustic measurements in Table 1 indicates that our neutral speech sentences were more similar to the negative sentences than to the positive ones in terms of amplitude and pitch range. Thus, the tendency toward selecting the broken object in the neutral condition might also be influenced by the similarity of these speech characteristics.

Table 2

Percentage of points (and standard deviations) to each of the referents as function of affect type and age group

Vocal Affect Condition		Age Group	
		4-year-olds	5-year-olds
Negative			
	Broken	52.1 (29.4)	75.0 (28.5)
	Enhanced	47.9 (29.4)	25.0 (28.5)
Neutral*			
	Broken	67.7 (31.7)	65.6 (32.8)
	Enhanced	31.3 (32.3)	34.4 (32.8)
Positive			
	Broken	45.8 (30.0)	49.0 (31.7)
	Enhanced	54.2 (31.0)	51.0 (31.7)

**4-year-old data for the neutral condition do not sum to 1, as one participant did not point at either the broken or enhanced object during a neutral trial.*

In summary, the pointing data suggest that 5-year-olds, but not 4-year-olds, appropriately apply a speaker's vocal affect cues in their overt decisions about the referents of novel words. However, in view of the patterns observed in past research, it is possible that a more implicit on-line measure of children's ability to learn words using a speaker's vocal affect may capture sensitivities that are not reflected in children's explicit and final decisions. An effect of this type would be of particular interest in the case of the 4-year-olds, who seem insensitive to the vocal affect cues based on overt responses.

Eye Gaze Patterns

Fixations to display objects were calculated every 20 ms, and were aligned with selected landmarks in the speech stream. We analysed these data in a number of steps. First, we examined children's fixations to the novel objects when they were presented in their original states during Phase 1. This analysis provides a basic index of children's visual exploration of the two objects as they appear and as a general orienting instruction is heard. Recall that the presentation of these objects was accompanied by an instruction directing children to "Look, look at these!", produced with neutral affect. To capture children's general looking patterns across this period, we created a difference score reflecting children's tendency to fixate one object (arbitrarily designated the target) over the other (arbitrarily designated the competitor). Note that neither object was broken or enhanced during this phase. We then conducted one-sample t-tests comparing these scores to 0 (no difference in fixations to one object over the other). This analysis yielded no significant effects, *ps* > .22. Inspection of the target advantage means (Negative: M = -.001, SD = .13; Neutral: M = -.024, SD = .16 and Positive: M = .033, SD = .18) demonstrates that children looked relatively equally between the two objects, as evidenced by target advantage scores close to 0. Thus, children did not display a preference for one object over another during Phase 1.

Next, to examine children's real-time interpretation upon hearing the ambiguous noun on critical trials, we focus on eye gaze patterns during the first sentence accompanying the "changed "objects (e.g., *"Look at the blicket"*), and more specifically during the specific time interval when the novel noun was heard. The average length of the novel noun on critical trials was 840 ms. We identified an analysis interval beginning

200 ms after noun onset and ending 1040 ms after noun onset. (The 200 ms margin added to the onset and offset times represents the typical lag for the eyes to react to unfolding linguistic information in this experimental paradigm see, e.g., Allopenna, Magnuson, & Tanenhaus, 1998). Fixations initiated before the beginning of this interval were excluded from analysis to ensure that observed eye movement behaviors could plausibly be associated with the interpretation of noun information, rather than reflecting a continued bias to fixate a particular object that attracted attention at an earlier time point.

Figure 2 shows the proportion of fixations to the two types of novel object, calculated at each successive 20 ms time point across the noun interval. As mentioned above, eye movements initiated prior to the beginning of this interval were removed. For this reason, both gaze curves begin with a proportion of zero and then rise throughout the time interval as children begin to direct or redirect their gaze to one of the two objects. The degree to which fixation proportions rise for a given object indicates the extent to which that object is incrementally considered as the noun's speech sounds unfold in time.

Separate plots are shown for the positive, neutral, and negative affect conditions. (Age groups are not represented separately as subsequent statistical analyses revealed that age did not influence fixation patterns.) Note that the proportion of fixations to the broken and enhanced objects do not sum to 1 due to the possibility of fixating blank regions of the display.

Using the broken object as a reference point, it is clear that fixation patterns differ across the three vocal affect conditions. In the negative affect condition (top panel), the new eye gaze patterns initiated as the noun unfolds are increasingly more likely to reflect consideration of the broken referent compared to the enhanced object, beginning approximately 550 ms after the onset of the noun. Visual consideration of 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 show a stronger tendency to fixate the enhanced object over the broken object beginning around 550 ms after noun onset.

To provide a single statistical measure capturing whether children's gaze behaviors were modulated by the speaker's affect, we used a target advantage measure reflecting the relative tendency to fixate the broken referent over the enhanced object within the depicted speech interval. This was obtained by calculating a difference score (i.e., the average probability of fixating the broken referent minus the probability of fixating the enhanced referent) for each participant and each vocal affect condition, across the specific time interval described earlier, namely the noun interval. These scores were then submitted to the same ANOVA model used for the pointing measures. This analysis yielded a significant linear main effect of vocal affect (negative > neutral > positive: F (1, 46) = 14.29, η_p^2 = .24, p < .001). Neither the main effect of age nor the age by vocal affect interaction reached significance (*ps* > .32). The significant effect of vocal affect indicates that children in both age groups were consistently less likely to look at the broken referent as speech became increasingly positive-sounding across conditions. These analyses indicate that both 4- and 5-year-olds used the vocal affect cues to increase their consideration of an appropriate referent for a novel noun as this noun unfolded in time. For the 4-year-olds, these results stand in contrast to their pointing behavior, which did not show an appropriate sensitivity to vocal affect.

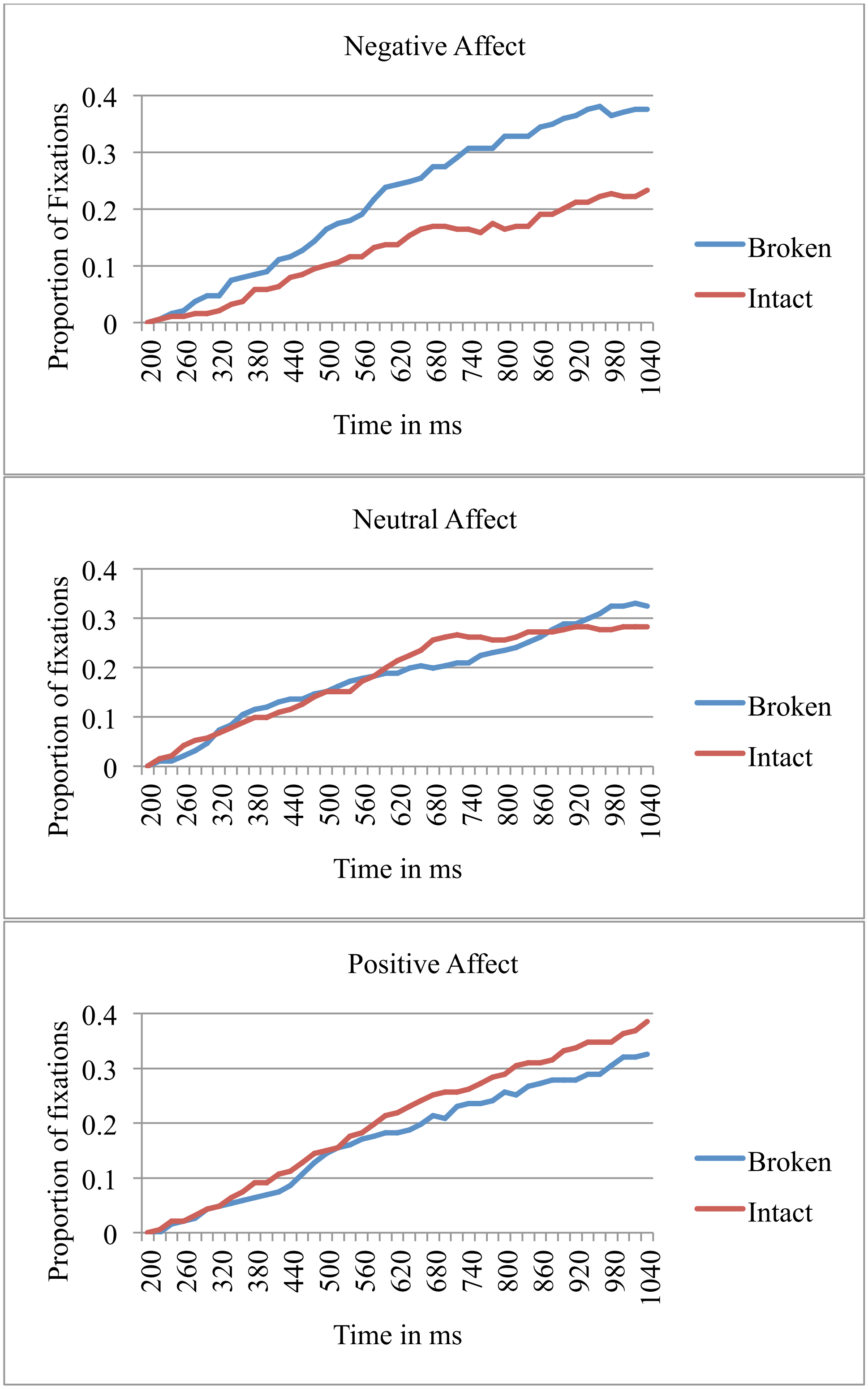


Figure 2. The probability of fixating display objects across the noun window (Expt. 1)

Finally, because the relevant speech cues are distributed across the entire utterance, and are not strictly localized to the noun window, it is relevant to consider whether children began to identify affect-compatible objects earlier in the utterance, before the noun was heard. That is, it is possible that children did not wait until the noun (whose syntactic environment would highlight it as an object-referring term) to begin identifying an appropriate display object in both the positive and negative affect conditions. To assess this possibility, we isolated a pre-noun interval beginning 200ms after onset of the first instruction and ending at 200 ms after noun onset (540 ms in total; See Appendix B for fixation profiles). The fixation profiles show a general preference to fixate the broken referent across affect conditions that remains stable across the pre-noun interval. This preference is perhaps not surprising given the greater visual complexity associated with this changed object (see Berman et al., 2010, for a similar result) and the previously reported bias to point to this object in the neutral affect condition.

To conduct a statistical evaluation of the data patterns, we again calculated a target advantage score reflecting the relative tendency to fixate the broken referent over the enhanced object within the pre-noun interval (540 ms). These data were submitted to the same ANOVA model used for the noun interval. The analysis revealed no significant effects or interactions (ps > .30). Thus, there were no reliable differences across conditions in children's consideration of the two candidate referents in the speech region before the noun was heard. This outcome clearly demonstrates that children's use of vocal affect cues upon hearing the noun is not the result of "carryover" effects stemming from vocal affect cues encountered earlier in the utterance.

Discussion

The pointing data suggested that 5-year-olds, but not 4-year-olds, successfully used vocal affect cues to map the novel label to the relevant state-changed novel object. At a more implicit level, the eye gaze data suggested that both the 4- and 5-year-old children were able to use vocal affect to link the correct state-changed object with the novel label. These results thus indicate that preschoolers will use vocal affect to establish mappings with objects that likely reflect a speaker's referential intentions -- a key ingredient in the process of word learning. It is possible, however, that preschoolers do not take the additional step of associating the novel label to an object concept. In other words, children might simply use vocal affect to direct their attention to a particular object, and not to build a linguistic-level association between a sound pattern and a meaning. One way to explore this possibility is to examine whether preschoolers will extend newly established word mappings to objects when they are returned to their original state (and no longer reflect the affect-relevant characteristics used in the initial mapping) and whether children can identify new exemplars of the same kind. In Experiment 2, we address this question, focusing on 5-year-olds.

Experiment 2

Given the evidence for 4-year-olds' somewhat fragile ability to use vocal affect to identify relevant referents, Experiment 2 focuses on 5-year-olds. Here, we assess whether 5-year-olds will extend or generalize the newly learned labels to (1) earlier-encountered objects returned to their original state and (2) new, unaltered exemplars of the same kind. To do this we first presented 5-year-olds with a task that required them to map the label to state-changed objects using either positive or negative affect (i.e., a replication of

Experiment 1). We next tested their ability to use the novel label when the objects had been returned to their original, unaltered state, and when sentences contained neutral vocal affect. Because neither the speech nor the objects reflect the cues used to deduce the speaker's referential intent in the initial mapping phase, this provides an important test of the association between the newly-learned word form and its intended meaning. Finally, we then tested children's ability to generalize the newly-learned label to a novel exemplar of the same kind, again using sentences with neutral affect. Generalization tests of this sort are regularly used to gauge the extent to which a word-concept association has been firmly established.

We did not include neutral affect trials in this experiment. This was because we wished to avoid situations where children would become frustrated, confused, or uncooperative due to the successive presentation of referentially ambiguous instructions (e.g., in the three phases of a given trial in this condition).

Method

Participants

Fifteen English-speaking children were included in the final sample: (8 males, 7 females; M = 5.43 years, SD = .33 years). Participants were predominately Caucasian and spoke English as their first language. An additional 7 children were tested but were excluded from the sample due to insufficient data collected from the eye tracker.

Design and Stimuli

Critical Trials: The 6 critical trials each consisted of 3 phases: a mapping phase, an extension phase and a generalization phase, each of which is described in detail below. Three critical trials were presented in each affect type: negative and positive.

***Mapping Phase*: **This phase was an exact replication of that used in Experiment 1 with one exception: no neutral affect trials were included. Thus, children first saw the objects in their original states, accompanied by general attention-directing phrases spoken in neutral affect (e.g., “Look, look at these!”). Children then saw the objects in their altered states (i.e., enhanced or broken) accompanied by instructions directing them to find the referent of a novel noun. These sentences occurred with either negative or positive sounding vocal affect.

Extension Phase: Directly following the mapping phase, participants were presented with a black screen and heard the instruction, “Now we are going to find the X again. Let’s find the X” spoken with neutral affect (where X was the novel noun used on the mapping trial). The objects then appeared on the screen in their original (unaltered) states but on different sides of the screen than they appeared in the mapping phase (see top panel of Figure 3). Here, the instruction directed children to find the referent of the novel noun, using neutral affect (“Where’s the X now? Point to the X”).

Generalization Phase: Following the extension phase, children were presented with a black screen and heard the instruction, “Now we are going to find another X. Let’s find another X” spoken with neutral affect. Next, two new exemplars (i.e., different colored versions) of the original objects appeared on the screen (see bottom panel of Figure 3) and children heard, “Can you find another X? Point to the X” spoken with neutral affect. This entire sequence was repeated for all six object sets on critical trials.

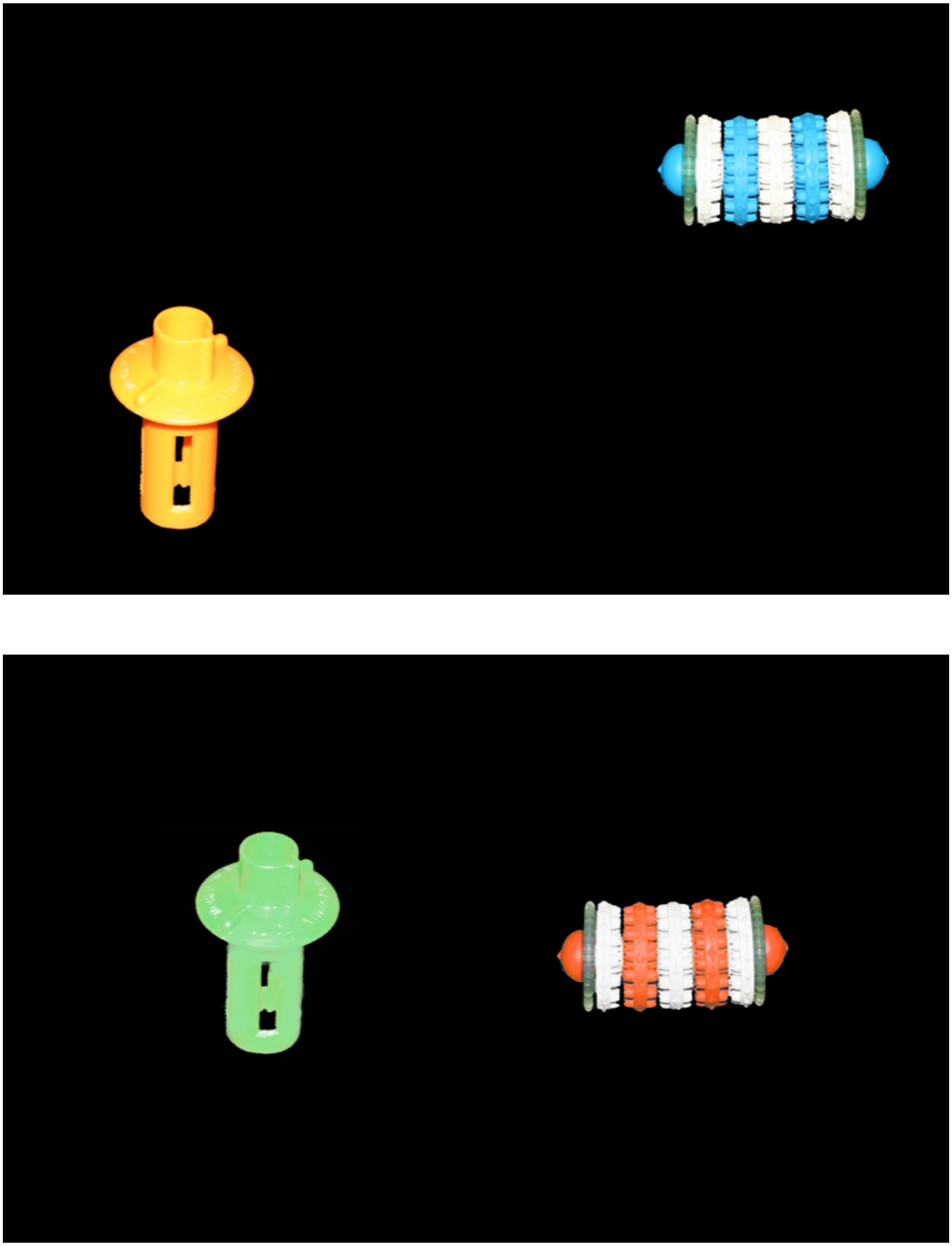


Figure 3. Sample of Phase 2 (extension: top panel) and Phase 3 (generalization, bottom panel) object displays for a critical trial in Experiment 2.

Filler Trials: Six filler trials using novel and familiar objects were also included (See Appendix A). Overall, children identified the correct target 92.21% (*SD* = 13.90%) of the time on the filler trials. Critical trials and fillers were presented in an interleaved format where each critical trial (with the three phases described above) was alternated with a subsequent filler trial. In addition, the critical trials were ordered such the affect type (positive or negative) was alternated over the course of the experiment.

Procedure

The testing procedure was identical to Experiment 1, except twelve trials rather than twenty-four trials were presented.

Results

Pointing Behaviors

Children's pointing behavior was again used as a measure of their explicit ability to detect and use a speaker's vocal affect cues to identify and map a referent to the novel word. Children's points to display objects were coded from videotape by an assistant who was unaware of the hypotheses and who had no information about the vocal affect condition corresponding to any given trial. Inter-rater reliability for 20% of the data (n = 3, coded by a second assistant) was excellent (Cohen's Kappa = .97; $p < .001$).

First, we calculated the percentage of trials where children pointed to the target object. Target objects were considered the object that 'matched' the vocal affect of the speaker during the mapping trials (i.e., the broken object for the negative affect trials and the 'enhanced' object for the positive affect trials). We summed the number of points to this object over the three trials included in each affect condition and calculated percentage scores (see Figure 4). Results of paired-samples t-tests indicated that children pointed to the target object significantly more in the negative affect condition than in the positive affect condition across all three tasks. First for mapping trials, negative trials ($M = .87$, $SD = .25$) were significantly different from positive trials ($M = .51$, $SD = .40$), $t(14) = 2.31$, $p < .05$, $d = 1.08$. The same pattern emerged for extension trials, negative trials ($M = .78$, $SD = .33$) were significantly different from positive trials ($M = .44$, $SD = .30$), $t(14) = 2.24$, $p < .05$, $d = 1.07$. Lastly, for generalization trials, negative trials ($M = .84$, $SD = .25$) were significantly different from positive trials ($M = .49$, $SD = .33$), $t(14) = 2.62$, $p < .05$, $d = 1.22$. Thus, children pointed to the affect-congruent target object significantly more in the negative affect condition than in the positive affect condition across all three tasks.

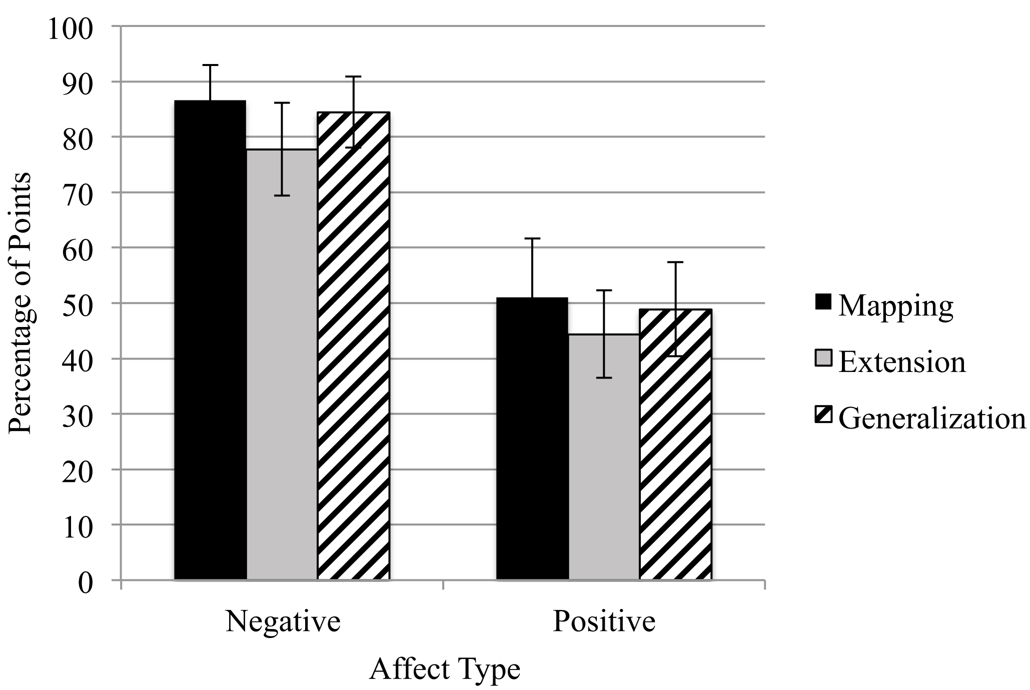


Figure 4. Experiment 2: Percentage of points to target object as a function of affect type and task.

We next compared target selections to chance level responding (50%). These analyses indicated that children pointed to the broken object on the negative affect trials at rates significantly above chance for all three trial types (*ps* < .01). In contrast, in the positive affect condition, children's pointing did not differ significantly from chance on any of the trial types (*ps* > .48) indicating that children pointed equally to the broken and enhanced objects.

The data so far suggest that 5-year-olds are able to successfully use negative-sounding vocal affect to map, generalize and extend a novel label to a novel object, but are unable to use positive vocal affect to successfully complete the same task. Further, comparisons with the performance of 5-year-olds in Experiment 1 show fully consistent patterns. Specifically, the percentage of points to the broken object in the negative vocal affect condition of Experiment 1 (75%) and in the corresponding condition for the mapping phase in Experiment 2 (87%) are both comparatively high. In contrast, the percentage of points to the enhanced object in the positive vocal affect condition of Experiment 1 (51%) and in the same condition for the mapping phase in Experiment 2 (56%) are both at chance level. However, given the results from Experiment 1, it is possible that the implicit and more fine-grained measures of behavior provided by eye gaze will reveal evidence of children's use of positive vocal affect.

Eye Gaze Patterns: Mapping phase.

Similar to Experiment 1, we were most interested in the time period covering children's real-time interpretation upon hearing the ambiguous noun on critical trials. We present gaze data only for the mapping phase as children often pointed before hearing the instruction on both the extension and generalization phases, which resulted in a loss of eye movement data because the child's arm would occlude the eye camera. Recall that children received an additional prompt (i.e., "Now we are going to find another *blicket…*), before the objects appeared on the screen. It is perhaps unsurprising that children did not wait until the end of the subsequent prompts to make their final selection.

For the mapping phase, we examined eye gaze patterns during the first utterance (e.g., *"Look at the blicket"*). As before, we first focused on the noun region. Following the procedure used in Experiment 1, we excluded fixations initiated before the beginning of this interval to ensure that gaze behavior could plausibly be associated with interpretation at the time of the noun, rather than reflecting a continued bias to fixate a particular object. Figure 5 shows the proportion of fixations to the two types of novel objects, calculated at each successive 20 ms time point across the noun interval. Separate plots are shown for the positive and negative affect conditions. Note that the proportion of fixations to the broken and enhanced objects do not sum to 1 due to the possibility of fixating blank regions of the display.

Using the broken object as a reference point, it is clear that the fixation patterns reveal differences in eye patterns between the two vocal affect conditions. In the negative affect condition (top panel), gaze is more likely to be directed toward the broken referent instead of the enhanced object, beginning approximately 400 ms after the onset of the noun. This trend reverses in the positive affect condition (bottom panel), where children now show a stronger tendency to fixate the enhanced object over the broken object beginning approximately 320 ms after noun onset.

To provide a single statistical measure capturing whether children's looking behaviors were modulated by the speaker's affect, we calculated the same target advantage measure as Experiment 1. Recall that this difference score reflects the relative tendency to fixate the broken referent over the enhanced object within the depicted speech interval. Results of a paired-samples t-test indicated that children were using vocal affect cues to increase their consideration of the appropriate referent for a novel noun, t (14) = 2.64, p = .02, d =.68. That is, children were more likely to fixate the broken object over the intact object in the negative affect condition (M = .12, SE = .04) and were less likely to fixate the broken object (i.e., more likely to fixate the enhanced object) in the positive affect condition (M = -.11, SE = .07).

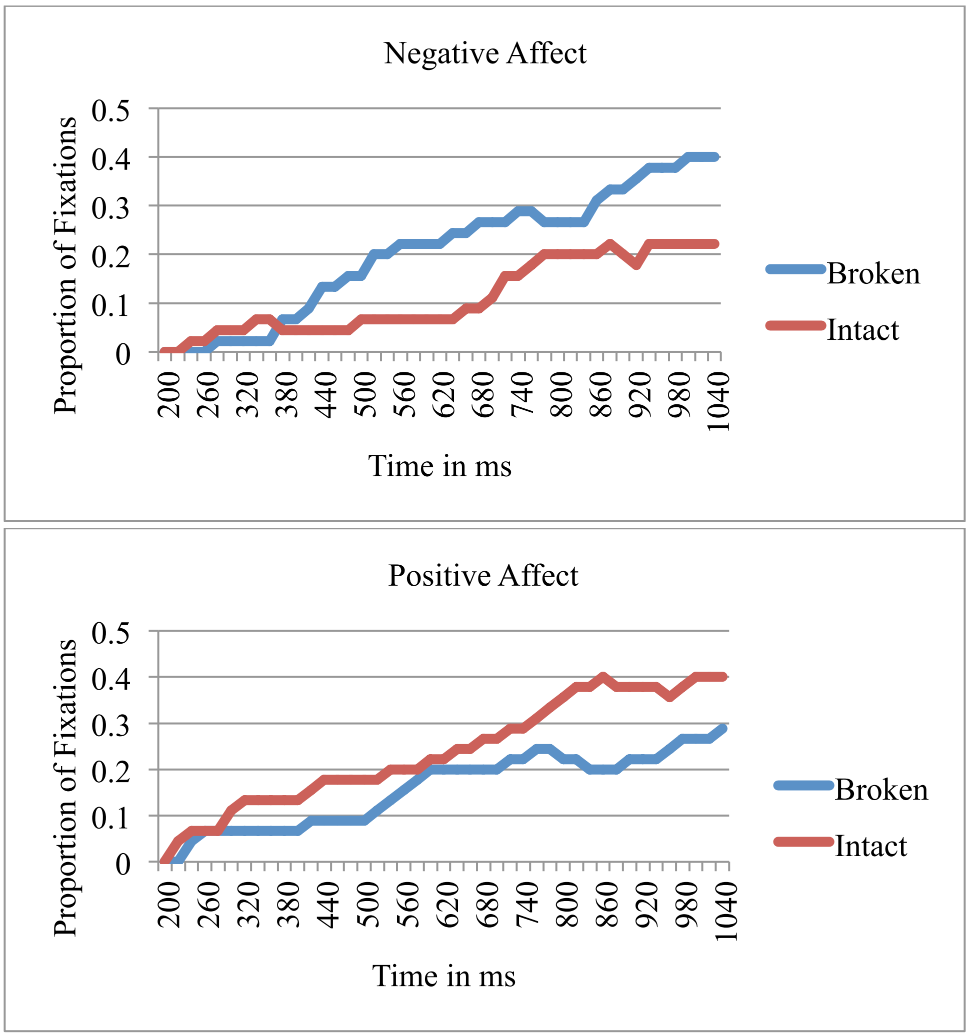


Figure 5. The probability of fixating display objects across the noun window (Expt. 2).

Second, we examined eye movements during the pre-noun region, isolating a pre-noun interval beginning 200ms after onset of the first instruction and ending at 200 ms after noun onset (540 ms in total; See Appendix C for fixation profiles). The fixation profiles show a general preference to fixate the broken referent in both affect conditions that remains relatively stable across the pre-noun interval. Target advantage scores did not differ significantly between the negative affect condition ($M = .16$, $SE = .10$) and the positive affect condition ($M = .28$, $SE = .05$), $p = .21$. Thus, there were no reliable differences in children's consideration of the two candidate referents in the speech region before the noun was heard as a function of affect type.

To summarize, the pointing data suggested that 5-year-olds successfully used negative, but not positive vocal affect cues to map, generalize, and extend a novel label to a novel object. In contrast, at a more implicit level, eye-gaze data for the mapping phase suggested that 5-year-olds could use both negative and positive vocal affect to map the correct state-changed object with the novel expression.

General Discussion

In two experiments, we examined preschoolers' use of a speaker's vocal affect to guide decisions about the referent of a novel word. Our findings offer a number of key insights into preschoolers' emerging ability to attend to a speaker's vocal affect in the presence of a novel word and two unfamiliar objects.

First, our findings demonstrate that 4- and 5-year-olds can use vocal affect to establish initial word-referent mappings with novel nouns. However, although the influence of this information is uniformly apparent in the children's implicit word-object mappings (as speech unfolds), it does not always affect their explicit decisions about the referent of a novel noun. In both experiments, 4- and 5-year-olds' gaze patterns suggest that they can use vocal affect to link a novel word to a novel object at an implicit level. However, only 5-year-olds demonstrated this sensitivity in their explicit selection of a referent while pointing.

The finding that children as young as 4-years-old show sensitivity to vocal affect during word-referent mapping is particularly remarkable in light of the complex reasoning required in the mapping phase of the task. That is, children first had to recognize changed objects as corresponding to the original exemplars seen earlier, correctly interpret the type of state change, concurrently classify the valence of the speaker's vocal affect, and match this information the objects' altered physical states. The final stage in reasoning is critical insofar as children had to integrate their appraisals of the emotional disposition of a speaker with characteristics of the objects in order to successfully learn a new word. Previous research has demonstrated that children as young as 18 months will use facial affect to gauge an actor's desires for a particular food (Repacholi & Gopnik, 1997) and that by 2 ½ years of age, children will use a speaker's facial and vocal affect to understand and recognize her associated emotion (Wellman, Phillips, & Rodriguez, 2000). Our findings suggest that children's ability to use expressed emotion as a contextual cue in more complex forms of behavior becomes apparent only later in time, namely when this cue is taken as reflecting the speaker's intention to refer to an object with an emotionally-congruent physical state.

Second, the results of Experiment 2 demonstrated that 5-year-olds used negative (although not positive) vocal affect to extend and generalize a newly learned label to objects returned to their original state (e.g., when they no longer possessed the properties that were linked with vocal affect cues). This finding clearly indicates that 5-year-olds were not simply linking the vocal affect to the state-changed object (when it did possess the relevant characteristics). Rather, they took the process one step further, creating a linguistic association between the novel label and the object. This in turn allowed the label to be used to refer to a different instantiation of the object, and in a context where affect cues were no longer present in the utterance. The generalization phase further demonstrated the robustness of the word-object association. Specifically, children were able to use the newly learned label to correctly identify a new exemplar that did not exhibit the physical features used to build the initial mapping. This outcome indicates that the linguistic association was created between a word and an object category, rather than just one specific object.

Turning now to the observed valence differences, the finding that 5-year-olds use negative but not positive affect to explicitly map, extend and generalize newly learned labels is consistent with the results of recent research using a variety of different tasks. Nelson and Russell (2011) demonstrated that preschoolers display an earlier emerging sensitivity to negatively-valenced affect than positively-valenced affect when asked to explicitly label a speakers' emotional state from vocal information. Berman, Graham, & Chambers (in press) demonstrated that 5-year-olds were more attuned to negative vocal affect than positive vocal affect, particularly early in an utterance, in contexts where no linguistic ambiguity was present. However, despite the similar pattern across studies, there are additional factors that could account for our observed differences between positive and negative vocal affect. For example, the earlier-mentioned bias toward the broken object may have resulted from the greater novelty of that object. If so, children in the positive-affect trials would have to overcome their curiosity about this object before redirecting their attention toward the positive-valenced object. It is also possible that the positive-valenced objects were not as obviously positive as the negative-valenced ones were negative in terms of their physical characteristics. Thus, the transparency of the object's association with a given affect type may have contributed to the observed differences between negative and positive vocal affect.

It is important to note, however, that the valence effects observed in the current study were more pronounced in explicit measures (pointing) than in implicit measures (eye gaze). Specifically, in both experiments, 4-and 5-year-olds spent more time fixating the enhanced object in the positive affect condition and spent more time fixating the broken object in the negative affect condition. Thus, an implicit sensitivity to positively and negatively-valenced affect may be detected earlier in development with implicit measures, revealing a clear pattern of continuity underlying children's increasing command of vocal affect cues. These findings are consistent with previous work in this area highlighting the "transitional" status of 4-year-olds in their coordination of vocal affect and linguistic information in studies using familiar nouns (Berman et al., 2010). Further, this consistent pattern suggests that studies that rely exclusively on children's ability to explicitly label emotions may underestimate children's capacity to detect emotion in vocal affect.

What might account for the seeming incongruity between children's looking and pointing behaviors? We first consider the 4-year-olds in Expt 1 who demonstrated sensitivity to vocal affect only in their eye gaze behavior. Here, one possibility is that the conscious integration of lexical and affective information that is required to behaviorally enact a final referential decision may have overwhelmed children's cognitive resources. Examples of children's overextended cognitive resources may be found in a number of investigations into children's developing capacities to combine the linguistic and affective aspects of speech (e.g., Friend, 2000; Morton & Trehub, 2001; Morton & Munakata, 2002; Morton et al., 2003). A related explanation for why children's pointing behavior was not always consistent with their eye gaze is that pointing requires the programming and implementation of complex motor movements whereas information-seeking behaviors such as gaze shifts are unconscious and have low metabolic demands (e.g., Abrams & Jonides, 1988). Thus, the apparent discordance seen between eye movements and pointing behaviors might arise from the motoric demands involved in pointing, and the greater opportunity for uncertainty to build as a result of their slower execution.

Second, we consider the 5-year-olds who demonstrated sensitivity to both positive and negative affect in their eye gaze, but only to negative affect in their pointing. As discussed above, there are various possibilities for why negatively-valenced speech may be easier for children to link to referential entities in these types of tasks (Berman et al., in press); therefore, if positive affect cues were detected in speech yet were accompanied by some degree of uncertainty, the observed pattern would be obtained. This dissociation could also signal that children at this age have representations of linguistically-relevant affect cues that are graded in nature (see Morton & Munakata, 2002; Munakata, 2001 for discussions of this approach). On this account, children would demonstrate sensitivity to speaker affect in their eye movements but not in their pointing, as successful pointing would require a stronger and more discrete representation.

In summary, our findings suggest that, by four years, children will use a speaker's vocal affect to identify the referents of new words. However, it is not until five years of age that children are able to translate these word mappings into a conscious referential decision, allowing newly-learned words to be extended and generalized.

References

Allopenna, P.D, Magnuson, J.S., & Tanenhaus, M.K., (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language. 38*, 419–439

Abrams, R. A., & Jonides, J. (1988). Programming saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 428-443.

Akhtar, N. (2004). Contexts of early word learning. In D. G. Hall, & S. R. Waxman, (Eds.). *Weaving a lexicon.* Cambridge, MA: MIT Press.

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development, 72,* 416-430.

Akhtar, N., & Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology, 14,* 79-93.

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M., Golinkoff et al., (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115-135). Oxford, England: Oxford University Press.

Bachorowski, J., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6,* 219-224.

Baldwin, D. A. (1991). Infant’s contribution to the achievement of joint reference. *Child Development*, *62*, 875-890.

Baldwin, D. A. (1993). Early referential understanding: Infant’s ability to recognize referential acts for what they are. *Developmental Psychology*, *29*, 832-843.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614-636.

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.

Berman, J. M. J., Chambers, C., & Graham, S. A. (*2013*). Contextual influences on children's use of vocal affect cues. *Quarterly Journal of Experimental Psychology, iFirst, 1-22.*

Bloom, L. (2000). The intentionality model of word learning: How to learn a word, any word. In R. M., Golinkoff et al., (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 19-50). Oxford, England: Oxford University Press.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*, 1584-1595.

Creel, S. C., & Bregman, M. R. (2011). How Talker Identity Relates to Language Processing. *Language and Linguistics Compass*, *5,* 190-204.

Diesendruck, G. (2007). Mechanisms of word learning. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development.* Blackwell: Malden, MA.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 10,* 181-195.

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

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.

Graham, S. A., Stock, H., & Henderson, A. M. E. (2006). Nineteen-month-olds' understanding of the conventionality of object labels versus desires. *Infancy*, *9*, 341-350.

Hertzog, C., & Rovine, M. (1985). Repeated-Measures Analysis of Variance in Developmental Research: Selected Issues. *Child Development, 56*, 787-809.

Henderson, A. M. E., & Graham, S. A. (2005). Two year-olds' appreciation of the shared nature of novel object labels. *Journal of Cognition and Development, 6*, 381–402.

Lieven, E. (1994). The child's language environment: Dyadic and polyadic. In Galloway, C., & Richards, B. J. (1994). *Input and Interaction in Language Acquisition,* 56-74. Cambridge University Press.

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology*, *35*, 60-68.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*, 834-843.

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., & Zelazo, P. D. (2003). Sources of inflexibility in 6-year-olds' understanding of emotion in speech. *Child Development*, *74*, 1857-1868.

Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development, 74,* 221-237.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*, 309-315.

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.

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1017-1030.

Paulmann, S., Titone, D., Pell, M. D. (2012). How emotional prosody guides your way: evidence from eye movements. *Speech Communication, 54*, 92-107.

Quam, C., & Swingley, D. (2011). Development in children's interpretation of pitch cues to emotion. *Child Development, 83,* 236-250.

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

Tomasello, M., & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology*, *30*, 639-650.

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, *23*, 157-176.

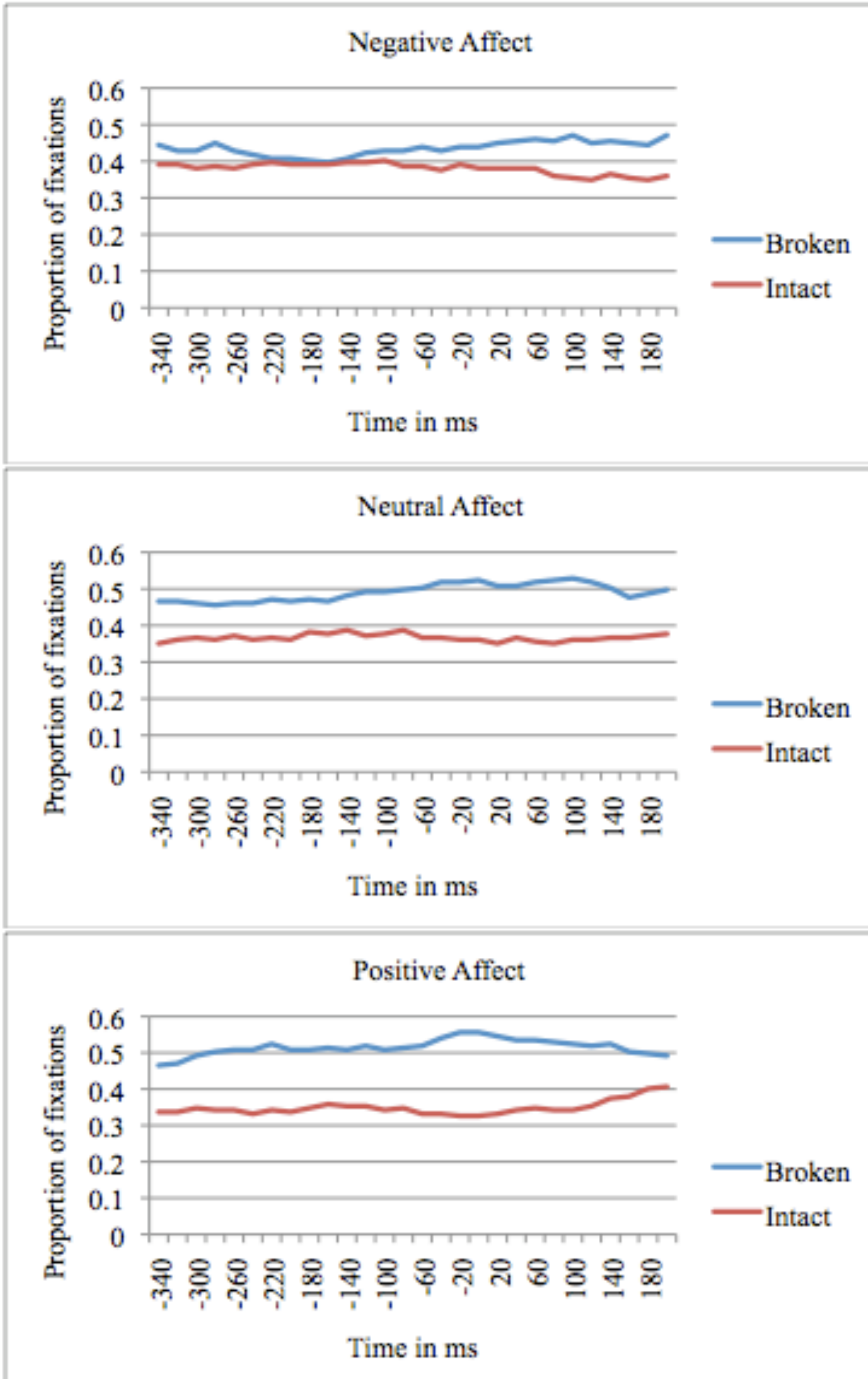
Waxer, M., & Morton, J. B. (2011). Children's judgments of emotion from conflicting cues in speech: Why 6-year-olds are so inflexible. *Child Development, 82,* 1648-1660.

Appendix A: Filler Trials

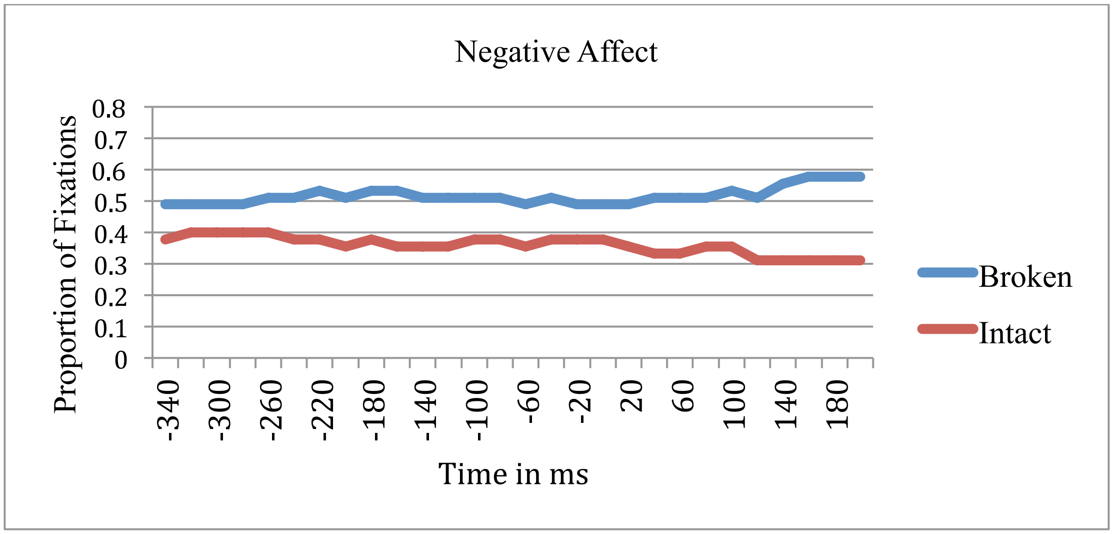
	Objects	*Change in Phase 2*	*Instruction*
1.	candle (f)	decorated	"Look at the candle"
	beach ball (f)	deflated	
2.	apple (f)	decorated	"Look at the apple"
	plate(f)	broken	
3.	doll (f)	broken	"Look at the dolly"
	hammer (f)	decorated	
4.	chicken toy (f)	broken	"Look at the chicken"
	scissors (f)	decorated	
5.	toy car (f)	broken	"Look at the muple"
	honey dipper (u)	broken	
6.	frog (f)	decorated	" Look at the sumal"
	book divider (u)	decorated	
7.	egg (f)	broken	"Look at the dillip"
	beverage cooler (u)	broken	
8.	fish (f)	decorated	"Look at the fishy"
	toe separator (u)	decorated	
9.	pencil (f)	broken	"Look at the pencil"
	plastic holder (u)	broken	
10.	monkey (f)	decorated	"Look at the monkey"
	juice top (u)	decorated	
11.	flower (f)	broken (no petals)	"Look at the flower"
	strainer (u)	broken	
12.	turtle (f)	decorated	"Look at the turtle"
	picture frame stand (u)	decorated	

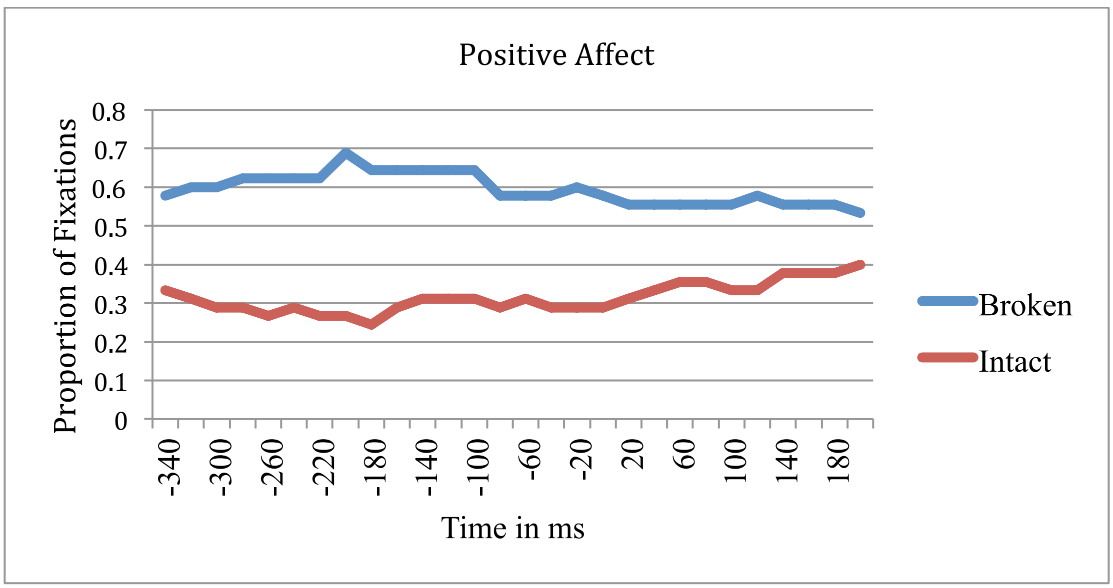
Note: (f) = familiar object, (u) = unfamiliar object

Appendix B. The probability of fixating display objects across the pre-noun window for Experiment 1.



Appendix C. The probability of fixating display objects across the pre-noun window for Experiment 2.





Chapter Four: Preschoolers' real-time integration of vocal and facial emotional information

Consider, for example, the utterance "It's snowing!". The meaning varies considerably depending whether the speaker is excited (perhaps a skier) or upset (perhaps a stranded commuter). During conversation, there are a number of sources of emotional information listeners can draw upon to disambiguate a speaker's communicative intent, including information conveyed via facial expressions or via vocal modulations. How children dynamically link these cues, however, is not well understood.

The building blocks for this integrative process are evident early in development. For example, 3-month-olds will use vocal affect to locate a matching emotional face, but only when the individual depicted is their mother (Kahana-Kalman & Walker-Andrews, 2001). During the preschool years, there is rapid development in children's sensitivity to various affective cues, particularly within the realm of vocal affect. For example, by 4 years of age, children can accurately label emotions from features of vocal affect alone (i.e., when listening to a foreign language or low pass filtered speech; e.g., Morton & Trehub, 2001). There are, however, observed differences in preschoolers' ability to recognize vocal affect across emotional categories. That is, preschoolers are more successful at recognizing sad vocal affect compared to happiness or fear (Nelson & Russell, 2011). Interestingly, differences of this type are also observed with adults, but are often manifested as differences in timing rather than success. For example, Pell and Kotz (2011) found that on average it took adults approximately 400 ms longer to recognize positive versus negative vocal affect.

The issue of timing in the recognition of vocal affect is highly relevant for understanding children's processing of emotion in speech. Because the affective cues expressed in speech often reflect a transient disposition towards objects or individuals (rather than a genuine emotional state), and because these cues may be conveyed via very brief utterances, the question of how quickly different emotions are detected over the course of child development has broad implications for understanding children's sensitivity to vocal affect in communication.

The goal of the current study is to examine the time course of children's ability to link auditory and facial expressions of affect across different types of emotions. Because children's latent sensitivity to vocal affect has been found to be more evident in measures of implicit processing (Berman, Chambers, & Graham, 2010; Berman, Graham, Callaway, & Chambers, 2013), we use eye gaze behavior to explore potential time course effects. Children's conscious selection of the emotional face that corresponds to speech (as measured pointing to a particular face) is also measured to provide an index of children's general success at matching visual and auditory cues.

Drawing upon a procedure developed by Paulmann, Titone and Pell (2012), we examined whether 3- and 5-year-olds could match sentences recorded with positive, negative, or neutral vocal affect to one of three photos (happy, sad, neutral) of the same individual. Critically, the content of the sentences was emotionally neutral, and each consisted of an instruction to look at an absent object (e.g., "Look at the doll/mido"). We included sentences with both novel and familiar nouns to control for potentially inherent valence biases (i.e., the noun ‘doll’ might plausibly hold more positive associations than ‘plate’, for example). We reasoned that differences in the basic ability to match vocal affect to visual information across emotions should be evident in both the eye gaze record and in children's points to the selected face. However, if the relevant differences are primarily related to the time course of processing, the results should be most evident in eye gaze patterns, and would be reduced or absent in the pointing data.

Method

Participants

The final sample included 32 English-speaking preschoolers: 16 3-year-olds (9 males, $M = 3.52$ years; $SD = .26$) and 16 5-year-olds (8 males, $M = 5.64$ years; $SD = .29$). Data from an additional five 5-year-olds and twelve 3-year-olds were excluded from analyses due to insufficient data collected from the eye tracker.

Stimuli

On each trial, the same three photographs of the face of a young woman were presented on a large display screen. The three photographs varied in terms of the emotional expression portrayed (happy, sad, or neutral) and were drawn from the Karolinska Directed Emotional Faces database (See Figure 1; Lundqvist, Flykt, & Ohman, 1998). The relative position of each emotional face in the display was systematically varied across trials.

The auditory stimuli consisted of 24 recorded sentences of the format, “Look at the X.” wherein X was either a familiar or unfamiliar noun (e.g., doll vs. mido). Each sentence was recorded in positive, neutral and negative vocal affect. The auditory stimuli were drawn from earlier studies in which the relevant affect differences had been statistically validated (Berman et al., 2010; Berman et al., 2013). For a given participant, each sentence occurred in only one of the three vocal affect types, and the pairing of vocal affect to sentences was counterbalanced across participants.

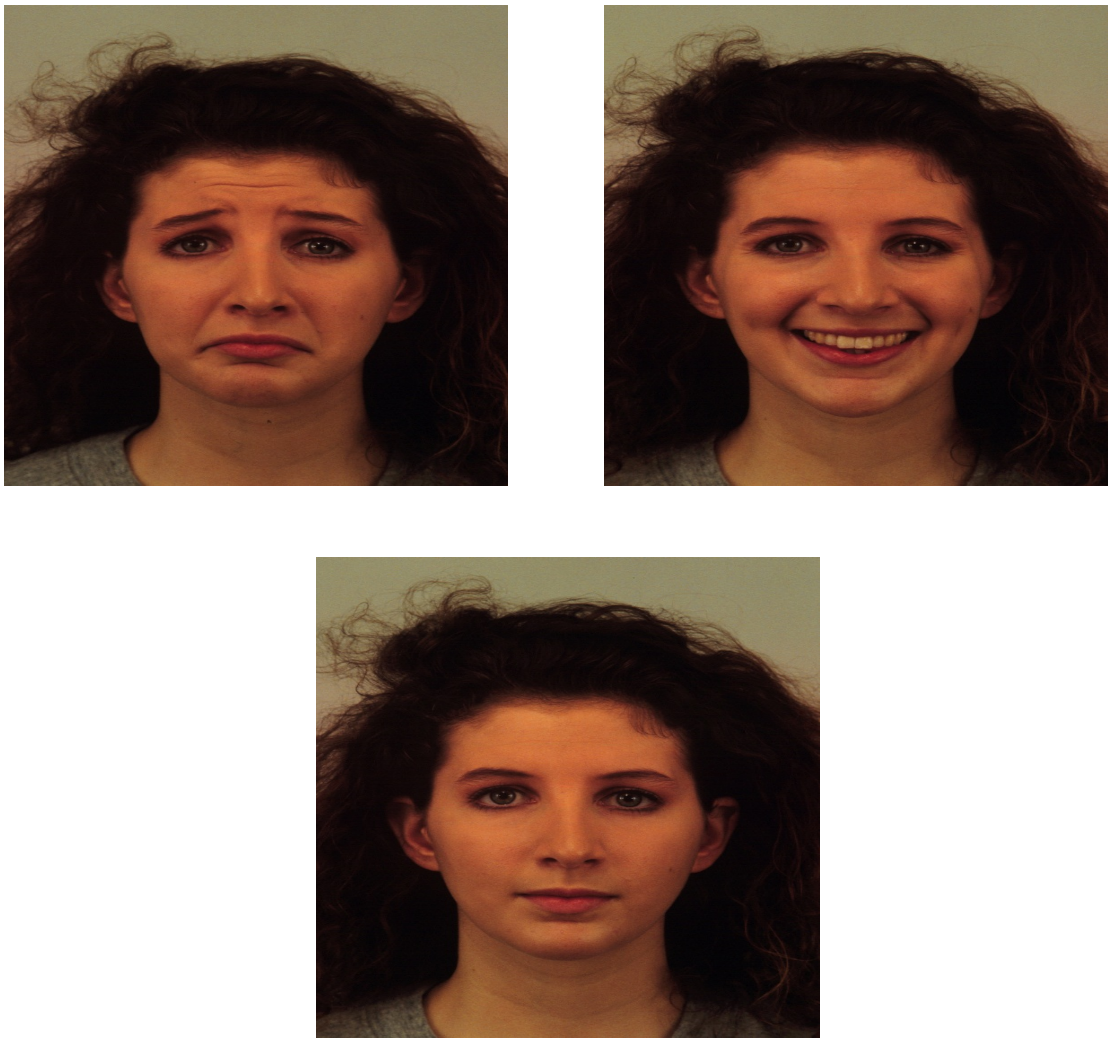


Figure 1. Sample array

Procedure

Each child was seated on a small chair facing a 46-inch computer monitor. The auditory stimuli were presented via speakers located behind the monitor. Eye movements were recorded using a Tobii x50 eye-tracker mounted below the display screen. A high-definition video camera placed behind the child recorded pointing.

After a brief calibration routine, each child was familiarized with the photographs of emotional faces (i.e., positive, neutral, negative). Each face was presented separately, accompanied by an utterance identifying the relevant emotion (e.g., "This is what I look like when I feel happy"). Following this familiarization, children were presented with 24 trials (8 for each vocal affect type) and were told to "listen to what [the speaker] says and then to point to what her face looked like when she said it". On each trial, the array of three faces was presented on the screen simultaneously with a recorded sentence. The array remained onscreen until the child pointed at one of the faces.

Results

Pointing Behaviors

Children's pointing behavior was coded as a measure of their explicit judgment about the emotional face that corresponded to the accompanying speech. Points were coded from videotape by assistants who were unaware of trial/condition information. Inter-rater reliability for 20% of the data ($n = 6$) was perfect (Cohen's Kappa = 1.00; $p < .001$). We calculated the percentage of points to the correct face matching the vocal affect in the accompanying speech (See Figure 2).

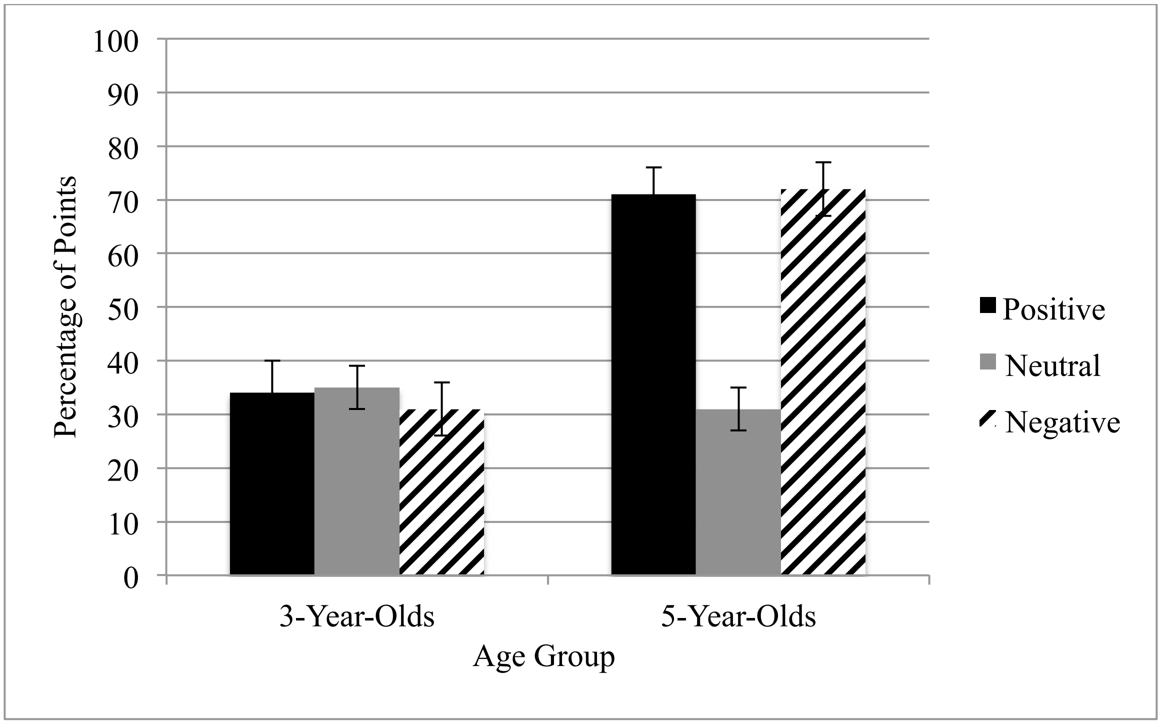


Figure 2. Percentage of points to corresponding face as a function of vocal affect type.

Data were submitted to an analysis of variance (ANOVA) with affect type (positive, neutral, and negative), and noun type (familiar vs. novel) as within-participant factors and age (3 vs. 5-year-olds) and gender (male vs. female) as between- factors. The analysis revealed a significant Age x Affect Type interaction (F (2,56) = 22.03, η_p^2 = .44, p < .01), indicating that the two age groups differed in their use of vocal affect information. Planned comparisons indicated 5-year-olds were more likely to point to the correctly corresponding face in the positive and negative affect conditions relative to the neutral condition (t(15) = 7.37, p < .001; t(15) = 7.05, p < .001, respectively), and there was no difference in correct selections across positive and negative affect conditions t(15) =.14, p = .89). Note that 5-year-olds' selection of the neutral face in the neutral condition corresponded to chance-level behavior (M = 31%, SE = .043, where chance is 33%) but their correct selections of the positive and negative faces were significantly greater than chance (ps < .001). Unlike the 5-year-olds, 3-year-olds did not match vocal affect to the correct face across conditions ps >.40). Neither gender nor noun type entered into any significant effects.

Eye Fixation Patterns

Figure 3 represents the probability of fixating the different faces over a 1600 ms interval (average utterance duration was 1586 ms). Because fixation patterns did not differ according to age group, gender, or whether the instructions contained novel or familiar labels (see below), the data are shown collapsed across these factors. The results for the positive vocal affect condition (top panel) show a preference to consider the happy face emerging around 600 ms after the onset of the utterance. This preference was sustained throughout the remainder of the analysis window. In the neutral affect condition (middle panel), the fixation profile suggests more or less equal consideration of all three faces. Finally, when the utterances were spoken with negative vocal affect (bottom panel), children displayed an increasing preference to fixate the sad face approximately 100 ms after the onset of the utterance and this preference remains until the 800 ms mark.

To evaluate these patterns statistically, we first calculated a difference score by subtracting looks to the happy face from looks to the sad face at each time point. (The exclusion of the neutral face from this score is motivated by children's apparent reluctance to consider this face in their pointing behavior and because the resulting measure will provide a more sensitive index of potential valence differences in the

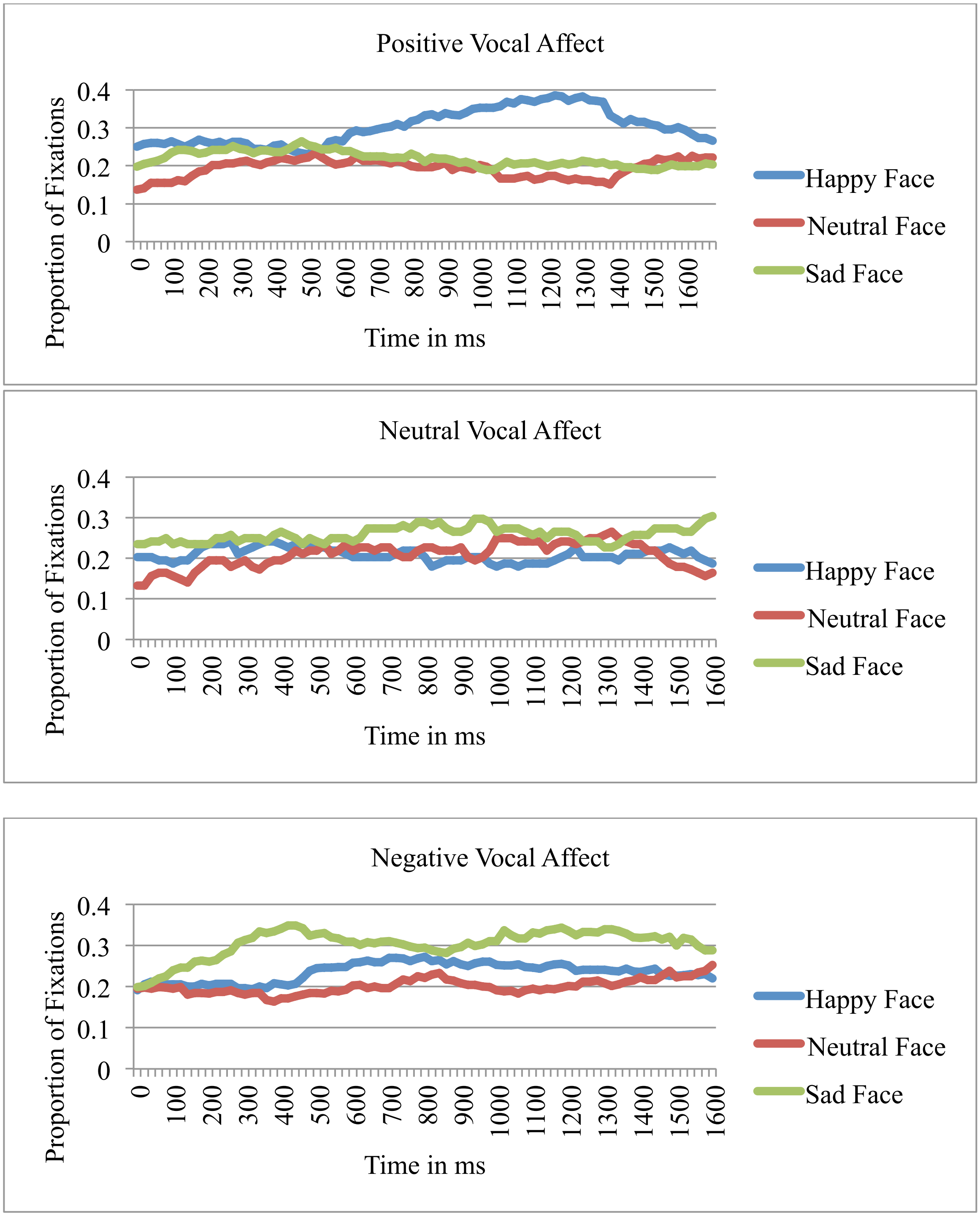


Figure 3. Proportion of fixations to faces over the course of the utterance, collapsed across age.

analyses that follow.) This measure provides a straightforward test of whether positive and negative vocal affect were reliably differentiated across the utterance. Timing effects were examined using four consecutive 400 ms intervals. We conducted separate ANOVAs for each interval, using the same model described above. For all 4 time intervals, there were main effects of vocal affect (all F's > 3.16, all p's < .05), suggesting consistent discrimination of positive and negative sounding speech across the utterance, and no significant effects of age, gender, or noun type.

To explore the apparent valence differences related to timing, we conducted follow-up comparisons against chance behavior (i.e., a difference score of 0) that showed that, in the negative affect condition, children looked more to the sad face than the happy face during the first two time intervals (ps < .01); whereas in the positive affect condition looks were not significantly different from chance in these time windows (*ps* > .18). In contrast, during the second half of the utterance (i.e., 800-1200 ms & 1200-1600 ms), children looked more to the happy face than the sad face in the positive affect condition during the final 800 ms interval (ps < .01). In contrast, children's looks in the negative affect condition did not differ significantly from chance (*ps* > .36). Finally, when the utterances were produced with neutral vocal affect there were no significant differences in looking behaviour throughout the entire utterance (*ps* > .19).

To summarize, the eye fixation data indicate that both the 3- and 5-year-olds use positive and negative vocal affect cues to locate the matching emotional face. However, children are able to use negative vocal affect significantly earlier in the utterance. The matching of positive affect information with a compatible face was not apparent until the final portion of the approximately 1600 ms utterance. Interestingly, neutral vocal affect led to consideration of all three faces for both age groups. Further, although 3-year-olds show sensitivity to vocal affect in their gaze behavior, their pointing behaviors suggest that the use of speaker affect is still at an emergent stage at this age: the overt selection of a referent did not reflect the accurate use of vocal affect information, consistent with earlier work in this area (Berman et., al, 2010). In contrast, the 5-year-olds successfully used positive and negative vocal affect to explicitly identify the corresponding face .

Discussion

The results of this experiment offer new insights into both time-course as well as into valence differences in preschoolers' use of vocal affect to gauge a speaker's emotional state. First, we found that children as young as three can match a speaker's vocal affect to her corresponding facial affect, as evidenced through their gaze behaviors. As the utterance unfolded, preschoolers used the speaker's vocal affect to locate the happy face in the positive affect trials and the sad face in the negative affect trials. This is an impressive outcome given 3-year-olds' documented difficulties in using vocal affect in previous studies (e.g., Berman et al., 2010). The simpler experimental scenario used in the current study is likely one reason for this difference, For example, in Berman et al. (2010), the task required 3-year-olds to link a speaker's vocal affect with her intention to refer to objects that were either intact or broken. The current results therefore suggest that 3-year-olds' difficulties are isolated to the process of linking vocal affect to referential intent, and do not simply reflect an inability to correctly recognize and classify a speaker's vocal affect.

Second, our results add to the growing literature that highlights differences in the sensitivity to negative versus positive affective information. Although this difference was not apparent in children's overt decisions about the match between an emotional face and emotional paralanguage, it was reflected in implicit processing, as reflected in eye movements. Specifically, although both 3- and 5-year-olds began to use negative vocal affect in the earliest moments to identify a sad face, it was not until approximately 800 ms into the utterance that they used positive affect to guide looking behavior. This timing effect in the integration of voice and faces is consistent with recent research suggesting that in the earliest moments of processing children and adults may be more attuned to negative vocal affect (e.g., Berman, Chambers & Graham, in press; Paulmann et al., 2012; Pell & Kotz, 2011).

Finally, the current results are broadly consistent with the well-documented finding of a negativity bias in children’s social-cognitive processing. In particular, Vaish Grossman, and Woodward (2008) propose that children differentially attend to negative versus positive information when making sense of the world around them. Further, negative stimuli have been hypothesized to carry a greater amount of information, which may require a greater amount of attention and processing (Peeters & Czapinski, 1990). Our results add to this literature and suggest that the negativity bias could be related to children’s faster processing of negatively-valenced vocal information.

References

Berman, J. M. J., Chambers, C., & Graham, S. A. (*2013*). Contextual influences on children's use of vocal affect cues. *Quarterly Journal of Experimental Psychology, iFirst, 1-22.*

Berman, J. M. J., Chambers, C. & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology, 107*, 87-99. doi:10.1016/j.jecp. 2010.04.012

Berman, J. M. J., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to learn new words. *Child Development. doi: 10.1111/cdev.12074*

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development*, *72*, 352-369. doi: 10.1111/1467-8624.00283

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*, 834–843. doi:10.1111/1467-8624.00318

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. doi:10.1016/j.jecp. 2011.03.014

Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: Evidence from eye movements. *Speech Communication*, *54*, 92–107. doi:10.1016/j.specom.2011.07.004

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (pp. 33-60). Chichester, England: Wiley.

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE 6(11)*: e27256. doi:10.1371/journal.pone.0027256

Vaish A., Grossmann T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*, 383–403. doi: 10.1037/0033-2909.134.3.383

Chapter 5: Conclusion

The chapters of this dissertation examined preschoolers' use of vocal affect in differing communicative situations with a specific focus on: (1) the contextual circumstances that promote preschoolers' use of vocal affect; (2) the developmental trajectory of this skill; and (3) how different vocal affect categories vary in terms of their effects on interpretation and the time course of processing.

In Chapter 2, I investigated how different visual and linguistic contexts influenced 5-year-olds' use of vocal affect to help interpret referring expressions. Across three experiments, 5-year-olds were presented with referential situations where the ambiguity of the referential expression either was or was not resolved by the end of the utterance. In Experiment 1, children were presented with two potential referents (e.g., two dolls) and one distracter object (e.g. a phone) and were asked to find the object requested by an ambiguous utterance (e.g., "Look at the doll…Point to the doll"). In Experiment 2, children were presented with contexts that contained only one potential referent (e.g., one broken doll) and two intact distracter objects (e.g., a cup and a plane) and were asked to find the object requested by an unambiguous utterance (e.g., "Look at the doll…Point to the doll"). Finally, Experiment 3 was analogous to Experiment 2, but used an embellished target (e.g., a fancy doll) and two broken distracter objects. Results from this set of experiments were twofold. First, in situations where the context entailed referential ambiguity (e.g., two possible targets), the findings suggested that 5-year-olds could explicitly use a speaker's vocal affect in a referential context. That is, children's gaze and pointing behaviour suggested an accurate use of both positive and negative (i.e., happy- or sad-sounding) vocal affect to resolve referential ambiguity and correctly interpret the speaker's request. Recall that in this type of task pointing is thought to reflect children's final referential decisions after all information has been integrated. Given previous research has indicated that 4-year-olds' sensitivity to vocal affect in a similar referential situation was evidenced only through their eye gaze and 3-year-olds exhibited no sensitivity in either their eye gaze or pointing behavior (Berman et al., 2010), the current findings suggest a more sophisticated use of vocal affect by age 5.

Second, results from Chapter 2 indicate that the referential context provided led to differences in timing across valence types. Here, when 5-year-olds were provided with unambiguous situations (e.g., one doll in the array and the referring expression requested children to "Look at the doll"), their eye movements suggested that in the earliest moments of the utterance (e.g., "Look at the…") negative but not positive vocal affect led children to anticipate reference to the broken object. In contrast, 5-year-olds' eye gaze in the positive vocal affect condition, when provided arrays with one embellished object and two broken distracter objects, reflected a relatively late use of vocal affect as evidenced by their gaze to the "odd-man-out" embellished distracter during the noun region (e.g., "doll"). Further, when 5-year-olds were provided with indeterminate ambiguity (e.g., multiple dolls when the referring expression requested children to "Look at the doll"), their eye movements suggested that they waited until the onset of the ambiguous noun (i.e., doll) to begin to make use of positive and negative vocal affect. These results point to a more efficient use of negative vocal affect versus positive vocal affect in unambiguous compared to ambiguous situations.

In Chapter 3, I explored 4- and 5-year-olds' use of a speaker's vocal affect to learn new words. In this task, children were presented with two novel objects that were shown to undergo a state change relative to their original depiction (e.g., one was subsequently shown as broken while the other was decorated by adding spots or stripes). Children's task was to map a novel label to one of the novel objects, drawing upon the speaker's vocal affect across three different affect conditions (positive, neutral or negative). Eye gaze measures indicated that both 4- and 5-year-olds used vocal affect to help map novel words to novel objects. However, in terms of their final behavioural decisions, only 5-year-olds displayed a more sophisticated use of vocal affect. Specifically, as the utterances moved from happy-sounding to more sad sounding, 5-year-olds, but not 4-year-olds were more likely to point to the broken object. When the strength of 5-year-olds' initial mapping was further tested through an extension task (i.e., when children were once again shown the original objects in their unaltered state) children were more likely to point to the target object if they had mapped the label using negative rather than positive vocal affect. Similarly, when 5-year-olds were asked to generalize their mappings to new category exemplars (i.e., a new exemplar whose colour was different than the original), children were more likely to point to the target object if they had mapped the label using negative rather than positive vocal affect. Taken together, results suggest that 4-year-olds are still at a transitional stage in their use of vocal affect to resolve ambiguity, whereas, 5-year-olds display a much more sophisticated use of this cue.

Finally, in Chapter 4, I examined whether 3- and 5-year-olds could match a speaker's vocal affect to her associated facial affect. This task required children to match the same utterances used in the experiments in Chapters 2 and 3 (e.g., "Look at the doll" or "Look at the blicket") to one of three emotional faces (i.e., happy, neutral, or sad).

This chapter was included to test whether a less complex task that bypasses the need to resolve referential mappings might reveal an earlier appreciation for vocal affect cues. Results from this chapter suggested that both 3- and 5-year-olds could use a speaker's vocal affect to recognize her associated facial affect. However, there were notable differences across affect types. More specifically, both 3- and 5-year-olds' eye gaze reflected a more rapid influence of negative compared to positive vocal affect. That is, both 3- and 5-year-olds looked at the sad face when the utterance was spoken with negative vocal affect within the first 400 ms of the utterance, whereas, it was not until approximately 800 ms into the utterance when children used positive vocal affect to locate the happy face. Across both age groups, children's pointing and eye movements suggested that children were not able to match a speaker's neutral vocal affect to an associated unemotional face. Furthermore, there were differences between 3- and 5-years-olds' explicit responses. Specifically, 5-year-olds, but not 3-year-olds, successfully pointed to the matching emotional face at above-chance levels in both the positive and negative vocal affect conditions. Together, these results suggest an emerging appreciation for vocal affect in 3-year-olds and once again point to a more sophisticated use of this cue by 5 years of age.

In this final chapter, I draw on these findings to refine our understanding of preschoolers' increasing ability to interpret and use a speaker's vocal affect in communicative contexts. I also discuss processing differences related to different emotional valence (i.e., positive and negative) across the various tasks used throughout this dissertation. I then discuss broader social implications of this work, with a focus on children's communicative competence. Finally, I discuss the limitations of the current studies and areas for future research.

The Development of an Appreciation for Vocal Affect

The results of this dissertation help reveal the specific ways in which children's use of emotional information (and particularly vocal affect) progresses with age. Previous research regarding children's coordination of vocal affect with more specific speech information when these two sources of information conflict point to a U-shaped developmental trend where infants use affective information to guide their behaviour, preschoolers appear to disregard this cue in favour of the content of the sentence, and school-aged children once again use a speaker's vocal affect to understand a speaker's underlying emotional state (Morton & Trehub, 2001; Friend, 2000). However, the results of this dissertation suggest that the previously described U-shaped developmental trend may have underestimated children's sensitivity to vocal affect during the preschool years. Specifically, the results indicate that children's use of vocal affect to recognize a speaker's associated emotional face emerges around the age of three and the ability to use a speaker's vocal affect in more referential situations emerges as children reach age 4. However, this trajectory is dependent on differences in context, task requirements, and measurement. Each of these areas will be described in more detail below.

Context. Throughout this dissertation, a number of different contextual situations were used to assess preschoolers' use of vocal affect. First, Chapter 2 provided evidence that 5-year-olds were able to explicitly use a speaker's vocal affect when the context involved referential ambiguity. In contrast, when the situation did not involve referential ambiguity (i.e., when there was only one possible referent for the description in the array), 5-year-olds' eye gaze suggested that during the earliest moments of an utterance (i.e., "Look at the…" when children heard "Look at the doll") they used negative, but not positive vocal affect to make real-time predictions regarding the eventually named referent. These results suggest that when 5-year-olds are presented with multiple referents, but ambiguous language, their eye gaze reflects a relatively late use of vocal affect (i.e., it was not until the noun "doll" where children's gaze reflected an apperception for vocal affect). However, when provided with an unambiguous context (i.e., a single referent), children can make immediate use of negative vocal affect cues.

Why might ambiguity delay the use of vocal affect? First, it is possible that when presented with multiple referents children expect their conversational partners to use disambiguating language. For example, perhaps children were waiting to hear "Look at the *broken* doll" and only when they encountered the word "doll" without the disambiguating adjective was the use of vocal affect triggered to resolve the ambiguity. A second possibility is that children needed additional time to inspect the differences between the broken and intact objects. From this perspective, children's delay was a result of comparing and contrasting the two same category referents before attempting to decide which referent the speaker may be requesting. A final, yet related possibility is that the ambiguity encountered required children to engage in additional processing and thus slowed their integration of vocal affect. Regardless of which explanation holds true, these results point to a delayed use of vocal affect when children are presented with ambiguous contexts.

Second, Chapter 3 tested children's ability to use vocal affect to learn new words. Here, the word learning context presents children with a different type of referential task where children do not bring their representations of familiar words or objects. That is, in this task, children were presented with novel rather than known words and objects. This difference might allow children to display a more robust appreciation for vocal affect as a cue to referential intent as they do not have any competing or conflicting linguistic information. Alternatively, the additional computational steps needed to learn new words may tax children's cognitive abilities and lead to a failure to use a speaker's vocal affect or a more modest influence relative to cases involving known words. This situation therefore provides an important complementary test for examining children's use of vocal affect. The results from the experiments in this chapter provided evidence that both 4- and 5-year-olds can use vocal affect to help learn new words. This is an impressive feat given the high demands this task likely placed on their developing cognitive capacities. Recall that children needed to hold representations of the original objects in their working memory and then compare these representations to the state changed objects in the mapping phase and then recognize and use the speaker's vocal affect to come to a referential decision. Across both age groups, children's eye gaze during the novel noun suggested that children used a speaker's positive and negative vocal affect to begin to make referential decisions regarding her intentions. However, when children were required to take the next logical step in word learning (i.e., extending and generalizing the newly mapped label), 5-year-olds' referent selection reflected the use of negative, but not positive vocal affect. More specifically, when the label was mapped using negative vocal affect children could find the correct referent in both the extension and generalization trials; however, when the referent was mapped using positive vocal affect children's referent selections were at chance levels.

Taken together, these results suggest that ambiguity across different contexts (i.e., familiar referent selection or novel word learning) may delay the use of a speaker's vocal affect to make referential decisions. Specifically, when children were faced with referentially ambiguous contexts, the speaker's vocal affect did not help narrow their referential options until the ambiguity was encountered in the speech stream. In contrast, when provided with unambiguous situations, 5-year-olds could use a speaker's negative, but not positive vocal affect in the earliest moments of processing to help guide their referential decisions.

Task requirements. Across this dissertation a number of different tasks were used to assess children's use of vocal affect. For example, when 3-year-olds were required to use a speaker's vocal affect to locate a photo of a face showing the relevant facial affect in Chapter 4, their eye gaze reflected sensitivity to both positive and negative vocal affect. These results suggest that using faces rather than objects allowed younger children to display their emerging sensitivity to vocal affect. One possibility for this finding is that faces are a more socially relevant cue for children than the known objects used in Chapter 2. That is, children may have been more likely to associate positive and negative vocal cues with a face rather than with objects that are broken or embellished. A second explanation for 3-year-olds' success is that using faces as stimuli reduced the number of inferential steps required for success. For example, it is likely easier for children to link a speaker's negative vocal affect to her associated sad face (as we know that even young infants can make this connection), compared to the process involved in associating her negative vocal affect with a broken object. This latter process involves an additional step whereby children need to match the speaker's emotional disposition to an object whose properties (e.g., brokenness) make it plausible for a given emotional reaction (e.g., sadness). This explanation suggests that the 3-year-olds in Berman and colleagues (2010) task may, in fact, be able to infer the speaker's emotional state from her associated vocal affect, but failed to make the extra computational steps required to connect the vocal affect to a broken or intact object. Yet another explanation stems from the fact that the same three faces were used across all trials in the Chapter 4 experiments, whereas different known/unfamiliar objects were presented on each trial in the earlier studies.

Although the precise basis of this difference is unclear, the results of the experiments in Chapter 4 are broadly consistent with recent research with adults showing that both positive and negative vocal affect are rapidly and unconsciously used to match the vocal affect of a speaker to an associated emotional face. Paulmann, Titone and Pell (2012) demonstrated that adults are quicker to shift their gaze to a face depicting a particular emotion, across both positive and negative affect conditions, when the accompanying speech affect is congruent (e.g., "Click on the happy face" spoken with positive affect) than when it is incongruent. In addition, emotional information was less influential as the sentence progressed. That is, as disambiguating language was heard, the content of the sentence superseded the vocal affect of the speaker, suggesting that vocal affect might play a more important role in the earliest moments of adults' language processing (See Rigoulot & Pell, 2012 for a similar result). Recall, that both 3- and 5-year-olds' eye gaze suggested an immediate influence of negative but not positive vocal affect. These results suggest that negative vocal affect may play an important role in the earliest moments of 3- and 5-year olds language processing, however, there are significant differences between the Paulmann et al., task and the task used in Chapter 4 of this dissertation. Critically, Paulmann and colleagues provided disambiguating content (e.g., "Click on the happy face"), whereas children in Chapter 4 were not provided with similar information. Therefore, similarities across these two studies should be interpreted with caution. However, the location of the effects (i.e., the earliest moments of the utterance) for both children and adults' use of negative vocal affect were similar.

Measurement. Language is spoken at a remarkably fast pace – approximately 2.5 words per second. Research has confirmed that as utterances unfold, language is processed rapidly and incrementally. Listeners analyze language on a moment-to-moment basis, rather than waiting for the end of an utterance or even a word (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Although overt behavioral measures such as pointing or verbal responses can provide insight into the final outcome of language processing, they do not allow for an examination of the incremental processes that may influence an individual's final interpretation (e.g., Fernald, Zangl, Portillo, & Marchman, 2008; Trueswell, Sekerina, Hill, & Logrip, 1999). In addition, as motor responses, eye movements are less metabolically demanding and are significantly faster than movements such as pointing or nodding (McMurray & Aslin, 2004). With the above in mind, I consider the 4-year-olds in Chapter 3 and the 3-year-olds in Chapter 4 who demonstrated sensitivity to the speaker's vocal affect only in their eye gaze responses. Here, if we were to rely on children's final behavioural decisions, we would miss the developmental emergence of children's use of vocal affect. Specifically, both the 3-year-olds and 4-year-olds' pointing was at chance levels across these different tasks, suggesting that they were not influenced by the speaker's vocal affect. However, at the more implicit level their eye gaze reflected a sophisticated use of this cue. These results fit well with Garner and Perner's (2001) implicit knowledge hypothesis. Here, earlier visual attention is thought to reflect an unconscious process that is outside of the individual's awareness and therefore cannot be called upon during specific assessment or questioning. Thus, when children were required to point to a specific object or face they were unable to use this implicit knowledge to demonstrate their underlying understanding. That is, both the 3- and 4-year-olds implicit behaviour (i.e., eye gaze) reflected an understanding of the use of vocal affect, but they could not (at this time) explicitly use this cue.

Differences in Valence and Timing

The results of this dissertation also point to significant timing and valence differences across emotions when investigating preschoolers' use of vocal affect. Specifically, in Chapter 2, when 5-year-olds were presented with unambiguous contexts, they used negative affect early in the utterance to begin to make referential decisions regarding the intent of the speaker. Further, in Chapter 3, 5-year-olds extended and generalized the newly mapped labels when they were learned using negative vocal affect. Finally, in Chapter 4 both 3- and 5-year-olds used negative vocal affect more quickly than positive vocal affect to match the vocal affect of the speaker to an emotional face. Thus, these results point to a stronger and earlier use of negative vocal affect across contexts.

Why might preschoolers privilege negative over positive vocal affect? There are a number of possibilities for these findings. First, it is possible that children are simply more successful at recognizing negative than positive vocal affect (Nelson & Russell, 2011; See Pell, Monetta, Paulmann, & Kotz, 2009; Pell, Paulmann, Dara, Alasseri, & Kotz, 2009; Scherer, Banse, & Wallbott, 2001 for similar results with adults). The results from this dissertation could fit within this line of research, but the results from both Chapter 4 (using faces rather than specific objects) as well as in the second experiment in Chapter 2 (where referential ambiguity was removed) suggest there may also be a timing related explanation for preschoolers' behaviour.

This second and by no means mutually exclusive possibility is that preschoolers are significantly faster at recognizing negative vocal affect. This possibility stems from research suggesting that adults are quicker to identify sad compared to positive vocal affect (Pell & Kotz, 2011). More specifically, adults took approximately 400 ms longer to identify happiness compared to sadness. This timing fits well with the results of Chapter 4. Here, both 3- and 5-year-olds began to match negative vocal affect with the sad face approximately 400 ms after the onset of the utterance, whereas, it was not until approximately 800 ms after the onset of the utterance where children began to use positive vocal affect to locate the happy face. Further, the results of the second and third experiments in Chapter 2 suggest that during the prenoun region (i.e., "Look at the…") only negative vocal affect played a role early in the utterance, whereas, an effect of positive vocal affect did not exert an influence until much later in the sentence processing task. Finally, the finding that 5-year-olds could only extend and generalize a newly mapped label, when it was learned with negative vocal affect could also be explained by a more efficient use of negative vocal affect. That is, children's eye gaze profiles in the noun region of the mapping phase reflected an earlier use of negative vocal affect. More specifically, during the noun region children's eye gaze suggested that they settled on the broken target object approximately 100 ms into the utterance when the utterance was spoken with negative vocal affect, whereas, it was not until approximately 500 ms into the utterance where children began to settle on the embellished object when the sentence was spoken with positive vocal affect. One possibility for children's faster recognition and use of negative compared to positive vocal affect seen across the experiments in this dissertation is that perhaps the acoustic cues for negative vocal affect (i.e., slower speech rate, lower intensity, and lower overall F0) are richer and better recognized by children and therefore allow for a faster and more robust detection and use of this cue.

The stronger and faster influence of negative vocal affect observed across the experiments in this dissertation also fits well within the broad framework of a negativity bias in children's social emotional processing (Vaish, Grossmann, & Woodward, 2008). The infant studies discussed in Chapter 1, which suggested that children could use their parent's emotional signals during social referencing interactions to guide their behaviour, rarely mention differences between affect types (e.g., Mumme & Fernald, 2003). However, it has been argued that using positive emotional cues does not necessarily increase infants' exploration of novel toys rather the significant differences stem from infants' reliance on negative cues to decrease their exploration (Vaish et al., 2008). Furthermore, when social referencing studies include neutral conditions, the differences between negative and neutral are significant, but the observed differences between positive and neutral are not (Mumme, Fernald, and Herrera, 1996). These results suggest that, even at a very young age, children may be differentially attuned to negative vocal affect. Furthermore, Peeters and Czapinski (1990) have argued that there is a positive-negative asymmetry in the way humans process emotionally valenced information.

Specifically, these authors argue that due to the inherent vulnerability of humans as we evolved, we learned to place a greater emphasis on the negative consequences of our actions. Over time, this emphasis likely caused our cognitive systems to place a greater weight on incoming negative information. With the above results in mind, my proposal is that children's early recognition of negative vocal affect combined with the assertion that negative emotion carries more information allows children to use negative vocal affect earlier in referential communication tasks.

Social Implications

Children's use of vocal affect has important implications for their overall social development. The results of this dissertation suggest that the vocal affect of a speaker is an important cue that young children can use to understand a speaker's referential intentions. However, the majority of the work in this area has tested children's identification of facial rather than vocal affect.

During the preschool years children begin to engage in social interactions that will ultimately shape their growing personalities (Crick & Dodge, 1994). For example, research has suggested that children's ability to detect and understand another person's emotions is strongly related to cognitive development and social competence (Denham et al., 2003; Trentacosta & Izard, 2007). Furthermore, previous research has demonstrated an association between children's social skills and their performance on emotion tasks. More specifically, Boyatzis and Satyaprasad (1994) found that 4- to 5-year olds' recognition, but not production of facial affect, was correlated with peer popularity from the teacher's perspective. Additional research has discovered comparable relationships, in preschool aged children, between understanding of emotions and school readiness

(Webster-Stratton, & Redi, 2004) and positive peer relations (Hubbard & Coie, 1994). It has also been argued that more broad social difficulties in these early years are highly linked with adjustment issues later in life (Crick & Dodge, 1994).

There is a wealth of evidence suggesting impairments of emotion recognition in several non-typically developing groups of children. However, the bulk of this research has been conducted with children who have been abused or neglected. For example, Pollack, Cicchetti, Hornung and Reed (2000) found that a group of preschool aged children with a positive history for physical abuse and neglect were less likely than their same aged peers to match an emotional face to a corresponding emotional situation. Furthermore, more recent research has found significant differences in recognition of facial expressions between typically developing children and those on the autism spectrum (Wallace, Coleman, & Bailey, 2008).

Some researchers have begun to investigate the role of vocal affect and social competence. One such study suggested that young children who have difficulty identifying vocal emotions are likely to have higher levels of social avoidance and are more distressed in social situations (McClure & Nowicki, 2001). However, although this work suggests that there are significant consequences when children do not recognize a speaker's vocal affect, few studies have directly tested this hypothesis. As vocal affect is an important cue to understanding the emotions of others, and this understanding has significant consequences for children's emotional development, it will be important to continue to test both typically and non-typically developing children's use of vocal across different situations and contexts.

Limitations and Areas for Future Research

Although the results indicate that children can use a speaker's vocal affect across multiple contexts, the results of this dissertation leave a number of questions unanswered that should be addressed in future research. First, this dissertation investigated the use of positive, neutral, and negative vocal affect, which represent three markers on the continuum from happy- to sad-sounding speech. Although this is an important starting point, one could argue that there are a number of different emotions and emotional continua that children are faced with on a daily basis that are useful cues for social interactions (e.g., surprise, anger, fear, disgust). Further, according to Widen and Russell's (2003, 2008) model of emotional acquisition, children learn to identify happiness and sadness before moving to more specific emotions such as surprise, and disgust. This gradual refinement of emotional categories occurs across the preschool years. The authors' proposal is that children start with very broad emotional categories (i.e., happy or sad) and then narrow these categories as they encounter events that do not fit with the concept of either happiness or sadness. Further, when presenting emotional information through the voice, there are more consistent acoustic features for happiness and sadness compared to other emotions (i.e., disgust or surprise; Banse & Scherer, 1996). Therefore, given 3- and 4-year-olds' emerging use of positive and negative vocal affect seen across this dissertation, future work investigating more specific emotional categories should likely begin by testing older preschoolers (i.e., 5-year-olds).

Second, although the tasks used throughout this dissertation required children to either use vocal affect to resolve referential ambiguity or to match vocal affect to an associated emotional face, neither task involved a genuine social interaction. Therefore, there is a sense in which these tasks may have lacked ecological validity. For example, children were “interacting” with a recorded voice that was requesting them to look and point at specific objects rather than an actual conversational partner. Further, the visual environment employed in the experiments was arguably less complicated than the visual world preschoolers find themselves in on a daily basis. For example, in most naturalistic contexts, there are many competing cues for attention when children are engaging in communication. However, although the tasks used throughout this dissertation were not entirely naturalistic, the pattern of findings was extremely consistent across both contexts and tasks, suggesting that children’s use of vocal affect in these tasks likely approximates their real world behaviour. However, converging evidence from studies using more ecologically valid situations would help ensure the current results capture children's abilities in the real world.

In summary, the findings of this dissertation expand our understanding of how preschoolers use vocal affect in referential situations. Indeed, by 5 years of age, children can display a sophisticated use of vocal affect to understand a speaker’s referential intentions. In addition, the results of this dissertation suggest that even at 3 years of age children are beginning to show a nascent use of vocal affect to recognize a speaker’s emotional state. Further, these results speak to the important role negative vocal affect plays in children’s resolution of referential ambiguity. Finally, this research demonstrates the vocal affect of the speaker is a critical cue that children need to incorporate into their ever-growing catalogue of speaker-produced cues to language use.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language. 38***,** 419–439

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614-636.

Berman, J. M. J., Chambers, C. & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology, 107*, 87-99.

Bloom, P. (1997). Intentionality and word learning. *Trends in Cognitive Sciences, 1,* 9-12.

Boyatzis, C., & Satyaprasad, C. (1994). Children's facial and gestural decoding and encoding: Relations between skills and with popularity. *Journal of Nonverbal Behavior, 18,* 37–55.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74-101.

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence. *Child Development, 74,* 238–256.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye-movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen

(Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). Philadelphia, PA: John Benjamins.

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*, 238-252.

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.

Grossmann, T., Striano, T., & Friederici A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science, 9,* 309-315.

Grossmann T., Striano T., & Friederici, A. D. (2007) Developmental changes in infants' processing of happy and angry facial expressions: A neurobehavioral study. *Brain and Cognition*, *64*, 30-41.

Hubbard, J. A., & Coie, J. D. (1994). Emotional correlates of social competence in children's peer relationships. *Merrill-Palmer Quarterly, 40,* 1-20.

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development*, *72*, 352-369.

Klinnert, M. (1984). The regulation of infant behavior by maternal facial expression. *Infant Behavior and Development, 7,* 447-465.

McClure, E. B., & Nowicki, S. (2001). Associations between social anxiety and nonverbal processing skill in preadolescent boys and girls. *Journal of Nonverbal Behavior, 25*, 3–19.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6,* 203-229.

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.

Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development, 74,* 221-237.

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development, 67,* 3219–3237.

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.

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1017-1030.

Paulmann, S. & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: how much is enough? *Cognitive, Affective, and Behavioral Neuroscience*, *10*, 230-242.

Paulmann, S. & Pell, M.D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, *35*, 192-201.

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology, 1,* 33–60.

Pell, M. D. (2005a). Nonverbal emotion priming: evidence from the 'facial affect decision task'. *Journal of Nonverbal Behavior, 29,* 45-73.

Pell, M. D. (2005b). Prosody-face interactions in emotional processing as revealed by the facial affect decision task. *Journal of Nonverbal Behavior*, *29*, 193-215.

Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: disentangling the effects of prosody and semantic cues. *Cognition & Emotion*, *25,* 834-853.

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.

Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication, 50*, 519-530.

Pollack, S., Cicchetti, D., Hornung, K., & Reed, A. (2000). Recognizing emotions in faces: Developmental effects of child abuse and neglect. *Developmental Psychology, 36,* 679-688.

Quam, C., & Swingley, D. (2011). Development in children's interpretation of pitch cues to emotion. *Child Development, 83,* 236-250.

Rigoulot, S., & Pell, M. D. (2012). Seeing emotion with your ears: Emotional prosody implicitly guides visual attention to faces*.* *PLoS ONE, 7*, e30740.

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.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

Trentacosta, C.J., & Izard, C.E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion, 7,* 77–88.

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.

Vaish A., Grossmann T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*, 383–403.

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science, 7,* 261-269.

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology*, *22*, 373-377.

Wallace, S., Coleman, M., & Bailey, A. (2008). An investigation of basic facial expression recognition in autism spectrum disorders. *Cognition and Emotion, 22,* 1353–1380.

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.

Webster-Stratton, C., & Reid, M. J. (2004). Strengthening social and emotional competence in young children – The foundation for early school readiness and success. *Infants and Young Children, 17,* 96-113.

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology, 39,* 114–128.

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development, 23,* 291–312.

APPENDIX A: COPYRIGHT PERMISSION LETTER:

CHILD DEVELOPMENT

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

May 28, 2013

This is a License Agreement between Jared Berman ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3150360971663
License date	May 15, 2013
Licensed content publisher	John Wiley and Sons
Licensed content publication	Child Development
Licensed content title	Preschoolers Use Emotion in Speech to Learn New Words
Licensed copyright line	© 2013 The Authors. Child Development © 2013 Society for Research in Child Development, Inc.
Licensed content author	Jared M. J. Berman,Susan A. Graham,Dallas Callaway,Craig G. Chambers
Licensed content date	Feb 25, 2013
Start page	n/a
End page	n/a
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Total	0.00 USD
Terms and Conditions	