Preschoolers Flexibly Shift Between Speakers' Perspectives During Real-Time Language Comprehension

Melanie Khu
University of Calgary

Craig G. Chambers (iD)
University of Toronto

Susan A. Graham (iD)
University of Calgary

In communicative situations, preschoolers use shared knowledge, or "common ground," to guide their interpretation of a speaker's referential intent. Using eye-tracking measures, this study investigated the time course of 4-year-olds' ($n = 95$) use of two different speakers' perspectives and assessed how individual differences in this ability related to individual differences in executive function and representational skills. Gaze measures indicated partner-specific common ground guided children's interpretation from the earliest moments of language processing. Nonegocentric online processing was positively correlated with performance on a Level 2 visual perspective-taking task. These results demonstrate that preschoolers readily use the perspectives of multiple partners to guide language comprehension and that more advanced representational skills are associated with the rapid integration of common ground information.

A fundamental aspect of successful communication is the ability to recognize and track the information that is (or is not) shared with a conversational partner, such as whether a conversational partner is aware of facts x and y, or whether object z is located within the partner's field of view. One critical component of the ability to monitor shared knowledge, or "common ground" (Clark, 1992; Stalnaker, 1975), and use this information to guide communicative behavior is *communicative perspective taking*. Here, we address three critical questions about the development of perspective-taking abilities. First, can preschoolers flexibly manage distinct common ground representations for two different speakers in the same communicative environment? Second, is perspective information used in the early

moments of processing linguistic reference? Finally, which broader cognitive abilities support preschoolers' real-time perspective taking? We address these questions by examining 4-year-olds' use of partner-specific common ground during online language processing and by investigating how individual differences in this ability relate to information processing and representational skills.

During the preschool years, children undergo critical gains in the ability to use shared knowledge to guide both the production and interpretation of referential language (Köymen, Mammen, & Tomasello, 2016; Matthews, Lieven, Theakston, & Tomasello, 2006; Nayer & Graham, 2006; Nilsen & Graham, 2009, Nilsen, Graham, Smith, & Chambers, 2008; see Bohn & Köymen, 2018, Graham, San Juan, & Khu, 2017, San Juan, Khu, & Graham, 2015 for reviews). Most relevant to the current study is research examining preschoolers' integration of perspective information in real-time language comprehension. These studies typically combine eye-tracking paradigms with referential communication tasks in which conversational partners have different knowledge or different perspectives regarding the objects available for reference. When a speaker requests an object using

an utterance like "Pass me *the cup*," the listener's goal is to identify a candidate referent within the mutually available common ground, excluding, for example, any objects that are perceptible to the listener yet are hidden from and are therefore unknown to the speaker. The efficiency of this process is revealed by the timing and pattern of eye movements as the description unfolds (e.g., the speed with which the target object is identified and which other objects are temporarily considered). Using this paradigm, some studies have suggested children are initially egocentric in the earliest moments of processing, considering referential candidates unknown to the speaker (i.e., objects in the child's "privileged ground"). For example, Epley, Morewedge, and Keysar (2004) presented children aged 4–12 years and adults with displays containing three objects of the same kind that differed in size (e.g., a large, medium, and small truck). The speaker instructed the listener to pick up one of the set of three objects using a singular definite noun phrase (e.g., "Please move the *small truck* above the glue"). Critically, the best referential match with the description's semantics was not in common ground (e.g., the smallest truck was exclusively visible to the child, but the medium truck and large truck were mutually visible). Gaze latencies measured from noun onset showed children first considered the hidden, smallest truck (1,497 ms, on average) and did not identify the target referent (i.e., the medium truck) until well past the offset of the noun (more than 3,647 ms later). This was taken as evidence that children initially interpret language from their own perspective, and only integrate information about common ground at a later point if linguistic input is insufficient to disambiguate the referent. Adults in the same study similarly showed an initial tendency to first consider the object that was not in common ground (i.e., the smallest truck). Similar patterns of initially "egocentric" processing have been reported in other studies of children and adults (e.g., Fan, Liberman, Keysar, & Kinzler, 2015; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003; Kronmüller & Barr, 2007).

However, other work has shown more effective use of common ground information, even in the early moments of processing. Nadig and Sedivy (2002) used a paradigm in which 5- and 6-year-old children were instructed to select an object from displays containing two objects of the same kind (i.e., a target and competitor), which matched the speaker's description to the same degree. For example, the child might be asked to select a glass from a display containing two glasses and two unrelated objects (e.g., baseball and crayon). The target

(referred to as "the glass") was always presented in common ground—that is, visible to both the child and the speaker. On some trials, the competitor was also in common ground, whereas on others the competitor was obscured from the speaker by a panel such that it was visible only to the child. Children consistently demonstrated sensitivity to the speaker's perspective as the description was processed, rarely considering the competitor when it was obscured from the speaker's view. Furthermore, children's preferential fixations to the target over the competitor were evident even in the earliest moments (i.e., 200–760 ms after the onset of the noun in the critical instruction). Other evidence of early integration of perspective inferences has been observed in multiple studies conducted with adults and children (e.g., Brown-Schmidt, 2012; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Ferguson & Breheny, 2012; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008; Mozuraitis, Chambers, & Daneman, 2015; Nilsen & Graham, 2009).

One way of reconciling the discrepant findings is provided by constraint-based accounts of language processing (e.g., Hanna et al., 2003; Heller, Parisien, & Stevenson, 2016; Tanenhaus & Trueswell, 1995). According to the constraint-based framework, listeners incorporate multiple sources of information simultaneously. In the case of communicative perspective taking, this means the simultaneous weighing of probabilistic cues that signal the relevant perspective to adopt for language comprehension. The fact that utterances such as questions and requests can require the listener to consider privileged knowledge rather than common ground demonstrates the need for comprehension mechanisms to continuously assess which perspective to consider from moment to moment (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Smyth, 1995). A second type of informational constraint involves the linguistic fit of a referential expression for a given referent ("the small truck" vs. "the truck"). To illustrate, in the Epley et al. (2004) study, the smallest truck in the display (i.e., the smallest of three trucks, and one that was not mutually visible to the speaker and listener) was the best linguistic match for the referential expression, namely "the small truck." Linguistic fit therefore provided a constraint favoring an object in *privileged* ground. Recall that in this case, children were not as successful at excluding consideration of the privileged-ground candidate.

Conversely, in the Nadig and Sedivy (2002) study, children were faced with two identical

objects that equally matched the referential expression ("the glass")—one in common ground and the other in privileged ground (i.e., hidden from the speaker's visual perspective). Here, children were more successful excluding privileged-ground candidates, demonstrating a preference for shared referents (i.e., sensitivity to speaker perspective) within the earliest moments of processing. A constraint-based approach thus can account for a wide variety of patterns reported in the literature, ranging from studies showing apparently early integration of common ground to those showing apparently late integration of common ground in both adults and children.

The idea of drawing on multiple sources of information simultaneously raises the question of how listeners deal with a situation involving *multiple* speakers and their differing perspectives. There is evidence to suggest that even young children have a basic ability to track the perspectives of distinct communicative partners. That is, infants can track the experiences and information shared with different individuals, and use this information to correctly identify the referent of an ambiguous request (e.g., Akhtar, Carpenter, & Tomasello, 1996; Moll & Tomasello, 2007). Preschoolers also take partner-specific information into account when producing referring expressions (Köymen, Schmerse, Lieven, & Tomasello, 2014; Matthews, Butcher, Lieven, & Tomasello, 2012) and demonstrate sensitivity to partner-specific referential precedents (Graham, Sedivy, & Khu, 2014; Matthews, Lieven, & Tomasello, 2010). In these studies, however, children only had to consider the perspective of one speaker at any given time, and in some cases, physical co-presence was confounded with perspective (i.e., the relevance of a given speaker's perspective was signaled by that speaker's physical presence). As a result, little is known about whether preschoolers can flexibly and accurately manage different common ground representations for alternating, co-present speakers, and how switching between distinct perspectives might impact the time course of language processing. Here, we pursue this question by examining 4-year-olds' sensitivity to two different speakers' perspectives during real-time language processing. Furthermore, although previous research (Nadig & Sedivy, 2002) has demonstrated robust and rapid perspective integration in school-aged children (i.e., 5- to 6-year-olds), no study has examined the time course of common ground use in a younger cohort.

A focus on 4-year-olds in particular is relevant because there are dramatic improvements in communicative perspective taking over the course of the preschool years (Nilsen & Graham, 2009). These occur alongside gains in representational skills, executive function, and other forms of perspective taking (e.g., visual perspective taking) between 3 and 5 years of age (Garon, Bryson, & Smith, 2008; Wellman & Liu, 2004). Exploring individual differences within this developmental period can therefore yield a more complete understanding of the link between these abilities and the development of communicative perspective taking.

Previous explorations of the role of individual differences have tended to focus on executive function, which refers to the deliberate, top-down processes regulating the goal-directed control of emotions, thoughts, and actions, and which is described as involving the maintenance and updating of information in working memory, inhibitory control, and cognitive flexibility (Miyake et al., 2000). Specifically, it has been posited that better executive function would help support the process of adjudicating between one's own perspective and an internal representation of a partner's perspective (Nilsen & Fecica, 2011), and thus one would expect positive relations between these abilities. In studies to date, use of common ground has been associated with both working memory (e.g., Lin, Keysar, & Epley, 2010; Schuh, Eigsti, & Mirman, 2016; Wardlow, 2013: Wardlow & Heyman, 2016) and inhibitory control (e.g., Brown-Schmidt, 2009; Nilsen & Fecica, 2011; Nilsen & Graham, 2009, 2012; Symeonidou et al., 2016). However, the precise link between communicative perspective taking and executive function is far from clear. Indeed, inconsistent patterns can be found across the perspective-taking literature. For example, Nilsen and Graham (2009) found that children with better inhibitory control were less egocentric in communicative perspective taking, yet they found no link between communicative perspective taking and working memory or cognitive flexibility. In contrast, Nilsen and Bacso (2017) found that working memory strongly predicted adolescents' communicative perspective taking (see Schuh et al., 2016 for similar findings).

Adding to this already complicated state of affairs, the information processing abilities captured under the umbrella of executive function are often understood as formally distinct from the ability to accurately represent the mental states of another person (e.g., Nilsen & Fecica, 2011). This latter ability relates to the actual content of mental representations rather than to the maintenance or manipulation of information by cognitive

mechanisms. As noted above, children's ability to explicitly reason about and represent others' mental states emerges during the preschool years (Wellman, Cross, & Watson, 2001; Wellman & Liu, 2004). Thus, it is possible that individual variation in communicative perspective taking depends not only on the ability to manage and manipulate information but also the ability to build an accurate internal representation of another person's knowledge or perspective. Accordingly, children with more advanced representational skills may demonstrate stronger communicative perspective taking. To our knowledge, no study to date has examined the separate contributions of executive function and representational skills to children's use of common ground in real-time language comprehension.

In the current experiment, we presented 4-year-olds with a referential visual perspective-taking task in which two speakers took turns giving the child instructions. On competitor trials, the child was presented with a visual display containing two same-category objects (a "target" and a "competitor," e.g., two shoes) along with two unrelated objects. Displays were accompanied by spoken instructions to point to objects on the screen (e.g., "Point to the shoe"). During *shared-competitor* trials, both the active speaker (i.e., the person giving the instructions) and the child had visual access to the target and competitor. During *privileged-competitor* trials, the child again had visual access to both objects, but the competitor was occluded from the active speaker by a barrier. The target and competitor objects were equivalent in terms of their referential fit with the description provided by the speaker ("the shoe") to ensure the potential influence of this cue was held constant across conditions (cf. Heller et al., 2016).

We used real-time eye movement measures to assess whether preschoolers consider their partners' perspectives as the instructions unfolded. Children's points to the target relative to the competitor were also measured to assess the degree to which implicit referential hypotheses were reflected in children's controlled, conscious judgments. Our predictions were as follows: If preschoolers do not take speaker-specific common ground into account (i.e., they interpret the utterance from their own perspective), then they should perceive the active speaker's description to be referentially ambiguous in both the shared-competitor and privileged-competitor conditions. In addition to random points, this would be reflected in the visual consideration of both the target and the competitor as potential

referents. In contrast, if preschoolers use the speakers' perspectives to inform their interpretations, then they should favor the candidate that is visually accessible to the speaker in the privileged-competitor condition rather than the competitor object that is hidden from the speaker's view. With regard to time course, if preschoolers' interpretations are initially egocentric, in keeping with the proposals advanced by Epley et al. (2004), we should observe no difference in fixations to the competitor in the shared-competitor and privileged-competitor conditions during the early moments of processing the noun in the speaker's instruction. Alternatively, if information about speakers' perspectives serves as an immediate constraint, in keeping with the multiple constraints perspective, children should show early consideration of the target over the competitor in the privileged-competitor, but not the shared-competitor, condition.

To assess the abilities that contribute to successful communicative perspective taking, we measured executive function, receptive vocabulary, and representational skills (using two visual perspective-taking tasks). If variability in preschoolers' communicative perspective taking reflects differences in general information processing, rather than differences related to representing the speakers' perspectives, then we should find significant positive correlations between children's executive function and their success at communicative perspective taking. Alternatively, if individual differences in communicative perspective taking are related to differences in the ability to maintain distinct representations of common ground, children with more advanced representational skills should exhibit better communicative perspective taking.

Method

Participants

Participants were ninety-five 4-year-olds (48 girls; M = 4.3, SD = .15, range = 4 years 0 months to 4 years 6 months), recruited as part of a larger study examining communicative development. None of the children had reported developmental disorders, and all were native speakers of English. An additional seven children completed the testing procedure but were excluded from the final sample due to insufficient eye gaze data resulting from complete track loss for one or more experimental conditions (n = 5) or uncooperativeness/difficulty following task instructions (n = 2). The majority of

parents reported completing postsecondary education (88% had completed postsecondary and/or graduate studies), and the majority of families self-identified as Caucasian (Caucasian = 86%; multiracial = 9%; Asian or Other = 5%). Parents and children attended two 1-hr long testing sessions within 1 week of each other. Parents completed standardized questionnaires and children participated in one additional communicative perspective-taking task (data reported in Khu, Chambers, & Graham, 2018). Children completed one communication task per testing session, which was followed by the individual difference measure tasks.

Communication Task

Children were seated on a chair facing a 46-in. display screen located on a short table. The experimenter stood behind the screen, facing the child. One speaker (S1) was seated to the child's left while a second speaker (S2) sat to the child's right (see Figure 1). The experimenter introduced the child to S1 and S2, and placed a large opaque barrier directly in front of the child such that it divided the screen in half vertically. The experimenter explained that although the child could see both sides of the screen when the barrier was in place, S1 and S2 could only see things displayed on their own side of the screen. To highlight that S1 and S2 could only see objects on their respective sides, the session began with a "guessing game" during which the experimenter and child generated clues for S1 to guess what was on the left side of the screen, and for S2 to guess what was on the right side of the screen. To underscore that S1 and S2 could not see objects on the other side of the barrier, their first guesses provided by these speakers were always incorrect, followed by a correct guess after two clues had been given.

After the guessing game, the child was presented with six trials in each of three within-subject conditions: a *shared-competitor* condition, a *privileged-competitor* condition, and a *no barrier* condition. Note that condition names refer to the competitor object from the perspective of the *child*. On each trial, a visual display containing four objects (one per quadrant) appeared on screen for 2.5 s to allow the child time to inspect the objects. Next, a black
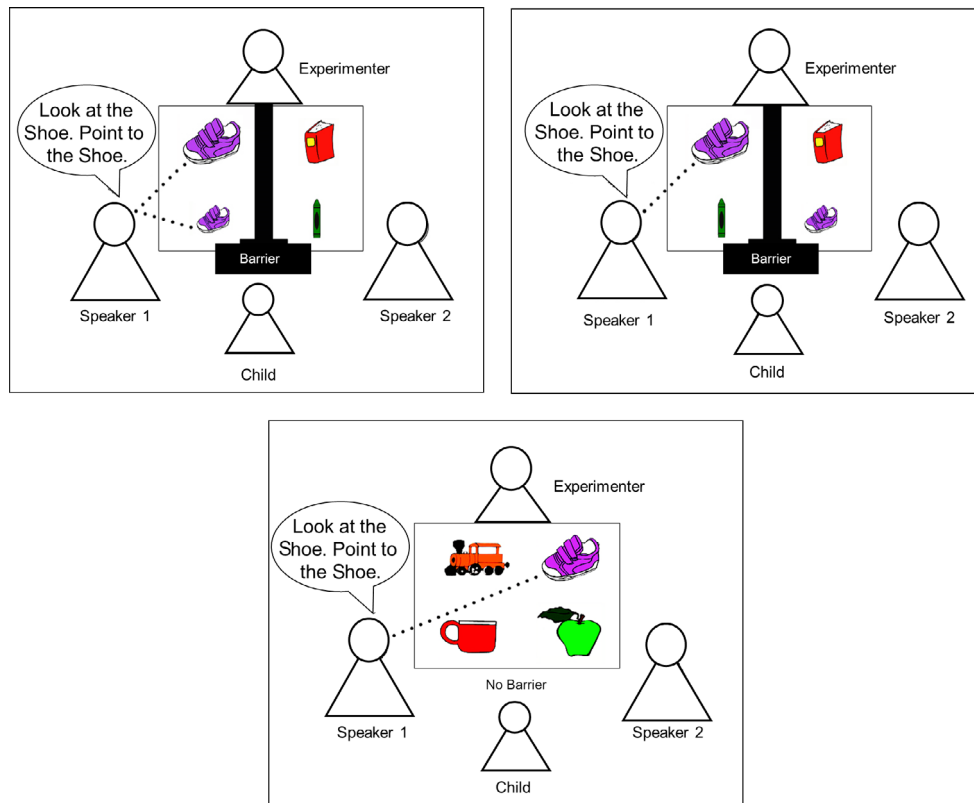

Figure 1. Sample trials from the shared-competitor condition (top left), privileged-competitor condition (top right), and no barrier condition (bottom).

screen appeared for 1 s, followed by the reappearance of the initial visual display. Upon the reappearance of the display, either S1 or S2 gave the child instructions relating to an object within the array ("Look at the X. Point to the X."). The experimenter controlled the presentation of images on the screen and removed or replaced the barrier between trials according to the experimental condition (described below) but did not speak during the main task. Trials were presented in one of four fixed pseudo-randomized orders, with no more than two trials of the same condition presented sequentially. S1 and S2 were each the active speaker for half the trials in each condition. Content for each experimental order and a list of visual stimuli is presented in Supporting Information.

Competitor Trials

On these trials, displays contained two objects, a target and a competitor, that belonged to the same category but differed in size (e.g., a large shoe and a small shoe), as well as two unrelated objects (e.g., book and crayon). The barrier was present on every competitor trial (i.e., both shared-competitor and privileged-competitor trials). As noted above, the child participant always had visual access to objects presented on either side of the barrier. On shared-competitor trials, the active speaker (i.e., speaker giving instructions) had visual access to the target and competitor, as these objects were presented on the side of the barrier close to that speaker (Figure 1, top left panel). The inactive speaker could not see these objects. On privileged-competitor trials, the active speaker had visual access to the target only (i.e., the competitor was displayed on the other side of the barrier; Figure 1, top right panel). To a skilled listener, the instruction given by S1 (e.g., "Look at the shoe. Point to the shoe") would be referentially ambiguous only in the shared-competitor condition because the presence of both potential referents (e.g., both shoes) was known to this speaker. In the privileged-competitor condition, a skilled listener should understand that the intended referent is the sole member of the same-category pair that S1 can see, and thus, linguistically ambiguous description is in fact in relation to the speaker's perspective. The object pairs were counterbalanced across conditions (e.g., for one group of children, a pair of cups would appear in the shared-competitor condition, and for the other group it would appear in the privileged-competitor condition). The size (big vs. small), side (left vs. right), and vertical positioning (bottom vs. top) of the target were counterbalanced.

No Barrier Trials

The purpose of these trials was to attenuate the expectation that the speakers would always refer to one member of a pair of objects or to objects on the side of the screen closest to them. For the no barrier trials, all objects were visible to both speakers, as well as to the child (see Figure 1, bottom panel). Displays contained either four unrelated objects (four trials) or a pair of same-category objects plus two unrelated objects, as in the shared-competitor and privileged-competitor conditions (two trials). S1 or S2 unambiguously referred to one of the unique objects in the display (i.e., the same-category objects were not referred to on these trials). For two of the no barrier trials, the target was located on the same side of the screen as the active speaker, and for the other four trials, the target was located on the opposite side of the screen.

Apparatus

The experiment was programmed and presented using E-Prime software with Tobii Extensions. Children's eye movements were recorded using a Tobii x50 (Tobii Group, Stockholm, Sweden) table-mounted eye tracker. A calibration procedure was conducted using Clearview software. Ninety-seven percent of children recorded 5/5 successful calibration points per eye, with the remainder ($n = 3$) recording 4 points per eye. Specific areas of interest corresponding to the location of display objects were defined prior to data collection. Eye movement data were recorded every 20 ms following the reappearance of the objects following the blank screen, and fixations were logged when the eye was stable for 100 ms or longer.

An HD camera, positioned behind the child, was used to record the live speakers and children's pointing behavior. Videos were coded on a frame-by-frame basis using FinalCut Pro 5.0.4 (Apple, Cupertino, CA, USA) to identify the onset and offset of each word, as well as the beginning and end of each trial. The trial bounds and speech landmarks were synchronized with the eye gaze data using Eye-gaze Language Integration software (Berman, Khu, Graham, & Graham, 2013).

Individual Differences Measures

Individual differences measures were administered after the communication task. The procedures used to collect these measures are described below.

Vocabulary

The Peabody Picture Vocabulary Test, 4th ed. (PPVT–4; Dunn & Dunn, 2007) was administered to measure receptive vocabulary. Children's raw scores were converted to standard scores using age-based norms.

Executive Functions

Working memory. The Picture Memory subtest of the Wechsler Preschool and Primary Scale of Intelligence, 4th ed. (Wechsler, 2012) was administered to assess working memory. Age-based Canadian norms were used to convert raw scores to standard scores using (Wechsler, 2012).

Conflict inhibitory control. The Stroop-like "day–night" task (Gerstadt, Hong, & Diamond, 1994) was used to evaluate conflict inhibitory control. First, the child was asked what is seen in the sky during the day (the sun) and what is seen in the sky at night (moon and stars). Next, the child was shown a card depicting the stars and moon in a dark night sky and another card showing the sun in a light daytime sky. The child was instructed to say "night" when presented with the sun card, and "day" when presented with the moon and stars card. Following three practice trials during which corrective feedback was provided, 16 test trials were administered in pseudorandom order. Self-corrections were counted as incorrect (e.g., Gerstadt et al., 1994). The data for 20% of children ($n = 20$) were coded by a research assistant. Interrater reliability was excellent (Cohen's $\kappa = .946$; $p < .001$).

Delay inhibitory control. A reward task, based on Beck, Schaefer, Pang, and Carlson (2011), was used to assess delay inhibitory control. Children made a series of decisions about receiving a smaller reward immediately or a larger reward later. Specific rewards were colorful balls and small erasers. 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; 1 versus 2 erasers. The child received a point for each trial in which they chose the delayed reward, for a total score out of 6.

Representational Skills

Visual perspective taking. As a primary measure of representational skills, children were tested using the Level 2 visual perspective "turtle task" (Flavell, Everett, Croft, & Flavell, 1981; Masangkay et al., 1974). Visual perspective-taking tasks were

chosen instead of measures assessing other mental state representations because children's spontaneous use of differences in visual perspective was most directly relevant to the communication task. The child sat across the table from the experimenter, who placed a picture of a turtle (depicted in profile) between them. The picture was placed such that the turtle appeared upside-down to the experimenter and right side up to the child. Using a forced choice format, the child was asked to describe how they saw the turtle and how the experimenter saw the turtle (i.e., "standing on its feet" or "lying on its back"). The image was then rotated 180° and the child was asked to again describe how they, and then how the experimenter, saw the turtle. The procedure was repeated for an additional two depicted animals (bird, pig), resulting in six trials in total. For each correct answer about the experimenter's perspective, the child received 1 point, yielding a total score out of 6.

To ensure children possessed adequate Level 1 visual perspective taking abilities to complete the communicative perspective-taking task, the classic Level 1 task of Masangkay et al. (1974) was also administered as a screening measure. A card with different animals on either side was held vertically between the child and the experimenter, and the child indicated what animal each person could see. This procedure was repeated for two additional trials, yielding a score out of 3. As expected given the participants' age, all children obtained a score of at least 2/3, with 87 of 95 (92%) obtaining a perfect score.

Results

Our analyses focus first on children's performance on the communication task. Of particular interest was children's consideration of the competitor across the two competitor conditions. We describe pointing behaviors first, followed by eye gaze patterns. We then examine children's performance on the representational skills and executive function tasks in relation to communicative perspective taking.

Pointing Behaviors

Pointing behavior was used as an explicit measure of children's perspective taking. A research assistant who was unaware of the experimental hypotheses coded children's points to the four on-screen objects for each trial from video recordings

without sound. A second assistant recoded 20% of videos to establish interrater reliability ($n = 20$), which was high (Cohen's $\kappa = .822$; $p < .001$). Children's selections were not mutually exclusive; that is, if they simultaneously pointed to both the target (e.g., shoe the speaker could see) and the competitor (e.g., shoe blocked from the speaker by the barrier), a point was counted for each object (see analysis below). The number of points to the competitor was summed and divided by the total number of points to the target and competitor for each condition for a given participant and then converted to a percentage score. In the shared-competitor condition, where the two same-category objects were on the same side of the barrier (e.g., two shoes), the object coded as the competitor was counterbalanced for size and location on screen (left vs. right; top vs. bottom). For the no barrier condition, the object coded as the competitor was a randomly selected nontarget item. When trials containing points to both objects were discounted, children pointed to the competitor on 36% of shared-competitor (ambiguous) trials and 9% of privileged-competitor (unambiguous) trials. In the no barrier condition, children pointed to the unambiguous target at near ceiling rates (95% of trials).

Three analyses were performed. First, a one-way repeated measures analysis of variance (ANOVA) was conducted to compare points to the competitor relative to the target across the three conditions, yielding a significant effect of condition, $F(1.68, 6.26) = 258.03$, $p < .001$, $\eta_p^2 = .733$. The Greenhouse–Geisser correction was used as Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(2) = 19.56$, $p < .001$. A follow-up paired-samples t test, with alpha corrected for multiple comparisons to yield an adjusted value of .017, confirmed that children pointed to the competitor significantly more often in the shared-competitor ($M = 46.9\%$, $SD = 15.7\%$) than in the privileged-competitor ($M = 17.7\%$, $SD = 20.3\%$) condition, $t(94) = 12.04$, $p < .001$, $d = 1.25$, with the large effect size providing robust evidence that children had taken the speakers' visual perspectives into account. As would be expected given the lack of linguistic ambiguity, follow-up paired-samples t tests also confirmed that children pointed to the competitor significantly less in the no barrier condition ($M = 0\%$, $SD = 0\%$) than in the shared-competitor, $t(94) = 28.66$, $p < .001$, $d = 3.96$, and privileged-competitor, $t(94) = 8.33$, $p < .001$, $d = 1.18$, conditions (both large effects). Next, we compared the number of points to the competitor in the shared-competitor condition and the privileged-competitor condition to the level that would be expected by chance alone (50%) using single-sample t tests. Consistent with expectations, children's pointing to the competitor in shared-competitor condition did not differ significantly from chance, $t(94) = 1.91$, $p = .059$, $d = 0.20$ (small effect). That is, when faced with the ambiguous instruction, children were as likely to choose the competitor as they were to choose the target, both of which were mutually visible. Recall that in the privileged-competitor condition, the competitor was the same-category object blocked from the speaker's view. In this condition, children pointed to the competitor significantly less often than would be predicted by chance (50%) $t(94) = 15.94$, $p < .001$, $d = 1.59$ (large effect). Thus, children were much less likely to choose the object that was hidden from the speaker than the mutually visible object, even though both objects were equally appropriate semantic matches with the referential description.

In the third analysis, we compared how often children pointed to both objects across the shared-competitor and privileged-competitor conditions. This measure provides another way to quantify children's sensitivity to the referential ambiguity in the shared-competitor condition. Children pointed to both objects significantly more in the shared-competitor condition ($M = 17.4\%$, $SD = 25.9\%$) relative to the privileged-competitor condition ($M = 12.1\%$, $SD = 22.1\%$), $t(94) = 2.16$, $p = .033$, $d = 0.23$. This pattern suggests that children tended to understand how the barrier prevented each active speaker from seeing the same-category object located on the opposite side, which in turn determined the objects available for reference by that speaker, although it should be noted that the effect size was relatively small.

Eye Gaze Patterns

To examine the time course of processing, we examined children's fixations to the competitor object relative to the target as the critical noun was heard. Our primary analysis focused on children's gaze behavior in the shared-competitor versus the privileged-competitor conditions. The no barrier condition was included in this analysis as these trials provide a benchmark for gaze patterns in a completely unambiguous referential context, and can be used to evaluate whether children adopt a simple strategy of ignoring the objects located farther away from the active talker. Recall that that the instructions had the format "Look at the [noun].

Point to the [noun]." Our analyses focused on the processing of the noun in the first sentence only.

Prior to conducting our main analysis, we assessed how the presence of the barrier led children to generate predictions about the side of the object the speaker would refer to, within the initial portion of the unfolding utterance (i.e., "Look at the . . . ," 640 ms in duration on average). We calculated a difference score reflecting the tendency to fixate objects on the active speaker's side by subtracting the mean proportion of time within this interval spent fixating objects on the opposite side of the barrier from the speaker from the mean proportion time spent fixating objects on the speaker's side. Here, a positive score indicates more time spent fixating objects on the speaker's side (i.e., objects in common ground when the barrier was present), a negative score indicates more time spent looking at objects on the opposite side of the barrier (i.e., objects in privileged ground when the barrier was present), and a score of zero indicates equal consideration of objects on both sides of the barrier. We collapsed across the two types of trials in which barrier was present (i.e., shared-competitor and privileged-competitor trials) as no information about the intended object was available at this point in the utterance, making the trials similar. A preliminary paired-samples t test indicated that difference scores between these two conditions were not significantly different, $t(94) = 0.21$, $p = .833$, $d = 0.02$, as would be expected. Difference scores from the barrier-present trials were then compared to the no barrier trials. A paired-samples t test indicated that children were more likely to fixate objects on the speaker's side of the screen on trials in which the barrier was present ($M = .21$, $SD = .20$) than on trials in which the barrier was absent ($M = .14$, $SD = .25$), $t(94) = 2.11$, $p = .037$, $d = 0.22$ (small effect). Thus, prior to hearing the target noun, the presence of the barrier led children to consider objects in the common ground shared with the active speaker to a greater extent than objects in privileged ground. Although children tended to look toward the side of the speaker in the absence of the barrier, they did so less than when the barrier was present.

Our primary analysis focused on the time interval during which the noun was heard. This interval extended from the onset of the noun and continued for the average noun length (760 ms). A 200-ms margin was added to each boundary to account 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 noun interval to ensure that the eye gaze data reflected a reaction to the noun information rather than a continued tendency to fixate a particular side. This ensures the measures reflect the processing of the description independent of any anticipatory effects, which is the primary question in studies testing the real-time integration of perspective information with linguistic processing. Note that a core prediction of the egocentric account is that fixations to a privileged (i.e., nonshared) competitor occur upon hearing the description, despite any patterns of anticipation that would suggest perspective taking is occurring (e.g., Barr, 2008). Figure 2 presents fixation patterns for the target, competitor, and distractor objects (averaged across the two distractors) during the critical noun for each condition. Consistent with the pointing analyses, in the no barrier condition, the object coded as the competitor was a randomly selected nontarget item. Critically, children clearly show a reduced tendency to consider the competitor object in the privileged-competitor condition (middle panel) compared to the shared-competitor condition (top panel), even from the earliest moments. Consideration of the privileged competitor is, however, greater than when the item coded as the competitor bears no relationship to the noun being heard (no barrier condition, bottom panel). This is expected because the semantics of the target description (e.g., "shoe") provide a probabilistic constraint supporting the competitor as a potential referent (see Heller et al., 2016).

To capture children's gaze behaviors in a single statistical measure, we calculated a ratio score that reflected the relative tendency to fixate the competitor object relative to the target. This ratio score was calculated by dividing the mean proportion of time spent fixating the competitor by the total time spent fixating the target and competitor during the noun interval (see condition averages in Figure 3).

A one-way repeated measures ANOVA conducted on these values yielded a significant effect of condition $F(2, 188) = 18.20$, $p < .001$, $\eta_p^2 = .162$. A follow-up paired-samples t test, with alpha corrected for multiple comparisons to yield an adjusted value of .017, revealed that the time spent fixating the competitor was significantly higher in the shared-competitor condition ($M = .49$, $SD = .29$) than the privileged-competitor condition ($M = .28$, $SD = .28$), $t(94) = 5.35$, $p < .001$, $d = 0.55$, medium effect). Thus, as the noun unfolded, children spent an equal proportion of time fixating the competitor and the target in the shared-competitor condition (i.e., when both objects were mutually visible) and
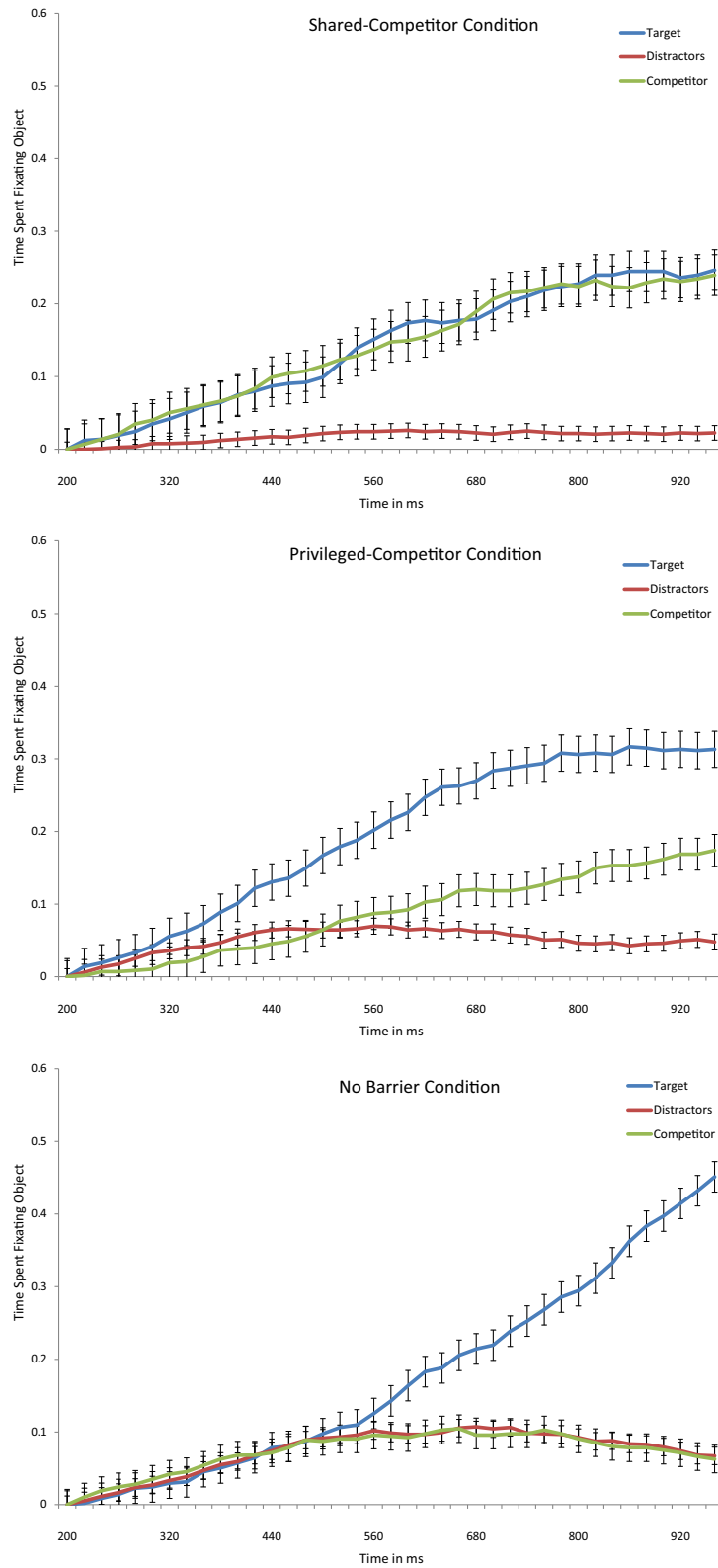
Figure 2. Fixation patterns for the target, competitor, and distractor objects (averaged across the two distractors) during the critical analysis period (i.e., 200–960 ms after the noun onset) for each condition.
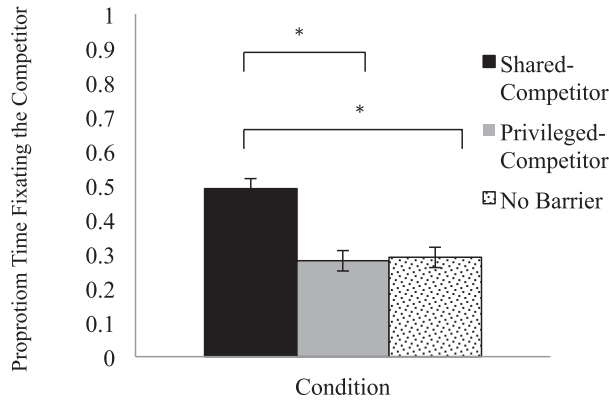
Figure 3. Proportion of time spent fixating the competitor, relative to the target, during the critical analysis period for the shared-competitor, privileged-competitor, and no barrier conditions. *$p \leq$ .001.

spent significantly less time fixating the competitor relative to the target in the privileged-competitor condition (i.e., when only the target was mutually visible). The proportion of time children fixated the competitor object in the privileged-competitor condition and the nontarget object in the no barrier condition (M = .29, SD = .27) was not significantly different, $t(94)$ = 0.33, p = .744, d = 0.03. Note that the similar proportions in this case reflect the mathematical equivalence of a comparatively stable target-competitor difference over the time interval (privileged-competitor condition) versus an initially small but eventually large difference over the same interval (no barrier condition). In contrast, children spent significantly more time fixating the competitor in the shared-competitor condition than the nontarget in the no barrier condition, $t(94)$ = 4.98, p < .001, d = 0.51 (medium effect). Note also that there was no early bias against the target (relative to other display objects) in the no barrier position, where the majority of trials involved reference to the object on the opposite side of the screen from where the speaker was seated. Thus, the effect observed in the privileged-competitor condition is

Table 1
Descriptive Statistics for Individual Difference Measures

	M	SD	Range
Representational skills	2.8	2.4	0–6
Executive function			
Inhibitory control	11.3	4.1	0–16
Working memory	11.1	2.6	6–16
Delay	3.4	1.9	0–6
Peabody Picture Vocabulary Test, 4th ed.	118.9	12.5	91–150

unlikely to reflect a heuristic or sustained expectation that active speakers would refer to objects on the side of the monitor closest to them. We do note that it is possible that the presence of the barrier and children's orientation toward the speaker's side of the barrier helped them to ignore the competitor in the privileged-ground condition.

Individual Differences and Performance on the Communication Task

Descriptive statistics for the individual difference measures are presented in Table 1. Inspection of the data indicated that the distribution of children's scores on the inhibitory control measure was moderately negatively skewed (i.e., 20% of children scoring 15 or 16 out of 16). To address this violation of the assumption of normality, we reflected the scores and applied a square root transformation. For ease of interpretation, untransformed scores are reported in the text. We first examined the relations among the individual difference measures. There was a moderate significant negative correlation between the square root transformed day–night task and receptive vocabulary, as measured by the PPVT, $r(92)$ = −.455, n = 95, p < .001. Note that this should be interpreted as a positive correlation because of the transformation described above. The other measures were not correlated with one another, ps > .073.

Our research questions focus on the relations between children's representational skills, executive function, and vocabulary skills and the degree of egocentricity they exhibited on the communication task. As such, we focused on the privileged-competitor condition, in which children had visual access to the competitor but the speaker did not. On these trials, a greater tendency to consider the competitor (i.e., visually or by pointing) corresponds to a more egocentric interpretation of the speaker's instructions (i.e., a lack of communicative perspective taking), whereas a lesser tendency corresponds to more effective perspective taking. We used the same pointing and eye gaze measures described above. Table 2 displays bivariate and correlations between the individual difference measures and the pointing and eye gaze measures.

There was a moderate negative correlation between the proportion of time spent fixating the competitor during the noun and scores on the Level 2 visual perspective-taking task, $r(93)$ = −.299, n = 95, p = .003, r^2 = .089. Thus, children with more developed cognitive perspective-taking skills were less likely to consider the member of the same-

Table 2

Bivariate Correlations (Pearson Correlation Coefficients) for All Individual Differences Measures

Measure	1	2	3	4	5	6	7
1. Pointing (PC)	—						
2. Eye gaze (PC)	**.31**	—					
3. PPVT	.07	.01	—				
4. Picture memory	.07	.06	.07	—			
5. Delay	−.09	.01	.18	−.04	—		
6. Day/night	.08	.03	−**.46**	−.12	−.14	—	
7. Turtle task	−.09	−**.30**	.19	−.07	.01	−.06	—

Note. Boldface type indicates $p < .05$. PC = privileged-competitor condition; PPVT = Peabody Picture Vocabulary Test.

category pair located outside the speaker's line of sight. This remained significant even when controlling for receptive vocabulary $r(92) = -.303$, $n = 95$, $p = .002$, $r^2 = .092$. Critically, children's visual perspective-taking skills were not related to time spent fixating competitor in the shared-competitor condition ($p = .922$) or time spent fixating the nontarget object in the no barrier condition ($p = .362$). In addition, none of the other individual difference measures significantly correlated with the tendency to consider the competitor $ps > .140$. There was also no relation between children's pointing and any of the individual difference measures, $ps > .366$. In sum, children's representational skills, which were unrelated to the general information processing and vocabulary measures, were the only skills that predicted children's implicit use of speaker-specific common ground.

Discussion

Our findings provide key insights into the development of communicative perspective taking during the preschool years. First, preschoolers consistently took speaker-specific common ground into account when making their explicit referential decisions. Specifically, when children were faced with two objects that matched the active speaker's referential description, they strongly preferred to point to the one in common ground over the referential competitor that was visible only to themselves and the inactive speaker. This is impressive given the young age of the listeners as well as the presence of multiple speakers. These findings are broadly consistent with previous research demonstrating that 3- to 4-year-olds will use their communicative partner's perspective to guide their interpretation of referential statements (Graham et al, 2014; Matthews et al., 2010; Nilsen & Graham, 2009). Our results extend

our current understanding by demonstrating that children flexibly adapt their behavior based on a speaker's perspective, even when the individual who is actively speaking is changing from one moment to the next.

As noted earlier, there is significant debate around whether perspective taking is evident in the earliest moments of processing referential descriptions. In the present study, preschoolers showed evidence of taking the active speaker's perspective into account during these early moments, despite being presented with instructions spoken by two different speakers. More specifically, as 4-year-olds heard a description that was technically ambiguous from their own perspective, they used the active speaker's perspective to identify the intended referent. These results add significantly to the literature by demonstrating real-time integration of others' visual perspectives at a younger age than previously demonstrated and in a task with more complex demands. This outcome is impressive, given that the linguistic, representational, and cognitive systems of 4-year-olds are routinely described as being in flux. Also impressive is the finding that preschoolers were able to rapidly integrate speaker-specific common ground when the active speaker was changing from one moment to the next. This outcome is in line with research demonstrating relatively early effects of common ground established through partner-specific entrainment on a shared term (Graham et al., 2014), as well as children's rapid integration of speaker-specific cues (e.g., Creel, 2012; Thacker, Chambers, & Graham, 2018) and is consistent with the evidence demonstrating adult listeners' sensitivity to the speakers' awareness of visual objects (e.g., Hanna et al., 2003; Heller et al., 2008; Mozuraitis et al., 2015).

In the present study, children pointed to the mutually visible competitor on 36% of shared-competitor (ambiguous) trials and only 9% of privileged-competitor (unambiguous) trials. These results are

similar to previous research using similar methods (e.g., Nilsen & Graham, 2009). For example, in Nadig and Sedivy's (2002) study, children selected the competitor first in 39% of trials when ambiguity was present. Our findings, however, contrast with studies that show late integration and difficulty with explicit choices (i.e., Epley et al., 2004, Fan et al., 2015, Wang, Ali, Frisson, & Apperly, 2016). What might account for the observed differences in egocentric errors and timing with regard to the use of common ground information? One possibility is that the communication task employed by Fan et al. (2015) and Epley et al. (2004) was more complex than that used in the present study—in these tasks, children were faced with three rather than two potential referents, and the competitor was a better match to the speaker's description (e.g., "pick up the small truck" when a medium truck and large truck were in common ground and the smallest truck was blocked from the speaker). In support of this assertion, Wang et al. (2016) demonstrated that increasing linguistic complexity resulted in a greater number of egocentric object selections (24% low vs. 53% high complexity) by children 8–10 years of age. However, preschoolers in our study had to flexibly represent the visual perspectives of two different speakers, which could also be said to result in added task complexity. Thus, rather than attributing differences in performance to the level of complexity or difficulty of the task, we propose that these differences can be explained using constraint-based approaches to this general question (e.g., Heller et al., 2016). As noted earlier, the materials in Epley et al. (2004) and Fan et al. (2015) require children have to overcome two things: their privileged (egocentric) perspective, plus the fact that the linguistic expression is a better fit for the privileged competitor than the mutually shared candidates. In our study, the linguistic fit was the same for target and competitor objects, thereby eliminating this influence of this additional factor.

Turning to the examination of the abilities that support communicative perspective taking, the results showed that children who performed better on the Level 2 visual perspective-taking task demonstrated less egocentricity in their online processing. How should this finding be interpreted? Given that preschoolers showed robust evidence of having acquired the form of visual perspective taking required for the communicative perspective-taking task (i.e., Level 1 perspective taking: recognizing that a partner can see X but not Y), the results cannot be understood as suggesting that some children were lacking adequate representational skills. Instead, the correlations are best understood as reflecting

differences in children's *level* of representational understanding, as captured by the Level 2 task (i.e., recognizing that people with different viewpoints may view the same object in different ways). We propose three possible explanations. First, it is possible that children with a more advanced representational understanding of others' mental states engage in perspective reasoning in a more efficient and automatic manner, facilitating the rapid integration of perspective information during referential processing. Alternatively, children with more advanced representational skills may not need to allocate as many cognitive resources to evaluating or internalizing the perspectives of others, allowing more resources to be devoted to the task of coordinating this information with incoming linguistic input. Third, it is possible that the need to monitor shared knowledge in the communicative interactions may itself bolster mentalizing skills (see Bohn & Köymen, 2018). Interestingly, in contrast to the pattern found with gaze measures, there was no relation between children's representational skills and their overt behavioral responses (points to objects). That is, having more advanced representational skills exerted a greater influence only on how quickly preschoolers integrated common ground information in the earliest moments and not their final referential decisions.

A notable finding in the current study was the absence of any detectable relation between children's executive functioning and implicit or explicit performance on the communication task. This stands in contrast to previous research showing a relation between inhibitory control and successful communicative perspective taking in children (Nilsen & Graham, 2009) and adults (Brown-Schmidt, 2009). One possible explanation is that the visual displays in the present study placed fewer demands on children's inhibitory control than those in the task of Nilsen and Graham (2009). However, given the demands of not only inhibiting one's own perspective but also having to shift between speakers, this explanation is not particularly compelling. We propose instead that a different measure of executive function, namely one that captures both inhibitory control and cognitive flexibility (mental shifting) such as the dimensional change card sort (e.g., Zelazo, 2006), might better capture the executive functioning demands involved in the need to shift between speaker perspectives. Another limitation of the current research was that it relied on single measures to assess each construct. It is possible that the relations between children's communicative perspective taking and executive functions may have been more adequately captured by a

multimethod approach in which each component had been assessed with several different tasks (e.g., Blankson et al., 2013). Furthermore, to investigate whether relations between different subcomponents of executive function (i.e., inhibitory control/cognitive flexibility/working memory) and successful communicative perspective taking vary based on representational demands, future research could directly compare the correlates of successful performance on communication tasks involving only one speaker (and correspondingly the common ground shared with a single individual) with cases involving multiple speakers and multiple common ground representations.

The type of referential communication task used in the current study also leaves open alternative interpretations about precisely when children generated a representation of a speaker's perspective. Specifically, given the presence of the barrier, it is possible that children formed a heuristic during the guessing game stage of the task (i.e., Side 1 is relevant for Speaker 1 and Side 2 is relevant for Speaker 2). This issue, however, is independent of the theoretical question at hand, which concerns 4-year-olds' ability to integrate the contents of these nonlinguistic representations during real-time referential interpretation. Furthermore, our measures are independent of visual anticipation that might arise from such a heuristic because our procedure for calculating fixation patterns during the critical time interval (i.e., the noun) included only newly programmed saccades. Nonetheless, one might argue that the potential to apply heuristics is responsible in part for children's strong performance. Future research could address this question by using communicative tasks that would prevent the development of such heuristics, such as alternating between *instructions* and *questions* across trials, given that the former requires language to be interpreted against knowledge and perspectives shared with the current speaker and the latter requires consideration of nonshared information (Brown-Schmidt et al., 2008). Yet, regardless of when children generated the speaker's perspective (i.e., in advance vs. during the utterance), our findings demonstrate that preschoolers can use others' perspectives in the earliest moments of processing referential descriptions in a context involving multiple speakers and where the relevant perspective changes from moment to moment.

In summary, the results of this study provide several key insights into preschoolers' online communicative perspective taking. First, preschoolers used distinct common ground representations for two different speakers to constrain referential interpretation within the earliest moments of processing. The robust and flexible real-time perspective-taking demonstrated by children whose linguistic, cognitive, and representational systems are still developing poses a challenge to theoretical accounts that assert that online perspective taking during communication is necessarily cognitively effortful or initially egocentric in the earliest moments of processing. Second, our findings demonstrate the importance of children's representational skills in supporting children's perspective taking during online spoken language processing. The findings also highlight young children's ability to successfully manage the informational demands associated with social interactions involving more than one communicative partner.

References

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

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. https://doi.org/10.1006/jmla.1997.2558

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40. https://doi.org/10.1016/j.cognition.2008.07.005

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., Khu, M., Graham, I., & Graham, S. A. (2013). ELIA: A software application for integrating spoken language and eye movements. *Behavior Research Methods*, *45*, 646–655. https://doi.org/10.3758/s13428-012-0302-1

Blankson, A. N., O'Brien, M., Leerkes, E. M., Marcovitch, S., Calkins, S. D., & Weaver, J. M. (2013). Developmental dynamics of emotion and cognition processes in preschoolers. *Child Development*, *84*(1), 346–360. https://doi.org/10.1111/j.1467-8624.2012.01841.x

Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives*, *12*, 104–108. https://doi.org/10.1111/cdep.12269

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

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*(1), 62–89. https://doi.org/10.1080/01690965.2010.543363

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*, 1122–1134. https://doi.org/10.1016/j.cognition.2007.11.005

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: Chicago University Press.

Creel, S. C. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development*, *83*, 2042–2056. https://doi.org/10.1111/j.1467-8624.2012.01816.x

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

Ferguson, H., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology*, *48*, 257–263. https://doi.org/10.1016/j.jesp.2011.08.007

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

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31–60. https://doi.org/10.1037/0033-2909.134.1.31

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

Graham, S. A., San Juan, V., & Khu, M. (2017). Words are not enough: how preschoolers' integration of perspective and emotion informs their referential understanding. *Journal of Child Language*, *44*, 500–526. https://doi.org/10.1017/S0305000916000519

Graham, S. A., Sedivy, J., & Khu, M. (2014). That's not what you said earlier: Preschoolers expect partners to be referentially consistent. *Journal of Child Language*, *41*, 34–50. https://doi.org/10.1017/S0305000912000530

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43–61. https://doi.org/10.1016/s0749-596x(03)00022-6

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*, 831–836. https://doi.org/10.1016/j.cognition.2008.04.008

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behaviour as the probabilistic weighting of multiple domains. *Cognition*, *149*, 104–120. https://doi.org/10.1016/j.cognition.2015.12.008

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211

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

Khu, M., Chambers, C., & Graham, S. A. (2018). When you're happy and i know it: Four-year-olds' emotional perspective taking during online language comprehension. *Child Development*, *89*(6), 2264–2281.

Köymen, B., Mammen, M., & Tomasello, M. (2016). Preschoolers use common ground in their justificatory reasoning with peers. *Developmental Psychology*, *52*, 423–429. https://doi.org/10.1037/dev0000089

Köymen, B., Schmerse, D., Lieven, E., & Tomasello, M. (2014). Young children create partner-specific referential pacts with peers. *Developmental Psychology*, *50*, 2334–2342. https://doi.org/10.1037/a0037837

Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, *56*, 436–455. https://doi.org/10.1016/j.jml.2006.05.002

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mind-blind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*, 551–556. https://doi.org/10.1016/j.jesp.2009.12.019

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*, 184–210. https://doi.org/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

Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, *46*, 749–760. https://doi.org/10.1037/a0019657

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology, 43,* 309–317. https://doi.org/10.1037/0012-1649.43.2.309

Mozuraitis, M., Chambers, C. G., & Daneman, M. (2015). Privileged versus shared knowledge about object identity in real-time referential processing. *Cognition, 142,* 148–165. https://doi.org/10.1016/j.cognition.2015.05.001

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

Nayer, S. L., & Graham, S. A. (2006). Children's communicative strategies in novel and familiar word situations. *First Language, 26,* 403–420. https://doi.org/10.1177/0142723706064834

Nilsen, E. S., & Bacso, S. A. (2017). Cognitive and behavioural predictors of adolescents' communicative perspective-taking and social relationships. *Journal of Adolescence, 56,* 52–63. https://doi.org/10.1016/j.adolescence.2017.01.004

Nilsen, E. S., & Fecica, A. M. (2011). A model of communicative perspective-taking for typical and atypical populations of children. *Developmental Review, 31*(1), 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*(2), 220–249. https://doi.org/10.1016/j.cogpsych.2008.07.002

Nilsen, E. S., & Graham, S. A. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development, 83,* 1400–1415. https://doi.org/10.1111/j.1467-8624.2012.01762.x

Nilsen, E. S., Graham, S. A., Smith, S. L., & Chambers, C. G. (2008). Preschoolers' sensitivity to referential ambiguity: Evidence for a dissociation between implicit understanding and explicit behavior. *Developmental Science, 11,* 556–562. https://doi.org/10.1111/j.1467-7687.2008.00701.x

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 spoken language comprehension? *Child Development Perspectives, 9,* 245–249. https://doi.org/10.1111/cdep.1214

Schuh, J. M., Eigsti, I. M., & Mirman, D. (2016). Discourse comprehension in autism spectrum disorder: Effects of working memory load and common ground. *Autism Research, 9,* 1340–1352. https://doi.org/10.1002/aur.1632

Smyth, R. (1995). Conceptual perspective taking and children's interpretation of pronouns in reported speech. *Journal of Child Language, 22,* 171–187. https://doi.org/10.1017/S0305000900009685

Stalnaker, R. (1975). Presuppositions. In B. Freed, D. J. Hockney, & W. L. Harper (Eds.), *Contemporary research in philosophical logic and linguistic semantics* (pp. 31–41). Dordrecht, the Netherlands: Springer.

Symeonidou, I., Dumontheil, I., Chow, W. Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology, 149,* 81–97. https://doi.org/10.1016/j.jecp.2015.11.007

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition, Vol. 11. Speech, language, and communication* (2nd ed., pp. 217–262). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012497770-9.50009-1

Thacker, J., Chambers, C., & Graham, S. A. (2018). When it is apt to adapt: Flexible reasoning guides children's use of talker identity and disfluency cues. *Journal of Experimental Child Psychology, 167,* 314–327. https://doi.org/10.1016/j.jecp.2017.11.008

Wang, J. J., Ali, M., Frisson, S., & Apperly, I. A. (2016). 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, 149,* 62–71. 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

Wardlow, L., & Heyman, G. D. (2016). The roles of feedback and working memory in children's reference production. *Journal of Experimental Child Psychology, 150,* 180–193. https://doi.org/10.1016/j.jecp.2016.05.016

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence™* (4th ed.). San Antonio, TX: The Psychological Corporation.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75,* 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1,* 297–301. https://doi.org/10.1038/nprot.2006.46

Supporting Information

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

Table S1. List of All Visual Stimuli (i.e., Target, Competitor, and Distractor Objects)

Table S2. Speaker, Target Object (Embedded in Utterance "Look at the _. Point to the _."), and Condition (SC = Shared Competitor; PC = Privileged Competitor; NB = No Barrier) for Each of the Four Experimental Orders