

When You're Happy and I Know It: Four-Year-Olds' Emotional Perspective Taking During Online Language Comprehension

Melanie Khu
University of Calgary

Craig Chambers
University of Toronto

Susan A. Graham
University of Calgary

Using a novel emotional perspective-taking task, this study investigated 4-year-olds' ($n = 97$) use of a speaker's emotional prosody to make inferences about the speaker's emotional state and, correspondingly, their communicative intent. Eye gaze measures indicated preschoolers used emotional perspective inferences to guide their real-time interpretation of ambiguous statements. However, these sensitivities were less apparent in overt responses, suggesting preschoolers' ability to integrate emotional perspective cues is at an emergent state. Perspective taking during online language processing was positively correlated with receptive vocabulary and an offline measure of emotional perspective taking, but not with cognitive perspective taking, conflict or delay inhibitory control, or working memory. Together, the results underscore how children's emerging communicative competence involves different kinds of perspective inferences with distinct cognitive underpinnings.

Consider a situation in which a speaker, sitting across a cluttered table from a listener, says, "Pass me that paper." A skilled listener can use a variety of cues to infer the speaker's referential intent: Is there a specific paper of interest to the speaker, or is there a particular paper that the listener, but not the speaker, can reach? This use of inferences about a conversational partner's mental state to guide communicative behavior is known as *communicative perspective taking*. Given that situations that require perspective taking are ubiquitous in day-to-day conversation, a key developmental achievement is learning to attend to and use perspective cues to influence behavior during communicative interactions. Here, we take a novel approach to this issue by investigating children's sensitivity to a partner's *emotional state* when interpreting spoken language.

Specifically, we examined 4-year-olds' use of a speaker's emotional tone of voice to make inferences about that speaker's communicative intent. To provide a stringent test of perspective taking, our task required children to consider the viewpoint of another individual whose emotional reaction to an event would conflict with their own.

To date, research on children's communicative perspective taking has primarily focused on contexts that require reasoning about others' *cognitive states*—that is, an individual's consideration of both what a communicative partner knows and whether this matches or mismatches one's own knowledge. For example, research has investigated young children's sensitivity to shared knowledge and information, or *common ground* (Clark, 1996), established through shared experience (e.g., Ganea & Saylor, 2007), as well as through common cultural background (e.g., Liebal, Carpenter, & Tomasello, 2013). These studies have demonstrated that older infants can use shared experiences with other individuals, such as previously searching for or sharing an object together, to identify the referent of an ambiguous request (e.g., Ganea & Saylor, 2007;

This research was supported by an operating grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) awarded to Susan A. Graham and Craig Chambers and funds from the Canada Foundation for Innovation, the Canada Research Chairs program, and the University of Calgary awarded to Susan A. Graham. Melanie Khu was supported by a graduate fellowship from SSHRC and a doctoral scholarship funded by the Killam Trusts. We thank Michelle Zepeda, Charlene Parker, Katherine Gibbard, Nicola Yee, Elisea De Somma, and Shruthi Sekar for their assistance with this research.

Correspondence concerning this article should be addressed to Susan A. Graham, Department of Psychology, University of Calgary, Calgary, Alberta, T2N 1N4, Canada. Electronic mail may be sent to susan.graham@ucalgary.ca.

© 2017 The Authors

Child Development © 2017 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2018/8906-0025

DOI: 10.1111/cdev.12855

Moll, Richter, Carpenter, & Tomasello, 2008). Similarly, 3- to 5-year-olds will assume that a speaker is requesting information about a novel, and not a culturally well-known, object when ambiguously asked, "What's that?" (Liebal et al., 2013). When producing referring expressions, children as young as 3 years of age also take common ground into account (e.g., Menig-Peterson, 1975; Matthews, Butcher, Lieven, & Tomasello, 2012).

Most relevant for the present work are studies that examine listeners' ability to consider others' perspectives during real-time language processing at a fine-grained temporal scale. In these studies, knowledge mismatches are typically created by varying what a communicative partner can or cannot see, with the contrast between mutually visible objects and objects that are hidden from one conversational participant serving as a proxy for knowledge that is shared (i.e., in common ground) or not shared during communicative interactions (e.g., Brown-Schmidt, 2009; Epley, Morewedge, & Keysar, 2004; Fan, Liberman, Keysar, & Kinzler, 2015; Keysar, Lin, & Barr, 2003; Wang, Ali, Frisson, & Apperly, 2015, among others). The question of interest is whether listeners' gaze patterns reflect a sensitivity to visual perspective differences in the early moments of language processing. Research using these paradigms has documented sensitivity to speaker perspective in children as young as 3–5 years (e.g., Nadig & Sedivy, 2002; Nilsen & Graham, 2009). Some studies, however, have found evidence for an egocentric bias—interference from a child's own point of view that limits the ability to fully consider a speaker's perspective—that decreases with age (e.g., Epley et al., 2004; Wang et al., 2015). The degree to which egocentricity is observed, and the point at which perspective integration occurs (i.e., early during language processing, as a sentence unfolds, vs. later, after a sentence is understood in its entirety), depends on several factors including task complexity and the degree of conflict between the listener and speaker's perspective (San Juan, Khu, & Graham, 2015).

To what extent do studies of common ground and visual perspective taking fully capture the kinds of sensitivities used in successful communicative interactions? This question is important because results from these types of tasks have been used to draw broader conclusions about the nature of children's communication and perspective-taking abilities (e.g., Fan et al., 2015; Nilsen & Fecica, 2011; Wang et al., 2015). However, dealing with discrepancies in knowledge and visual perspective involves only the cognitive aspects of

communicative perspective taking. As such, these paradigms do not address the *emotional* perspective of a communicative partner, which could conceivably be more commonly evaluated and also more consequential in children's everyday social interactions. To illustrate, consider the following situation: A child telephones a neighboring child to ask if she can come over and play, and the neighbor replies "okay." From the point of view of an experienced communicator, the child's understanding of this response should differ depending on whether it was spoken in an excited and cheerful manner versus an unenthusiastic and reluctant manner. At present, these emotionally shaded meanings have not been examined in empirical studies of online communicative perspective taking, despite evidence for differences in the development of cognitive and emotional perspective-taking abilities (e.g., Hoffmann, Singer, & Steinbeis, 2015). Given the possibility that children's success at communicative perspective taking might vary depending on the type of perspective inferences at play, a more comprehensive understanding of children's conversational abilities requires empirical investigation of their performance in emotionally relevant situations.

To engage in emotional perspective taking of this type, children need to be sensitive to the cues that signal a speaker's emotional state or emotional disposition toward objects and events. In recent years, there has been a growing body of research examining children's understanding of how this information is conveyed by emotional prosody (e.g., Berman, Chambers, & Graham, 2010; Quam & Swingley, 2012). Emotional prosody refers to non-linguistic modulations of the voice that are understood by listeners to convey emotionality (e.g., changes in pitch contours, pitch levels, and speech rate; see Banse & Scherer, 1996; Frick, 1985). Infants will use a speaker's emotional prosody, when paired with appropriate facial expressions, to guide their interpretation of an ambiguous referential statement (e.g., excitement, Akhtar, Carpenter, & Tomasello, 1996; Moll et al., 2008; dissatisfaction, Moll & Tomasello, 2006). By 4–5 years of age, children can use these prosodic cues without visual support to narrow the possible candidates for an unfolding referential description (Berman, Graham, & Chambers, 2013; Berman et al., 2010), to learn novel words (Berman, Graham, Callaway, & Chambers, 2013), to identify emotional faces (Berman, Chambers, & Graham, 2016), and to identify the emotional state of a speaker (Quam & Swingley, 2012). However, these studies do not provide

evidence of emotional perspective taking *per se*. This is because the use of emotional prosody cues does not require a consideration of the viewpoint of another individual whose emotional reactions might conflict with their own. For example, children's ability to use a speaker's sad tone of voice to help identify an intended referent (a broken doll) could rest on their own potential emotional reactions (e.g., I would be sad if my doll broke) or prior associations (e.g., my sister sounded sad when her doll broke), without engaging in perspective reasoning about the speaker in the current situation. Similarly, infants' use of a speaker's excited tone to identify an object the speaker and child had shared earlier (e.g., Moll et al., 2008) could arise because the child's own emotional reaction would likely be the same.

There is evidence from work outside the language processing and emotional prosody literature that suggests that preschoolers do, in fact, have a complex understanding of others' emotional states (e.g., Harris, de Rosnay, & Pons, 2016; Pons, Harris, & de Rosnay, 2004). By 3–4 years of age, children begin to understand the situations that elicit different emotions (e.g., Denham, 1986), and by 3–5 years of age, they start to appreciate the link between emotions and desires (i.e., that different desires lead to different emotional reactions; e.g., Harris, Johnson, Hutton, Andrews, & Cooke, 1989; Yuill, 1984). For example, when presented with vignette-style scenarios, preschoolers can correctly identify a puppet's emotional reaction to an emotion-evoking situation, even when this reaction differs from how the children themselves would react (e.g., the puppet enacts fear about attending the first day of school when, by parental report, the same situation would evoke excitement in the child; Denham, 1986; Denham & Couchoud, 1990). Thus, preschoolers appear to possess the basic competence required to engage in emotional perspective taking.

The above finding does not, however, reveal whether children can integrate inferences about another's conflicting emotional state rapidly enough to constrain their interpretation of spoken language. That is, in order to make use of another's emotional perspective to infer meaning and to react accordingly (as in the phone example described above), the integration of unfolding linguistic information with perspectival cues must occur quickly enough to influence children's comprehension "in the moment." Here, we pursued this question by presenting 4-year-olds with a task that required them to use a speaker's emotional prosody to infer the

speaker's emotional state and, correspondingly, the speaker's communicative intent. Our focus on 4-year-olds is motivated by the important developmental changes in communicative (e.g., Nilsen & Graham, 2009) and emotional perspective taking (e.g., Denham & Couchoud, 1990) that occur between 3 and 5 years of age, and the fact that considerable individual differences in cognitive abilities such as executive function (e.g., Carlson, 2005) are often reported for this age group.

The task was set up as a competitive game between the child and the speaker, in which a positive outcome for the child corresponded to a negative outcome for the speaker and vice versa. Here, we use the term "competitive" to indicate that on many trials, the child and the speaker desired different outcomes. Critically, this manipulation allowed us to create a direct contrast between the relevant emotional response for the child and that for the speaker. On each trial, children were presented with a visual display containing two doors: the child's door and the speaker's door. They were told that there would be a sticker behind only one of the doors and that the location of the sticker would determine which player would be awarded a sticker. Children were told that the speaker would receive advance knowledge of (but no control over) the location of the sticker on each trial. Based on the speaker's ambiguous sentence about the sticker's location (e.g., "Look! There it is."), spoken with either positive (happy) or negative (sad) emotional prosody, children had to identify the location of the sticker. If children are sensitive to the emotional state of the speaker, they should assume that the sticker would be behind the speaker's door when the speaker sounds happy and that the sticker would be behind their door when the speaker sounds sad. The opposite pattern of results would arise if children interpret the speaker's statements from a uniquely egocentric perspective.

As a further test of children's emotional perspective taking, a subset of trials involved scenarios where players could lose stickers. Children were told that if a large "X" appeared behind a player's door that player would lose a sticker. When the child lost a sticker, the speaker did not, and vice versa. Accordingly, on these trials, children should assume the "X" would be behind the speaker's door when the speaker sounded sad (i.e., because she would lose a sticker) and that the "X" would be behind their door when the speaker sounded happy (i.e., because the speaker would get to keep all of her stickers). Again, the opposite pattern of results would be expected if children were

interpreting the speaker's statements from a uniquely egocentric perspective. The inclusion of the potential loss trials ensured the design provided a strong test of children's perspective-taking abilities across scenarios.

Two measures were used. First, to examine online processing, we recorded children's real-time eye gaze behavior. Eye movement measures reflect moment-by-moment language processing at the millisecond level (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), often detecting effects and patterns that are not reflected in preschoolers' explicit behavioral responses, such as in the context of processing of emotional prosody (e.g., Berman et al., 2010; Berman, Graham, Callaway et al., 2013), sentence comprehension (e.g., Snedeker & Trueswell, 2004), and false-belief reasoning (e.g., Clements & Perner, 1994; Low, 2010). Specifically, we examined when, leading up to their explicit judgments about the sticker's location, children showed evidence of integrating the speaker's emotional perspective in their visual consideration of the two doors. Here, we were interested in whether children demonstrated a possible initial egocentric bias, followed by a recovery phase (as in Epley et al., 2004), or instead whether an appreciation of their partner's perspective would be apparent in the early moments of processing (as in Nadig & Sedivy, 2002). Our second measure involved children's controlled and conscious judgments, as reflected in their pointing to their own door or the speaker's door. If 4-year-olds readily engage in emotional perspective taking during online language comprehension, we reasoned that both measures should reflect consideration of the "correct" door from the speaker's view. Conversely, if children at this age are generally unsuccessful at this form of perspective taking (e.g., adopting instead an egocentric interpretation), both measures should reflect consideration of the incorrect door. A third possibility, reported in earlier studies of 4-year-olds (e.g., Berman et al., 2010), is that children's eye gaze patterns may reflect implicit sensitivities that do not fully translate into conscious behaviors (i.e., pointing). It is also important to note that children's success at emotional perspective taking may differ depending on emotional valence (happy vs. sad) and the reasoning context (e.g., when the situation involves a potential win vs. a potential loss).

We also collected measures of individual differences. Previous research has found correlations between individuals' visual perspective-taking ability during communication and their inhibitory control (e.g., Brown-Schmidt, 2009; Nilsen & Graham,

2009). There is also a well-established association between theory of mind (as measured by false belief tasks) and inhibitory control during the preschool years (e.g., Carlson & Moses, 2001; Devine & Hughes, 2014). A question relevant to the current study is whether different kinds of communicative perspective taking draw on different components of executive function. To this end, we measured children's working memory, delay inhibitory control, and conflict inhibitory control, and examined their correlations with performance on the communication task. If online communicative perspective taking within the emotional and cognitive domains draw on similar components of executive function, then children's performance on the real-time communicative perspective-taking task should be positively correlated with their conflict inhibitory control skills.

To assess the potentially important role of mentalizing ability (Nilsen & Fecica, 2011), we included an offline measure of *emotional* perspective taking (the affective knowledge task of Denham, 1986) and an offline measure of *cognitive* perspective taking (the visual perspective task introduced by Masangkay et al., 1974). If the types of mentalizing abilities that underlie communicative perspective taking are domain general, we would expect children's performance to be related to both of the offline measures of perspective taking. However, if this relation is more domain-specific, performance would plausibly be correlated only with the offline measure of emotional perspective taking. Finally, to control for baseline differences in children's basic language skills, children's receptive vocabulary was assessed.

Method

Participants

Participants were ninety-seven 4-year-olds (49 males; $M = 4.30$, $SD = 0.15$, range = 4.0–4.5 years), recruited as part of a larger study on communicative development. Data were collected between January 2014 and April 2015. A large sample was tested in order to ensure sufficient variance for the individual difference measures. All were native English speakers and had no reported developmental disorders. Five additional children were tested but excluded from the final sample due to difficulties following task instructions or uncooperativeness ($n = 2$) or insufficient eye gaze data ($n = 3$). The majority of families self identified as Caucasian (Caucasian = 86%, multiracial = 9%, Asian or other = 5%), and the majority of parents had completed

postsecondary education (high school = 13%, postsecondary = 73%, graduate studies = 14%). Children were tested across two hour-long sessions within a week of each other. Children completed an additional communicative perspective-taking task and their parents completed standardized questionnaires (those data will be reported elsewhere). Individual difference measures were collected following the communication task.

Communication Task

Stimuli

Visual stimuli consisted of 15 image sequences presented in a set pretrial, trial, posttrial order (see Figure 1). The pretrial image for each sequence contained either a single object (on critical trials) or an identical pair of objects (on filler trials) paired with a recorded statement introducing the object(s) (e.g., "The next sticker is a fish sticker" or, on potential loss trials, "Next is the bad X"). These statements were recorded using child-directed speech with neutral emotional prosody. The trial image that was presented next in each sequence consisted of duplicate closed doors, one red and one blue. The trial image was first presented for 1,000 ms in silence, after which the child heard a referentially ambiguous utterance ("Look, there it is" or "Oh, I see it."), spoken with either positive (happy-sounding) or negative (sad-sounding) emotional prosody. Although they were prerecorded, the utterances appeared to be spontaneously produced by the (confederate) adult player located in another room (see Procedure). The posttrial image of each sequence showed the two doors opened to reveal the object(s) behind the door(s) and was presented without sound. The location of the doors on the screen was counterbalanced, such that half the participants saw the red door on the left and the blue door on the right, and the other half saw the reverse. The side and color of child's door versus the adult player's door was also counterbalanced across participants, but for a given participant, the location of each player's door remained constant across all trials. On critical trials, the location of the revealed object was balanced across players (i.e., the object appeared behind each player's door on 50% of trials).

All auditory stimuli were recorded by a female native English speaker. To ensure that the utterances conveyed the intended emotional prosody, 10 adults rated the utterances on a scale that ranged from 1 (*negative sounding*) to 7 (*positive sounding*),

with 4 (*neutral*) at the midpoint. Ratings confirmed that the positive and negative utterance types were significantly different from one another (negative: $M = 2.05$, range = 1.00–3.00; positive: $M = 6.55$, range = 5.50–7.00; $p < .001$), as well as from the neutral introductory phrases ($M = 4.15$, range = 3.15–4.54; $ps < .001$).

Additional Materials

Stickers depicting the same objects used in the image sequences were used, along with one red and one blue sticker bag, and one red and one blue T-shirt.

Apparatus

The experiment was programmed and conducted using E-Prime software (Psychology Software Tools, Inc., Sharpsburg, PA) with Tobii extensions. Visual stimuli were shown on a large 46 in. monitor, and auditory stimuli were presented through speakers located behind the monitor. Participants' eye gaze was recorded using a Tobii x50 eye tracker (Tobii AB, Danderyd, Sweden). A five-point calibration procedure was conducted using Clearview software (Tobii AB, Danderyd, Sweden). The majority of children ($n = 94$) recorded five calibration points for each eye, with the remainder ($n = 3$) recording four points per eye. Participants' eye movements were identified in relation to specific areas of interest, corresponding to the location of display objects, defined prior to data collection. Eye movement data were collected during the test trial portion of each trial only (i.e., not during the pretrial or posttrial phases). Gaze position was logged every 20 ms and fixations were defined as looks to a given spatial location that lasted longer than 100 ms. The gaze data were aligned with speech landmarks in the prerecorded utterances using Eye-gaze Language Integration software (Berman, Khu, Graham, & Graham, 2013). An HD camera, positioned behind the participant, recorded pointing behavior.

Procedure

The task was introduced as a competitive game that would be played against an adult player (confederate). First, the child was shown "the control room" (i.e., a small room adjacent to the testing room) where the adult player was sitting at a computer screen. The child was told that she or he would be playing a game against the adult player and that the adult player would be speaking to

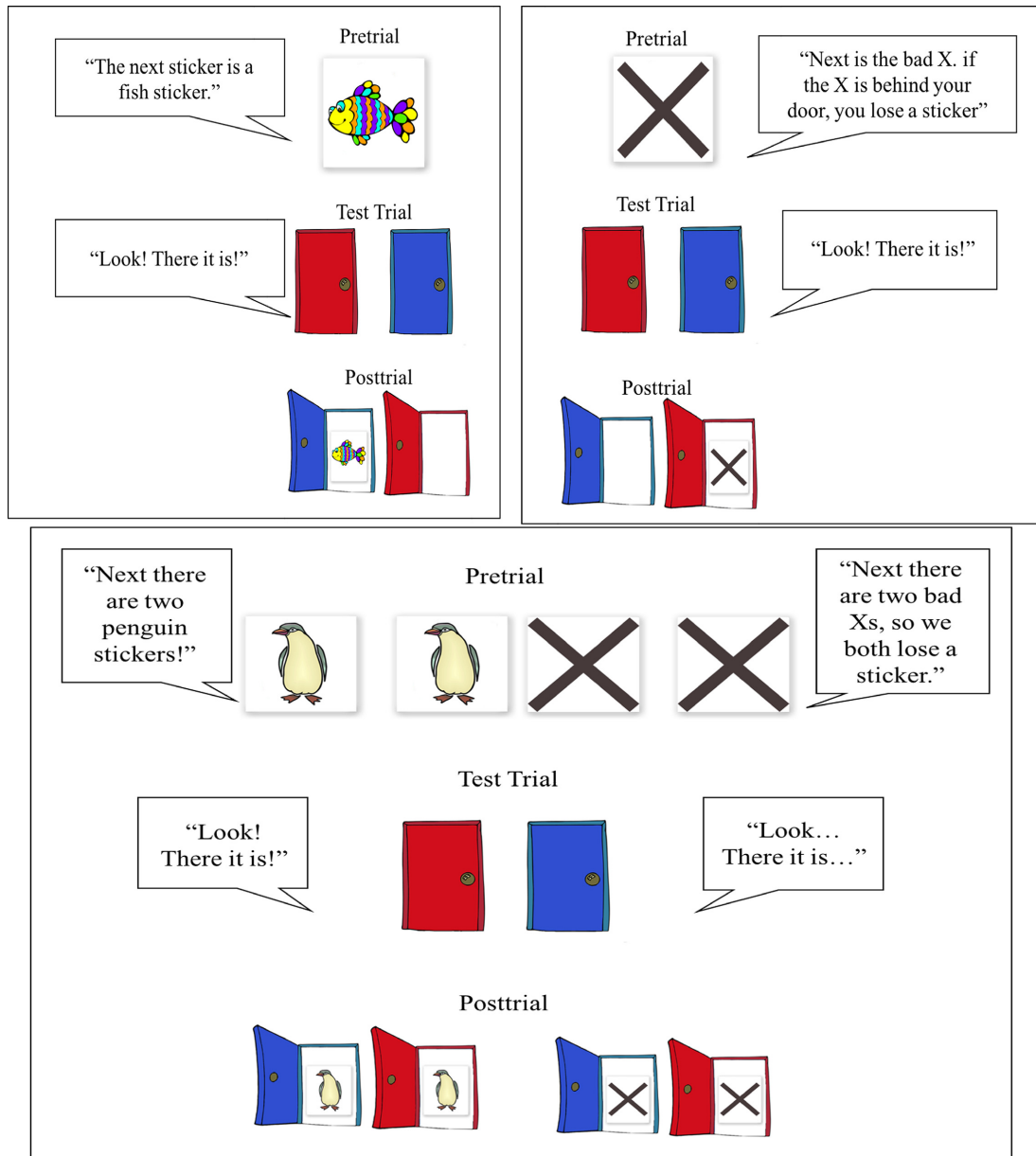


Figure 1. Sample procedure for potential win trials (top left panel) and potential loss trials (top right panel). Sample procedure for filler trials is presented in the bottom panel, with double win trials (spoken with positive emotional prosody) on the left and double loss trials (spoken with negative emotional prosody) on the right.

the child from the control room through a microphone. The purpose of this multiroom format was to allow for the use of prerecorded utterances while maintaining the impression that the child was interacting with a live speaker. The child and adult player were then each assigned one of two colors (e.g., child, blue; confederate, red) and given shirts and prize bags of that color. The color assigned to the child was counterbalanced across participants.

After meeting the adult player, the child was brought into the testing room and was seated in a chair approximately 4.5 feet away from a large monitor. Following a brief calibration procedure, images of a red door and blue door appeared side by side on screen. The experimenter explained the opportunity to win stickers depending on whether the sticker appeared behind the child's door (e.g., blue door) or the speaker's door (e.g., red door). The experimenter further explained that on each

trial the adult player (henceforth the “speaker”) would receive advance knowledge of (but no control over) the location of the sticker. Finally, the experimenter explained that the speaker would tell the child, through the microphone, which sticker would be coming next (e.g., fish, dog, “X,” etc.), and the child would have to decide which door the sticker would be behind. Control questions were used to verify that the child understood which door was his or her own.

The child then participated in four practice trials. The procedure for each trial was identical: During the pretrial phase, the child was informed about the type of upcoming object (i.e., a sticker or an “X”). During the trial, the child saw the two closed doors and heard an ambiguous utterance spoken in either a positive or negative prosody (e.g., “Look, there it is.”). The child was then instructed to point to the door he or she thought the object would be behind. The posttrial location of the sticker was revealed only after the child had pointed to a door. Based on the on-screen location of the object, a sticker was given to, or taken from, the child or speaker’s bag (e.g., if a sticker was behind the child’s door on screen, it was placed in the child’s bag). For the practice trials only, the experimenter also commented on how the child and speaker might be feeling (e.g., “It’s behind your door, so you get the sticker! You must be feeling pretty happy. [Speaker’s name] is sad because she didn’t get a sticker, but that’s okay, we’re going to play again.”). The four practice trials consisted of two potential win trials (one negative and one positive prosody), one potential loss trial (negative prosody) and one double win filler trial (positive prosody). Trial types are described in detail next (see Figure 1). After the practice trials, the child was presented with 16 critical trials intermixed with four filler trials and presented in one of four randomized orders.

Potential win trials. On these trials, players had the potential to win stickers. During the pretrial phase, the child was told by the prerecorded voice what the next sticker would be and saw a picture of the sticker. The child then saw two closed doors and heard the ambiguous prerecorded statements in either positive or negative emotional prosody. The posttrial location of the sticker varied by the prosody of the utterance: Negative utterances were followed by an image of the sticker behind the child’s door, and positive utterances were followed by an image of the sticker behind the speaker’s door.

If children interpreted the emotional prosody cues from an egocentric perspective, they would

assume that the sticker would be behind their door in the positive prosody condition (i.e., the child is happy because he or she will get a sticker) and behind the speaker’s door in the negative prosody condition (i.e., the child is sad because he or she will not get a sticker). By contrast, if children were taking the speaker’s perspective into account when interpreting the utterances, the reverse pattern would be expected. That is, children would assume that the sticker would be behind the speaker’s door in the positive prosody condition (i.e., the speaker is happy because she will get a sticker) and behind the speaker’s door in the negative prosody condition (i.e., the speaker is sad because she will not get a sticker). Each child was presented with eight potential win trials (four negative and four positive emotional prosody).

Potential loss trials. On these trials, the players had the potential to lose stickers. Here, the child was told by the prerecorded voice that the next trial would involve a “Bad X.” On every potential loss trial, the child was reminded that if the X were to be behind a player’s door, that player would lose a sticker. The child then saw two closed doors and heard the ambiguous prerecorded statements in either positive or negative emotional prosody. The posttrial location of the X varied depending on the prosody of the utterance: negative prosody utterances were followed by an image of the X behind the speaker’s door, and positive prosody utterances were followed by an image of the X behind the child’s door. As noted earlier, if children interpreted the emotional prosody cues from an egocentric perspective, in this scenario they would assume that the X would be behind the speaker’s door in the positive prosody condition (i.e., the child is happy because he or she will not lose sticker) and behind their own door in the negative prosody condition (i.e., the child is sad because he or she will lose a sticker). However, if children were taking the speaker’s perspective into account, the opposite pattern would be expected. Each child was presented with four potential loss trials (two negative and two positive emotional prosody).

Filler trials. Four filler trials were included to prevent children from establishing expectations about certain trial types and emotional prosody pairs. On the double win trials (two trials), the child saw two identical objects during the pretrial and was told that there would be two stickers, one for each player. The child saw the two doors and heard an ambiguous statement (always positive). During the posttrial reveal, both players had a

sticker behind their door, signifying that both players would be given a sticker. This type of filler trial was included to prevent children from adopting a strategy of discounting their door on positive prosody potential win trials. On the double loss filler trials (two trials), the child saw two Xs during the pretrial and was told that because of the two Xs, both players would lose a sticker. During the test phase, the child saw the two closed doors and heard an ambiguous statement (always negative). During the posttrial reveal, both players had an X behind their door, and so each player had a sticker removed from their prize bag. This type of filler trial was included to prevent children from discounting the speaker's door on negative prosody potential loss trials.

Individual Difference Measures

Executive Functions

Working memory. Children completed the picture memory subtest of the Wechsler Preschool and Primary Scale of Intelligence, 4th ed. (Wechsler, 2012). Raw scores were converted to standard scores using age-based Canadian norms.

Conflict inhibitory control. Children's performance was assessed using the Stroop-like "day-night" task (Gerstadt, Hong, & Diamond, 1994). The child was first asked what is in the sky at night (stars) and what is in the sky during the day (the sun). The child was then shown a card depicting a dark night sky with stars and a moon, and another card showing the sun in a light blue sky. Children were instructed to say "day" when presented with the moon and stars card, and "night" when presented with the sun card. Following three practice trials with corrective feedback, the child was administered 16 test trials in pseudorandom order. Accuracy (score up to 16) was recorded. Self-corrections were counted as incorrect (e.g., Gerstadt et al., 1994). A research assistant coded 20% of the data ($n = 20$). Interrater reliability was excellent (Cohen's $\kappa = .946$; $p < .001$).

Delay inhibitory control. Delay inhibitory control was assessed using a task based on Beck, Schaefer, Pang, and Carlson (2011). The child made a series of choices in which they could receive a smaller reward immediately or a larger reward later. Specific rewards were small erasers and colorful balls. Six test trials were administered in the following order: 1 versus 4 erasers, 1 versus 2 balls, 1 versus 6 erasers, 1 versus 4 balls, 1 versus 6 balls, and 1 versus 2 erasers. The number of trials on which the child chose the delayed reward (score up to 6) was recorded.

Mentalizing Ability

Emotional perspective taking. Vignettes of emotion-eliciting scenarios were presented in a puppet task (Denham, 1986). Puppets enacted the various situations, and children indicated how the protagonist puppet felt by affixing a felt face depicting one of four potential emotions (i.e., sadness, happiness, fear, or anger) to the puppet's face. The puppets enacted four vignettes that evoked unequivocal emotional responses (e.g., happiness at being given a cupcake), as well as six vignettes that involved more equivocal scenarios (e.g., going into the water at the swimming pool). In the latter situations, the puppet displayed an emotional response that was inconsistent with the child's typical response, as previously established by parental report (e.g., if the child is typically excited about going into the water, the puppet enacted fear). For each vignette, the child received a score of 2 if they correctly identified the displayed emotion, 1 if they identified an incorrect emotion of the correct valence (e.g., indicating angry instead of sad), and 0 if they identified an incorrect emotion of the incorrect valence (e.g., indicating happy for sad).

Cognitive perspective taking. The visual perspective "turtle task" (Flavell, Everett, Croft, & Flavell, 1981; Masangkay et al., 1974) was administered. The experimenter sat across the table from the child and placed a picture of a turtle between them. The picture was placed such that the turtle appeared right-side up to the child and upside down to the experimenter. The child was asked, using a forced choice format, to describe how he or she saw the turtle, as well as how the experimenter saw the turtle (i.e., "standing on its feet" or "lying on its back"). The turtle image was rotated 180° and the same questions were asked. This procedure was repeated with two additional animals (bird, pig), yielding six trials in total. The child received one point for each correct answer about the experimenter's perspective (score up to 6).

Vocabulary

Children were administered the Peabody Picture Vocabulary Test, 4th ed. (PPVT-4; Dunn & Dunn, 2007), and standard scores (age-based norms) were used in analyses.

Results

We first report children's performance on the communication task, focusing on performance during

the critical trials. Of particular interest is children's consideration of their door relative to the speaker's door as they hear the ambiguous statement spoken with positive or negative prosody. Children's pointing behaviors are described first, followed by children's eye gaze patterns. Given that predictions vary by trial type, children's performance is examined separately for the potential win and the potential loss trials. Preliminary analyses showed children's performance did not vary as a function of trial order (potential win: $p = .07$ for pointing, $p = .22$ for eye movements; potential loss: $p = .59$ pointing, $p = .17$ eye movements), and thus, order was not included in the analyses. We then examined the relations between performance on the mentalizing and executive function tasks and the communication task.

Pointing Behaviors

Children's pointing behaviors were used as an explicit measure of children's expectations about the location of the objects, based on the speaker's emotional prosody. A research assistant who was unaware of the experimental hypotheses coded the points to the doors from video recordings without sound. Interrater reliability calculated for 20% of the data ($n = 20$; coded by a second assistant) was excellent (Cohen's $\kappa = .955$; $p < .001$). To capture children's pointing behavior with a single measure, we compared points to the child's door across the positive and negative prosody trials within each trial type. The number of points to the child's door was summed and divided by the total number of points for each prosody and trial type, and then converted to a percentage score.

Potential Win Trials

Recall that for these trials, players want the sticker to appear behind their assigned door. Upon hearing the speaker's sad-sounding tone of voice in the negative emotional prosody condition, children should infer that the speaker is sad that she will not win a sticker and thus point to their own door. In the positive emotional prosody condition, children should infer that the speaker is happy because she will win the sticker, and they should therefore point to the speaker's door. As can be seen in Figure 2, children pointed to their own door significantly more than would be predicted by chance for both the negative, $M = 82.7$, $SD = 26.4$, $t(96) = 12.23$, $p < .001$, and positive prosody

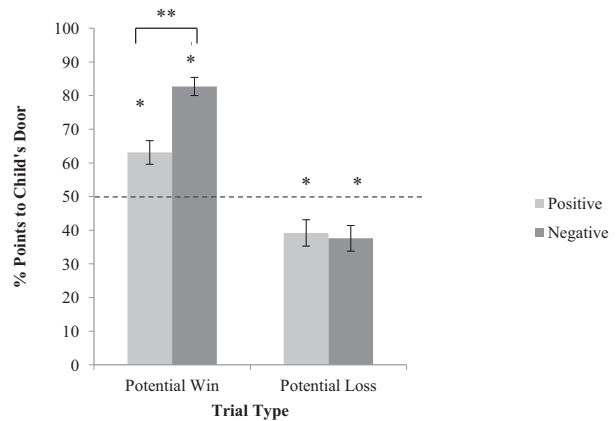


Figure 2. Percent points to the child's door as a function of trial type and emotional prosody condition. * $p < .001$ (vs. chance). ** $p < .001$.

conditions, $M = 63.1$, $SD = 34$, $t(96) = 3.78$, $p < .001$. There was, however, a significant difference between two conditions; children pointed to their own door significantly more often on negative prosody trials than positive prosody trials, $t(96) = 5.28$, $p < .001$, $d = 0.54$.

Potential Loss Trials

Recall that if children were interpreting the utterances from the speaker's perspective, on potential loss trials, negative emotional prosody should lead to the inference that the speaker is sad because she will lose a sticker. Accordingly, they should point to the speaker's door. On positive prosody trials, they should infer the speaker is happy because she will not lose a sticker (i.e., the child will get the X), and they should point to their own door. Children pointed to their door significantly less often than would be predicted by chance for both the negative, $M = 37.6$, $SD = 36.8$, $t(96) = 3.31$, $p = .001$, and positive prosody conditions, $M = 39.2$, $SD = 42.8$, $t(96) = 2.49$, $p = .015$, with no significant differences across conditions ($p = .727$).

In summary, the pointing data suggest that under certain circumstances, preschoolers engage in some degree of emotional perspective taking when making inferences about communicative intent. When children had the opportunity to win a reward, their overall tendency to point to their own door was shifted in the relevant direction by the speaker's emotional prosody. When faced with the prospect of losing a reward, however, children tended to point to the speaker's door regardless of whether the speaker sounded happy or sad. In the next set of analyses, we examine children's eye gaze

as the utterances unfolded and children prepared to make their selections. These implicit measures were expected to capture sensitivities not reflected in children's explicit decisions.

Eye Gaze Patterns

Children's fixations to the two closed doors were examined across four time intervals: baseline, utterance, postutterance, and action planning. The *baseline interval* captured children's visual consideration of the two doors during the 1,000 ms of silence that preceded the onset of the utterance (i.e., before emotional prosody information was available). The subsequent three intervals capture children's interpretation of the ambiguous positive or negative prosody utterances with varying degrees of immediacy. The second interval, *the utterance interval*, began 200 ms after the onset of the first word of the utterance (i.e., the first point at which children had access to emotional prosody information) and continued for the average utterance length (2,400 ms). The 200 ms margin was added in order to adjust for the time required for the eyes to react to unfolding linguistic information (e.g., Allopenna, Magnuson, & Tanenhaus, 1998). We excluded fixations initiated before the beginning of the interval in order to ensure that the eye gaze data reflected a reaction to the prosody information in the speech stream rather than a tendency to continue fixating a particular door. The third interval, *the postutterance interval*, examined children's eye gaze patterns immediately after the offset of the utterance and continued for the same duration as the average length of the utterance (2,400 ms). To capture children's gaze patterns in the seconds immediately preceding their explicit choice of a door, the last interval, *the action-planning interval*, examined children's eye gaze for an additional period of 2,400 ms after the offset of the postutterance interval. Inspection of children's mean latency to point indicated that 85% of children had begun to initiate pointing by the end of the action-planning interval (13% pointed within the postutterance interval). To quantify children's gaze behaviors to the doors using a single measure, we calculated the average proportion of time spent looking to the child's door (the average time spent fixating the child's door divided by the total time spent fixating both doors, within each time interval).

Potential Win Trials

Baseline. A one-sample *t* test against chance (0.50) revealed a tendency for children to fixate

their own door, $M = 0.57$, $SD = 0.29$, $t(96) = 2.28$, $p = .025$. As would be expected, given that there is no linguistic or prosodic information available, there was no difference between positive ($M = 0.58$, $SD = 0.33$) and negative prosody conditions ($M = 0.55$, $SD = 0.34$), $p = .450$.

Interpretation of emotional prosody. Figure 3 displays the proportion of time spent fixating the child's door across the three main time intervals of interest (utterance, postutterance, action planning) as a function of emotional prosody. The measures were submitted to a 2 (prosody: positive vs. negative) \times 3 (interval: utterance vs. postutterance vs. action planning) repeated measures analysis of variance (ANOVA). This analysis revealed significant main effects of prosody, $F(1, 96) = 51.13$, $\eta_p^2 = .348$, $p < .001$, and interval, $F(2, 192) = 7.67$, $\eta_p^2 = .07$, $p = .001$, qualified by a significant Prosody \times Interval interaction, $F(2, 192) = 35.178$, $\eta_p^2 = .268$, $p < .001$. To examine the significant

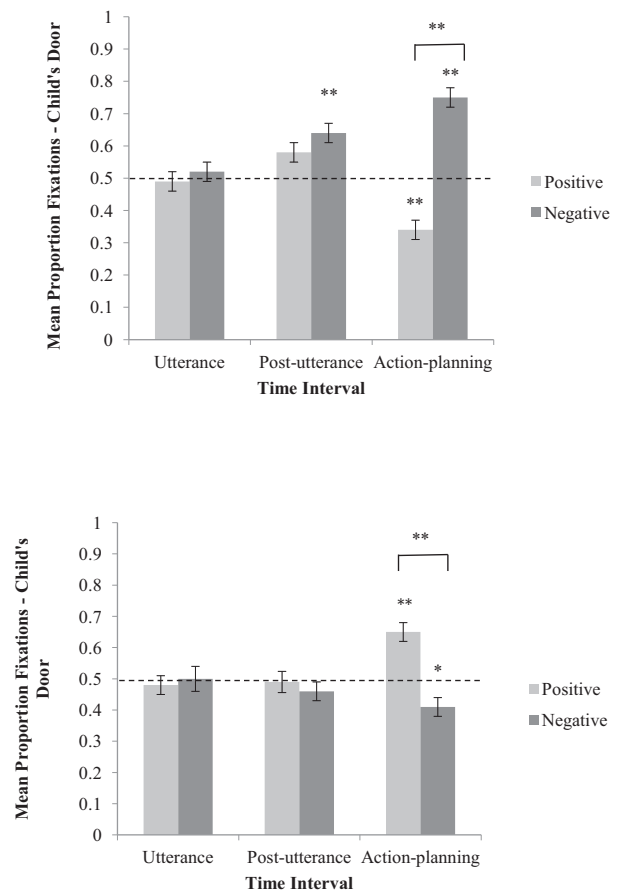


Figure 3. Proportion of time spent fixating the child's door across the three time intervals as a function of emotional prosody condition for potential win trials (top panel) and potential loss trials (bottom panel). * $p = .01$. ** $p < .001$.

interaction, we compared the time children spent fixating their own door in the positive and negative prosody conditions across the three time intervals. This measure was not significantly different across the positive and negative prosody conditions within the utterance (negative: $M = 0.52$, $SD = 0.29$; positive: $M = 0.49$, $SD = 0.29$), $t(96) = 0.76$, $p = .448$, or postutterance intervals (negative: $M = 0.64$, $SD = 0.28$; positive: $M = 0.57$, $SD = 0.33$), $t(96) = 1.75$, $p = .084$. During the action-planning interval, children spent significantly longer looking at their own door in the negative prosody condition ($M = 0.75$, $SD = 0.26$) than the positive prosody condition, $M = 0.34$, $SD = 0.27$), $t(96) = 11.74$, $p < .001$, $d = 1.19$.

We then compared the time children spent fixating their own door against chance (0.50) for each prosody condition and interval using one-sample t tests, with alpha corrected for multiple comparisons to yield an adjusted value of 0.017. Recall that if children were considering the speaker's perspective, they should fixate their own door more than would be expected by chance in the negative prosody condition (i.e., should expect the sticker to be behind their door), and less than chance in the positive prosody condition (i.e., should expect the sticker to be behind the speaker's door). For the negative prosody condition, children's fixations to their door during the utterance did not differ significantly from chance ($p = .441$). However, the proportion of time children spent fixating their own door was significantly greater than chance in the postutterance interval, $t(96) = 5.02$, $p < .001$, and action-planning interval, $t(96) = 9.41$, $p < .001$. For the positive prosody condition, children's looking to their door did not differ significantly from chance during the utterance interval ($p = .775$) or postutterance interval ($p = .023$). During the action-planning interval, however, the proportion of time children spent fixating their door was significantly below chance, $t(96) = 5.97$, $p < .001$. Thus, in the negative emotional prosody condition, children's looking toward their own door began immediately after the utterance had ended and increased over time. In the positive prosody condition, a consistent pattern of looking toward the speaker's door emerged only later, as children began planning their pointing response. There was a moderate positive correlation between children fixating a given door during the action-planning interval and subsequently pointing to that door (averaged across positive and negative emotional prosody), $r = .294$, $n = 97$, $p = .003$.

Potential Loss Trials

Baseline. A one-sample t test comparing children's looking to chance (0.50) revealed a tendency to fixate the speaker's door (i.e., to fixate their own door less than would be predicted by chance), $M = 0.42$, $SD = 0.30$, $t(96) = 2.53$, $p = .013$. The proportion of time children spent looking at their door did not differ between positive ($M = 0.38$, $SD = 0.39$) and negative prosody conditions ($M = 0.46$, $SD = 0.37$), $p = .077$.

Interpretation of emotional prosody. To capture changes in looking over time, the proportion of time children spent looking toward their own door was submitted to the same repeated measures ANOVA described earlier. This analysis revealed a significant main effect of prosody, $F(1, 96) = 10.71$, $\eta_p^2 = .10$, $p = .001$, that was qualified by a significant Prosody \times Interval interaction, $F(2, 192) = 8.80$, $\eta_p^2 = .084$, $p < .001$ (see Figure 3). To explore the significant interaction, we compared children's looking to their own door in the positive and negative prosody conditions across three time intervals. Children fixated their own door to a similar extent in the two emotional prosody conditions during the utterance (negative: $M = 0.50$, $SD = 0.36$; positive: $M = 0.48$, $SD = 0.34$), $t(96) = 0.29$, $p = .770$, and postutterance intervals (negative: $M = 0.46$, $SD = 0.34$; positive: $M = 0.49$, $SD = 0.33$), $t(96) = 0.73$, $p = .468$. They fixated their own door significantly more in the positive prosody ($M = 0.65$, $SD = 0.33$) than the negative prosody ($M = 0.41$, $SD = 0.33$) condition during the action-planning interval, $t(96) = 5.59$, $p < .001$, $d = 0.57$.

Recall that if children were considering the speaker's perspective, they should fixate their own door more than would be expected by chance in the positive prosody condition (i.e., they should expect the X to be behind their door) and less than would be expected by chance in the negative prosody condition (i.e., they should expect the X to be behind the speaker's door). For the positive prosody condition, one-sample t tests (adjusted alpha for three comparisons = .017) indicated that children's looking to their own door during the utterance ($p = .635$) and postutterance region did not differ from chance ($p = .753$). During the action-planning region, the proportion of time children spent fixating their own door was significantly above chance, $t(96) = 4.43$, $p < .001$. For the negative prosody condition, the time children spent looking to their own door during the utterance ($p = .937$) and postutterance region did not significantly differ from chance ($p = .195$). During the

action-planning region, children looked at their own door significantly less often than would be predicted by chance, $t(96) = 2.63$, $p = .010$. Thus, children correctly looked toward their own door in the positive prosody condition and to the speaker's door in the negative prosody condition. This eye gaze pattern emerged immediately prior to making their explicit decisions by pointing. Recall, however, that their pointing decisions did not reflect an appreciation of emotional prosody, with children consistently pointing to the speaker's door regardless of whether she sounded happy or sad. Examination of children's gaze and pointing behaviors revealed a moderate positive correlation between the door children fixated during the action-planning interval and the door to which they eventually pointed (averaged across positive and negative emotional prosody), $r = .377$, $n = 97$, $p < .001$.

Individual Differences and Performance on the Communication Task

Descriptive statistics for the individual difference measures are displayed in Table 1. First, we found no correlation between children's pointing and any of the individual difference measures ($ps > .134$). This may reflect the fact that children's explicit use of speaker perspective was not very apparent in the pointing measures, as discussed earlier. Next, we explored the relation between individual difference measures and the measures of implicit processing provided by the gaze data. To create a single measure representing children's communicative perspective-taking skills, we calculated a composite score by averaging children's mean visual consideration of the correct door across the two conditions (e.g., for potential win trials, looking to the child's door in the negative prosody condition and looking to the speaker's door in the positive prosody condition). Separate composite scores were calculated for each time interval and trial type. Bivariate and partial correlations (controlling for receptive vocabulary) between the individual difference measures and the mean proportion of time spent fixating the correct door are displayed in Table 2.

To further examine the relations among variables, two hierarchical multiple regression analyses (one per trial type) were conducted. The dependent variable was the average proportion of time children spent looking at the correct door during the action-planning interval (as this was when children had demonstrated the most robust evidence of

Table 1
Descriptive Statistics for Individual Difference Measures

	<i>M</i>	<i>SD</i>	Range
Perspective taking			
Turtle task	2.8	2.4	0–6
Puppet task	18.4	2.0	13–20
Executive function			
Day–night task	11.3	4.2	0–16
Picture memory	11.2	2.6	6–16
Delay task	3.4	1.9	0–6
PPVT-4	119.1	12.5	91–150

Note. Reported values reflect those obtained in our sample. PPVT-4 = Peabody Picture Vocabulary Test, 4th ed.

perspective integration). The predictor variable entered on the first step was receptive vocabulary (PPVT). The three executive functioning measures (i.e., day–night, gift delay, picture memory) were entered on the second step. The two mentalizing tasks (puppet and turtle) were entered on the third step.

The first regression analysis examined children's performance on the potential win trials (see Table S1 for model summary and Table S2 for correlations for all measures). The overall model, with all variables in the equation, predicted looking toward the correct door, $R^2 = .161$, $F(6, 90) = 2.88$, $p = .013$. On Step 1, receptive vocabulary contributed significantly to the model, $R^2 = .068$, $F(1, 95) = 6.97$, $p = .010$. The three executive function variables added to the model on Step 2 did not incrementally improve prediction, $\Delta R^2 = .024$, $F_{\text{inc}}(3, 92) = 0.82$, $p = .489$. The addition of the two mentalizing variables on Step 3 significantly improved the model, $\Delta R^2 = .069$, $F_{\text{inc}}(2, 90) = 3.68$, $p = .029$. With all variables included, only the emotional perspective-taking puppet task accounted for unique variance in the model (6.6%, $p = .009$). Thus, emotional perspective-taking and vocabulary measures, but not cognitive perspective-taking or executive function measures, contributed to the prediction of children's integration of the speaker's perspective on the potential win trials (although in the end only the emotional perspective-taking task uniquely predicted performance). The second regression analysis examined children's performance on the potential loss trials. On this analysis, the independent variables did not significantly predict the average proportion of time children spent looking at the correct door during the action-planning interval ($p = .310$ for the overall model).

Table 2

Bivariate and Partial Correlations Between Proportion of Time Spent Fixating the Correct Door During the Communication Task, and Offline Perspective-Taking and Executive Function Measures

	PPVT	Picture memory	Day-night	Delay	Turtle task	Puppet task
<i>Communication task: eye gaze</i>						
Potential win						
Utterance	.00	.01 (.01)	-.01 (-.01)	-.02 (-.02)	.00 (.00)	-.03 (-.03)
Postutterance	.20*	.13 (.12)	.03 (-.07)	.15 (.12)	-.05 (-.09)	.17 (.18)
Action planning	.26**	.17 (.16)	.15 (.04)	.05 (.00)	-.02 (-.07)	.33** (.29)**
Potential loss						
Utterance	-.03	-.08 (-.08)	-.14 (-.14)	-.05 (-.05)	-.04 (-.03)	-.02 (-.18)
Postutterance	.02	-.02 (.01)	.09 (.09)	.05 (.04)	.03 (.03)	.06 (.03)
Action planning	.08	.01 (-.17)	.08 (-.03)	-.01 (-.03)	-.14 (-.16)	.02 (.05)

PPVT = PPVT-4 = Peabody Picture Vocabulary Test. *Significant at < .05 level. **Significant at < .01 level.

Discussion

We investigated preschool children's emotional perspective taking using a novel communication task. By manipulating trial outcomes such that a desired outcome for an adult speaker meant an undesired outcome for the child and vice versa, we created a situation where the expected emotional reaction was the opposite for the adult and child. Correctly interpreting the emotionally tinged utterances produced by the speaker therefore depended on whether children could draw on perspective-taking abilities.

The results showed that children used perspective inferences about a speaker's emotional state to guide their interpretation of ambiguous statements, as evidenced by their implicit gaze behaviors. This occurred despite the salient and emotionally significant conflict between the speaker's perspective (e.g., speaker's happiness about winning) and their own egocentric perspective (e.g., child's sadness about not winning). These findings are broadly consistent with research demonstrating preschoolers' sensitivity to a communicative partner's knowledge (i.e., common ground) during language comprehension (e.g., Liebal et al., 2013; Nadig & Sedivy, 2002; Nilssen & Graham, 2009) and production (e.g., Matthews, Lieven, Theakston, & Tomasello, 2006; Menig-Peterson, 1975). The current study extends our understanding to situations that are emotionally salient and that require the consideration of a communicative partner's affective, rather than cognitive, state.

Interestingly, the question of *when* children demonstrated an appreciation of the speaker's emotional perspective (as evidenced in their eye gaze) depended on whether there was the potential to win versus lose a reward, as well as the

emotional valence of the utterance. On potential win trials, children correctly looked toward the location where the reward would appear if they had won (i.e., their door) immediately after hearing a sad-sounding utterance and continued to do so until they began to point. For the happy-sounding utterances, however, it was not until just prior to pointing that children demonstrated an appreciation of the speaker's emotional perspective (i.e., by correctly looking toward the speaker's door). This gaze pattern suggests that children may initially have been uncertain about how to interpret the positive-sounding statements. When faced with the prospect of losing a reward, children did not show a clear pattern of fixations during or immediately after the utterance. It was not until the moments prior to the onset of children's explicit responses (i.e., pointing) that they showed evidence of integrating the speaker's emotional perspective, regardless of the valence of the utterance. Thus, children showed the earliest appreciation of speaker perspective when there was the potential to win a reward and the speaker sounded sad. A possible explanation for the earlier appreciation of speaker perspective on these trials is that this type of trial may have been most salient to children—relative to the other types of trials—because they signaled that the child was about to win a reward, resulting in greater engagement, excitement, and attention.

Although the emotional prosody cues were present from the onset of the utterance, it was not until roughly 2.5 s after the offset of the utterance that children showed evidence of having integrated the speaker's emotional perspective (with the exception of the potential win negative prosody trials, where the effect was observed earlier). Although at first glance these findings may seem

consistent with the “late integration” effects observed in some visual perspective-taking studies, there are several important factors suggesting a more cautious interpretation is warranted. First, the data do not illustrate a pattern whereby the early moments of processing are distinctly egocentric, only later shifting to an interpretation informed by a consideration of the speaker’s perspective. Rather, the pattern is one where the early moments of processing are neutral, possibly reflecting uncertainty. Second, it is relevant to note the more challenging nature of our task relative to the task used in studies of visual perspective taking (e.g., Epley et al., 2004; Nadig & Sedivy, 2002). Recall that children in the current experiment were presented with two critical trial types, as well as two different emotional prosody conditions, with the significance of the emotional prosody information entirely dependent on trial type. Accordingly, children had to hold the trial type in mind while concurrently processing the linguistic input, emotional prosody, and subsequent perspective inferences. To illustrate the high degree of perspective conflict, consider a trial on which the child had the opportunity to win a reward. Upon hearing a happy-sounding statement, the child had to simultaneously disregard his or her own salient, egocentric interpretation of the utterance (e.g., happy = reward for me) and instead consider the speaker’s incongruent emotional perspective (e.g., happy = no reward for me). In contrast, the way in which a speaker’s knowledge state is inferred in visual perspective-taking studies is typically consistent across all trials, and the relevant cue typically involves a highly salient element of the perceptual environment (namely, a panel or barrier that hides an object from the point of view of a communicative partner).

The results of this study also revealed that children’s explicit judgments and their implicit processing were not highly related. That is, although their gaze patterns suggested children had successfully used the speaker’s emotional perspective to guide their interpretation, this was not reflected in their pointing on the potential loss trials, and was only weakly reflected in their pointing on the potential win trials. Similar differences in implicit understanding versus explicit behavior have been demonstrated in previous research examining preschoolers’ sentence processing, communicative perspective taking, and use of emotional prosody to make referential decisions (e.g., Berman et al., 2010; Berman, Graham, Callaway et al., 2013; Berman, Graham, & Chambers, 2013; Snedeker & Trueswell, 2004).

What underlies the fact that implicit understanding was only moderately reflected in explicit behavior? First, we propose that children’s pointing did not result from initially egocentric language comprehension, but rather resulted from difficulty in behaviorally inhibiting the impulse to point to a desired outcome in an emotionally evocative situation. Preschoolers’ difficulty inhibiting their prepotent pointing to a desired reward has been repeatedly demonstrated within the executive functioning literature (e.g., pointing to an empty box to get a reward, Hala & Russell, 2001; pointing to a smaller reward to get a larger reward, Carlson, Davis, & Leach, 2005). Second, the complexity of our task, coupled with the demands of coordinating linguistic and perspective information, may have taxed preschoolers’ cognitive resources, thus interfering with their explicit performance (for similar disruption of explicit performance in middle childhood, see Wang et al., 2015). Third, 4-year-olds’ emerging sensitivity to emotional prosody cues, relative to the more robust understanding exhibited by 5-year-olds (e.g., Berman et al., 2010; Moore, Harris, & Patriquin, 1993), might have made their referential decision making more susceptible to disruption. Finally, the emotional load associated with the prospect of losing a reward may have impacted perspective taking on the potential loss trials.

Turning now to the issue of individual differences, none of the individual differences assessed contributed to the prediction of performance on the potential loss trials. The results from the potential win trials, however, indicate that children’s use of a speaker’s emotional perspective during communication is related to their emotional perspective-taking abilities. On these trials, we found a relation between children’s use of emotional prosody to understand an ambiguous utterance in real time and their *emotional*—but not *cognitive*—perspective-taking ability (as assessed using an offline task). These results offer insight into the proposal that mentalizing ability—the ability to recognize, represent, and reason about the internal states of others—plays an important role in the development of communicative perspective taking (Nilsen & Fecica, 2011). However, this ability may not be a unitary, domain-general construct. For example, children who demonstrate egocentricity in a visual perspective-taking task do not necessarily show egocentricity in an emotional perspective-taking task (Hoffmann et al., 2015). Relatedly, different brain regions and functions have been implicated when solving visual perspective versus false belief tasks (e.g., Aichhorn, Perner, Kronbichler, Staffen, &

Ladurner, 2006), and in tasks involving emotional perspective versus other kinds of perspective (e.g., Silani, Lamm, Ruff, & Singer, 2013; Steinbeis, Bernhardt, & Singer, 2015). Our findings are therefore consistent with a separation between cognitive and emotional perspective taking. This is compatible with recent adult work suggesting listeners' use of common ground does not rest on a domain-general perspective-taking ability (Ryskin, Benjamin, Tullis, & Brown-Schmidt, 2015), in contrast to earlier suggestions in favor of a domain-general explanation (Wardlow, 2013).

As noted earlier, we did not find a relation between conflict or delay inhibitory control and successful communicative perspective taking. This represents a notable difference between our findings and previous research in which participants with lower inhibitory control were significantly more likely to interpret a speaker's statements egocentrically in a visual perspective-taking task (Brown-Schmidt, 2009; Nilsen & Graham, 2009). It also contrasts with the well-established association between inhibitory control and false-belief understanding (e.g., Devine & Hughes, 2014). One clear possibility is that inhibitory control *per se* is not as important for taking a communicative partner's emotional perspective compared to the kinds of cognitive and visual perspective taking required in past studies. More specifically, suppressing one's cognitive or visual perspective may be accomplished through a kind of cognitive inhibitory control that is distinct from the mechanisms regulating emotional self-control. It could also be the case that our inhibitory control tasks did not measure the kind of "hot" executive functioning (i.e., high-stakes, emotionally charged) implicated in our communication task. That is, a "hot" task might better capture the type of executive function required to successfully engage in emotional perspective taking than the "cool" (i.e., decontextualized and emotionally neutral) day-night task we used to measure conflict inhibitory control (e.g., Zelazo & Carlson, 2012). One recent study (Kramer, Lagattuta, & Sayfan, 2015) found that a happy-sad inhibitory control task (i.e., saying "sad" for the happy face and "sad" for the happy face) was more challenging for both children and adults than other versions of the task (including the day-night version). Thus, including a "hot" task or a task with emotional stimuli may have better captured relevant individual differences. Future research employing a wider range of tasks that measure both cognitive and emotional control may help clarify the role of

various executive functions in children's emotional perspective taking.

Another direction for subsequent research could be to assess the extent to which children are explicitly aware of their emotional perspective taking or whether this processing occurs in a more automatic, implicit manner. For example, children could be asked to provide their reasoning for selecting a certain door. Furthermore, although we opted not to include a neutral emotional prosody condition in the present study (to avoid diluting the effects and saliency of the emotional prosody information), future research comparing online processing times for positive relative to neutral, and neutral relative to negative emotional prosody cues would be helpful in determining how a broader range of emotion categories facilitates or impedes performance. Finally, although we favor the interpretation that children's implicit processing was not reflected in their explicit behavior as a result of various cognitive demands, a potential limitation to this interpretation was not asking a verification question to ensure they understood that pointing to a specific door would not change the outcome.

Future research with older children will offer insight into how various developmental changes influence children's ability to use emotional cues in communicative perspective taking. As noted earlier, 5-year-olds' sensitivity to emotional prosody is more robust than that of 4-year-olds (e.g., Moore et al., 1993; Quam & Swingley, 2012), with significant improvements observed in explicit sensitivity to prosody and the coordination of emotional prosody with visual information between 3 and 5 years of age (Berman, Graham, Callaway et al., 2013; Berman, Graham, & Chambers, 2013; Berman et al., 2016). Children's emotional understanding similarly improves with age (Pons et al., 2004). Thus, even by 5 years of age, children may show much greater concordance between implicit processing and explicit judgments, as well as lower rates of selecting the incorrect door (i.e., desired outcome) for both the potential win and the potential loss trials.

In summary, our findings show that 4-year-olds possess an emergent ability to disregard their egocentric emotional perspective when using a speaker's emotional tone of voice to infer communicative intent. We found a link between this ability and children's performance on a separate measure of emotional perspective taking—one that did not require the rapid, real-time coordination of linguistic and perspective information. When compared with other studies of children's conversational abilities, the findings point to important differences in

the development of communicative perspective taking in the cognitive and emotional domains, mirroring known dissociations in cognitive and emotional mentalizing abilities. These outcomes highlight the value of using a broad range of contexts and methodologies to capture the communicative competence of both children and adults. Finally, enhancing our understanding of children's emotional perspective taking has considerable value, as studies have consistently shown relations between emotion understanding and social competence (Trentacosta & Fine, 2010).

References

- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, *30*, 1059–1068. <https://doi.org/10.1016/j.neuroimage.2005.10.026>
- Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, *67*, 635–645. <https://doi.org/10.1111/j.1467-8624.1996.tb01756.x>
- Alloppenna, 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. <https://doi.org/10.1006/jmla.1997.2558>
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636. <https://doi.org/10.1037//0022-3514.70.3.614>
- Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test-retest reliability. *Journal of Cognition and Development*, *12*, 169–193. <https://doi.org/10.1080/15248372.2011.563485>
- Berman, J. M., 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. <https://doi.org/10.1016/j.jecp.2010.04.012>
- Berman, J. M., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal affect and facial emotional information. *Journal of Experimental Child Psychology*, *142*, 391–399. <https://doi.org/10.1016/j.jecp.2015.09.014>
- Berman, J. M., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to map new words to objects. *Child Development*, *84*, 1791–1805. <https://doi.org/10.1111/cdev.12074>
- Berman, J. M., Graham, S. A., & Chambers, C. G. (2013). Contextual influences on children's use of vocal affect cues during referential interpretation. *The Quarterly Journal of Experimental Psychology*, *6*, 705–726. <https://doi.org/10.1080/17470218.2012.713367>
- Berman, J. M. J., Khu, M., Graham, I., & Graham, S. A. (2013). ELIA: A software application for integrating gaze data and real-time language. *Behavior Research Methods*, *45*, 646–655. <https://doi.org/10.3758/s13428-012-0302-1>
- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*, 893–900. <https://doi.org/10.3758/PBR.16.5.893>
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*, 595–616. https://doi.org/10.1207/s15326942dn2802_3
- Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science*, *16*, 609–616. <https://doi.org/10.1111/j.1467-9280.2005.01583.x>
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032–1053. <https://doi.org/10.1111/1467-8624.00333>
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–397. [https://doi.org/10.1016/0885-2014\(94\)90012-4](https://doi.org/10.1016/0885-2014(94)90012-4)
- Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: Contextual validation. *Child Development*, *57*, 194–201. <https://doi.org/10.2307/1130651>
- Denham, S. A., & Couchoud, E. (1990). Young preschoolers' ability to identify emotions in equivocal situations. *Child Study Journal*, *20*, 153–170. Retrieved from <http://denhamlab.gmu.edu/Publications%20PDFs/Denham%20%20Couchoud%201990a.htm>
- Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *5*, 1777–1794. <https://doi.org/10.1111/cdev.12237>
- Dunn, D. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: PsychCorp.
- Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*, 760–768. <https://doi.org/10.1016/j.jesp.2004.02.002>
- Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, *26*, 1090–1097. <https://doi.org/10.1177/0956797615574699>
- Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology*, *17*, 99–103. <https://doi.org/10.1037/0012-1649.17.1.99>
- Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*, 412–429. <https://doi.org/10.1037//0033-2909.97.3.412>

- Ganea, P. A., & Saylor, M. M. (2007). Infants' use of shared linguistic information to clarify ambiguous requests for objects. *Child Development, 78*, 493–502. <https://doi.org/10.1111/j.1467-8624.2007.01011.x>
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3.5–7 years old on a Stroop-like day-night test. *Cognition, 53*, 129–153. [https://doi.org/10.1016/0010-0277\(94\)90068-X](https://doi.org/10.1016/0010-0277(94)90068-X)
- Hala, S., & Russell, J. (2001). Executive control within strategic deception: A window on early cognitive development? *Journal of Experimental Child Psychology, 80*, 112–141. <https://doi.org/10.1006/jecp.2000.2627>
- Harris, P. L., de Rosnay, M., & Pons, F. (2016). Understanding emotion. In M. Lewis, J. Haviland-Jones & L. Feldman Barrett (Eds.), *Handbook of emotions* (4th ed., pp. 293–306). New York, NY: Guilford.
- Harris, P. L., Johnson, C., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition and Emotion, 3*, 379–400. <https://doi.org/10.1080/02699938908412713>
- Hoffmann, F., Singer, T., & Steinbeis, N. (2015). Children's increased emotional egocentricity compared to adults is mediated by age-related differences in conflict processing. *Child Development, 86*, 765–780. <https://doi.org/10.1111/cdev.12338>
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41. [https://doi.org/10.1016/S0010-0277\(03\)00064-7](https://doi.org/10.1016/S0010-0277(03)00064-7)
- Kramer, H. J., Lagattuta, K. H., & Sayfan, L. (2015). Why is happy-sad more difficult? Focal emotional information impairs inhibitory control in children and adults. *Emotion, 15*, 61–72. <https://doi.org/10.1037/emo0000023>
- Liebal, K., Carpenter, M., & Tomasello, M. (2013). Young children's understanding of cultural common ground. *British Journal of Developmental Psychology, 31*, 88–96. <https://doi.org/10.1111/j.2044-835X.2012.02080.x>
- Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development, 81*, 597–615. <https://doi.org/10.1111/j.1467-8624.2009.01418.x>
- Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development, 45*, 357–366. <https://doi.org/10.2307/1127956>
- Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two- and four-year-olds learn to adapt referring expressions to context: Effects of distracters and feedback on referential communication. *Topics in Cognitive Science, 4*(2), 184–210. doi: 10.1111/j.1756-8765.2012.01181.x
- Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics, 27*, 403–422. <https://doi.org/10.1017/S0142716406060334>
- Menig-Peterson, C. L. (1975). The modification of communicative behavior in preschool-aged children as a function of the listener's perspective. *Child Development, 46*, 1015–1018.
- Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen-month-olds know what 'we' have shared in a special way. *Infancy, 13*, 90–101. <https://doi.org/10.1080/15250000701779402>
- Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology, 24*, 603–613. <https://doi.org/10.1348/026151005x55370>
- Moore, C., Harris, L., & Patriquin, M. (1993). Lexical and prosodic cues in the comprehension of relative certainty. *Journal of Child Language, 20*, 153–167. <https://doi.org/10.1017/S030500090000917X>
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*, 329–336. <https://doi.org/10.1111/1467-9280.00460>
- Nilsen, E. S., & Fecica, A. M. (2011). A model of communicative perspective-taking for typical and atypical populations of children. *Developmental Review, 31*, 55–78. <https://doi.org/10.1016/j.dr.2011.07.001>
- Nilsen, E. S., & Graham, S. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58*, 220–249. <https://doi.org/10.1016/j.cogpsych.2008.07.002>
- Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1*, 127–152. <https://doi.org/10.1080/17405620344000022>
- Quam, C., & Swingle, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development, 83*, 236–250. <https://doi.org/10.1111/j.1467-8624.2011.01700.x>
- Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. *Journal of Experimental Psychology: General, 144*, 898–915. <https://doi.org/10.1037/xge0000093>
- San Juan, V., Khu, M., & Graham, S. A. (2015). A new perspective on children's communicative perspective taking: When and how do children use perspective inferences to inform their comprehension of spoken language? *Child Development Perspectives, 9*, 245–249. <https://doi.org/10.1111/cdep.12141>
- Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *Journal of Neuroscience, 33*, 15466–15476. <https://doi.org/10.1523/JNEUROSCI.1488-13.2013>
- Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238–299. <https://doi.org/10.1016/j.cogpsych.2004.03.001>

- Steinbeis, N., Bernhardt, B. C., & Singer, T. (2015). Age-related differences in function and structure of rSMG and reduced functional connectivity with DLPFC explains heightened emotional egocentricity bias in childhood. *Social, Cognitive and Affective Neuroscience, 10*, 302–310. <https://doi.org/10.1093/scan/nsu057>
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634. <https://doi.org/10.1126/science.7777863>
- Trentacosta, C. J., & Fine, S. E. (2010). Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Social Development, 19*, 1–29. <https://doi.org/10.1111/j.1467-9507.2009.00543.x>
- Wang, J. J., Ali, M., Frisson, S., & Apperly, I. (2015). Language complexity modulates 8- and 10-year-olds' success at using their theory of mind abilities in a communication task. *Journal of Experimental Child Psychology, 130*, 1–12. <https://doi.org/10.1016/j.jecp.2015.09.006>
- Wardlow, L. (2013). Individual differences in speakers' perspective taking: The role of executive control and working memory. *Psychonomic Bulletin & Review, 20*, 766–772. <https://doi.org/10.3758/s13423-013-0396-1>
- Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence™* (4th ed.). San Antonio, TX: PsychCorp.
- Yuill, N. (1984). Young children's co-ordination of motive and outcome in judgments of satisfaction and morality. *British Journal of Developmental Psychology, 2*, 73–81. <https://doi.org/10.1111/j.2044-835X.1984.tb00536.x>
- Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives, 6*, 354–360. <https://doi.org/10.1111/j.1750-8606.2012.00246.x>

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Table S1. Results of Hierarchical Multiple Regression Analysis Predicting Implicit Emotional Perspective Taking on the Communication Task From Vocabulary, Executive Functions, and Mentalizing Skills

Table S2. Bivariate Correlations (Pearson Correlation Coefficients) for All Measures