American Psychological Association

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

2020, Vol. 46, No. 10, 1922–1947
http://dx.doi.org/10.1037/xlm0000886

Referential Context and Executive Functioning Influence Children's Resolution of Syntactic Ambiguity

Zhenghan Qi
University of Illinois, Urbana–Champaign and University of Delaware

Jessica Love
University of Illinois, Urbana–Champaign and Northwestern University

Cynthia Fisher
University of Illinois, Urbana–Champaign

Sarah Brown-Schmidt
University of Illinois, Urbana–Champaign and Vanderbilt University

Classic studies reveal two striking differences between preschoolers and adults in online sentence comprehension. Adults (a) recruit referential context cues to guide syntactic parsing, interpreting an ambiguous phrase as a modifier if a modifier is needed to single out the intended referent among multiple options, and (b) use late-arriving information to recover from misinterpretation. Five-year-olds fail on both counts, appearing insensitive to the referential context and often failing to recover from parsing errors (Trueswell, Sekerina, Hill, & Logrip, 1999). But other findings suggest that 5-year-olds show *delayed* rather than absent sensitivity to the referential context, and that individual differences in executive functioning predict children's ability to recover from garden-path errors. In 2 experiments, we built on these findings, focusing on whether children recruit referential-context cues if given time to do so. Children heard temporarily ambiguous instructions (e.g., *Put the frog on the pond into the tent*), while we monitored their eye-gaze and actions. We used a slow speech rate, and manipulated referential context between rather than within subjects, to give children time to bring referential context cues into play. Across experiments, eye-movement and action analyses revealed emerging sensitivity to the referential context. Moreover, error rates and eye-movement patterns indicating failures to revise were predicted by individual differences in executive function (scores in Simon Says and Flanker tasks). These data suggest that children, like adults, use referential context information in syntactic processing under some circumstances; the findings are also consistent with a role for domain-general executive function in resolution of syntactic ambiguity.

Keywords: executive function, eye movements, kindergarten-path effect, referential context, visual world paradigm

Supplemental materials: http://dx.doi.org/10.1037/xlm0000886.supp

In adulthood, language comprehension is incremental and interactive (e.g., Altmann, 1998; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Gleitman, 2004; Trueswell & Tanenhaus, 1994). As each sentence unfolds, we integrate multiple knowledge sources to prepare for the most likely continuation, despite frequent ambiguity. This headlong processing style has clear benefits, allowing sentences to be perceived as meaningful before they are complete; but it also imposes costs. Sometimes listeners are led down the garden path, generating an interpretation based on partial input that turns out to be wrong; they must then reassess their initial (mis)interpretation in light of later-arriving disambiguating information. However, sometimes the interactive nature of the language processing system can reduce the likelihood of making garden-path errors, because a rich array of linguistic and contextual cues often lets us predict the right interpretation before disambiguating information becomes available.

This article was published Online First June 25, 2020.

Zhenghan Qi, Department of Psychology, University of Illinois, Urbana–Champaign, and Department of Linguistics and Cognitive Science, University of Delaware; Jessica Love, Department of Psychology, University of Illinois, Urbana–Champaign, and Kellogg School of Management, Northwestern University; Cynthia Fisher, Department of Psychology, University of Illinois, Urbana–Champaign; Sarah Brown-Schmidt, Department of Psychology, University of Illinois, Urbana–Champaign, and Department of Psychology and Human Development, Vanderbilt University.

The data and the analysis R codes are available on https://github.com/zhenghanQ/forestgit.

We thank all the participating families in this study. This research was supported by grants from the National Science Foundation (BCS1348522) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (HD054448).

Correspondence concerning this article should be addressed to Zhenghan Qi, Department of Linguistics and Cognitive Science, University of Delaware, 125 East Main Street, Newark, DE 19711. E-mail: zqi@udel.edu

For example, studies of the temporary syntactic ambiguity shown in (1a) illustrate both a garden-path effect and the integration of multiple knowledge sources in adults' online parsing. In (1a), the prepositional phrase *on the napkin* is temporarily ambiguous. It could modify the noun *frog* (specifying which frog to act on), or the verb *put*, specifying a destination for putting. If the visual context contains only one possible referent for *the frog*, adults at first prefer the destination interpretation (Crain & Steedman, 1985; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). For example, in a visual-world paradigm task in which adults act out such sentences while their eye movements are monitored, they look toward an empty napkin in the display when they hear *on the napkin*, implying that they are considering the destination interpretation. This initial interpretation must then be revised upon arrival of the disambiguating phrase *in the box*; as a result, correct interpretation of (1a) is delayed relative to the unambiguous control sentence (1b). This strong preference for a destination interpretation reflects the subcategorization bias of the verb *put*, which typically requires a destination phrase. However, adults also recruit top-down constraints derived from the referential context. In sentence (1a), if the visual context contains two frogs, adults are more likely to entertain the modifier interpretation—and thus less likely to look at the empty napkin—even before disambiguating information arrives (e.g., Novick, Thompson-Schill, & Trueswell, 2008; Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). In a two-frog referential context, the expression *the frog* is insufficient to pick out a unique frog to act on; adults use the contextual need for modification to guide interpretation of the ambiguous prepositional phrase. These and other findings provide strong evidence for adults' sensitivity to the *referential principle*: in cases of syntactic ambiguity, adults compute the referential consequences of multiple parses online, allowing them to select the parse that is best supported by the referential context (e.g., Crain & Steedman, 1985).

(1a) *Put the frog on the napkin in the box.*

(1b) *Put the frog that's on the napkin in the box.*

Preschoolers' comprehension of these garden-path sentences reveals both similarities and striking differences between preschoolers' and adults' parsing. Trueswell et al. (1999) first compared 5-year-olds' and adults' interpretations of sentences such as (1a) and (1b). They found that preschoolers, like adults, strongly preferred the destination interpretation when first encountering the ambiguous phrase *on the napkin* in a one-referent context. But children also showed two non-adult-like patterns. First, 5-year-olds often failed to use the disambiguating information at the end of the sentence to revise their initial interpretations. Adults almost never made errors in acting out sentences such as (1a) even in the one-referent context; instead, they used the disambiguating information to recover from the garden-path error. In contrast, young children made action errors on more than half of the ambiguous trials. For example, in response to (1a), many 5-year-olds placed a frog on an empty napkin, and then hopped it into the box. Second, 5-year-olds failed to use the presence of two frogs in the referential context to reduce their preference for a destination interpretation. They were equally likely to look at the empty napkin when hearing *on the napkin*, and to make action errors, whether there were two frogs or one in the display. Crucially, the same children successfully interpreted unambiguous control sentences such as (1b), suggesting that children's difficulty was due to the temporary ambiguity itself, not sheer sentence complexity.

These *kindergarten-path* effects (Trueswell et al., 1999) have been replicated many times (Anderson, Farmer, Goldstein, Schwade, & Spivey, 2011; Kidd, Stewart, & Serratrice, 2011; Snedeker & Trueswell, 2004; Weighall, 2008; Woodard, Pozzan, & Trueswell, 2016). Evidently, although preschool parsing is adult-like in many respects, 5-year-olds (a) struggle to integrate information from the referential context to help resolve syntactic ambiguities, and (b) are much less able to revise garden-path misinterpretations in response to later disambiguating information. Both of these abilities emerge gradually in older children (Trueswell et al., 1999; Weighall, 2008).

Why Do Children Fail to Use the Referential Context to Resolve Syntactic Ambiguity?

The young child's neglect of referential context in resolving syntactic ambiguity presents an interesting puzzle, because the use of the local context to infer a speaker's likely meaning has always been assumed to be a central mechanism in language acquisition (Gleitman, 1990; Pinker, 1984; Tomasello, 1995, 1999). For example, young children recruit one particular type of contextual information—knowledge about what objects the speaker can and cannot see—to resolve referential ambiguity. On hearing *Pick up the duck*, 6-year-olds (Nadig & Sedivy, 2002) and even 3- and 4-year-olds (Nilsen & Graham, 2009) largely ignored another duck in the display that was hidden from the speaker's view.

In principle it could be that 5-year-olds' insensitivity to the referential context manipulation in the classic frog on napkin experiments reflects ignorance of the pragmatic requirements of the definite determiner *the* (e.g., Deutsch & Pechmann, 1982). A definite determiner implies a referent that is uniquely identifiable in context; if children have not yet learned this requirement, they might not realize that *the frog* is insufficient in a two-frog context, and assume that any frog would do. However, preschoolers' language comprehension shows considerable sensitivity to the requirements of definite determiners (e.g., Cimpian, Meltzer, & Markman, 2011; Dale, Loftus, & Rathbun, 1978; van Hout, Harrigan, & de Villiers, 2010). For example, when following instructions such as *Find the orange*, 3- and 5-year-olds more often demanded clarification (e.g., *Which orange*?) in contexts with multiple referential candidates (multiple oranges) as opposed to only one (Morisseau, Davies, & Matthews, 2013; see also Nilsen, Graham, Smith, & Chambers, 2008).

Given that preschoolers have at least some knowledge of the pragmatic requirements of definite reference, we are left with our original puzzle: Why might children fail to use referential context information to guide syntactic ambiguity resolution? At least two possibilities, not mutually exclusive, have been proposed.

According to the *cue validity hypothesis* (e.g., Trueswell & Gleitman, 2004), young children learn to use cues to syntactic structure in an order that depends on the information-value of each cue in the input. For example, because the larger goal context sometimes determines in advance which referents are relevant to the ongoing action, speakers do not reliably produce modified noun phrases whenever there are two same-name referents in view (e.g., Brown-Schmidt & Tanenhaus, 2008; Engelhardt, Bailey, & Ferreira, 2006; Goudbeek & Krahmer, 2012). Thus the mere presence of two frogs in the visual context may only weakly predict a modifying phrase. In contrast, lexical items are quite reliable predictors of syntactic structure and meaning. For example, action verbs vary in how often they occur with a verb-attached prepositional phrase indicating an instrument of action, and both children and adults quickly recruit these verb biases to interpret ambiguous prepositional phrases (Snedeker & Trueswell, 2004). On the cue validity account, it should take longer for children to learn to rely on referential-context cues than on verb bias for syntactic disambiguation (e.g., Snedeker & Trueswell, 2004; Trueswell, Papafragou, & Choi, 2011). If so, then the 5-year-old's insensitivity to referential-context cues might reflect ignorance of their value, and we should expect children to benefit more from two-referent contexts as they gain language experience (Trueswell et al., 2011). Indeed, a replication of Trueswell et al. (1999) using mouse-tracking methods reported a link between children's vocabularies (as a proxy for prior language experience) and their tendency to make fewer action errors in the two-referent condition, suggesting that higher-vocabulary children had learned to expect a modifier in the presence of two objects with the same name (Anderson et al., 2011).

Alternatively, on the *top-down processing account* (Snedeker, 2013; Snedeker & Yuan, 2008), young children might possess the knowledge required to use top-down constraints such as the referential principle to guide parsing, but still often fail to bring those constraints to bear, especially when ranked against strong bottom-up constraints, simply because top-down processing takes time (Dell, 1986; see Dagerman, MacDonald, & Harm, 2006, for a similar argument regarding deficits in sensitivity to prior linguistic context in older adults). To illustrate, to use the referential principle in online parsing decisions, children must first identify syntactic options, compute their referential consequences—including the referential insufficiency of "the frog" in a two-frog scene—then let these consequences filter down to affect syntactic choices. From early childhood through adolescence, processing speed increases with age in a wide range of linguistic and nonlinguistic tasks (Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998; Kail, 1991; Kail & Salthouse, 1994). If children's slower processing poses a particular challenge for top-down processing, then we should expect that children often fail to use top-down information in time, especially where bottom-up constraints are powerful; however, children might show signs of referential context use when they have more time to do so.

Several existing results provide tentative evidence for this prediction; these provide the impetus for our study. First, and most centrally to the current study, Snedeker and Trueswell (2004) investigated the use of bottom-up lexical cues and top-down referential-context cues in 5-year-olds' and adults' syntactic processing. Participants listened to instructions with globally ambiguous *with* prepositional phrases, such as *Tickle the frog with the feather*, while looking at a display containing either one or two frogs (one holding a feather); the display also included a large feather that could serve as an instrument of action. Snedeker and Trueswell chose action verbs that varied in their likelihood of occurring with an instrument phrase. As noted earlier, both adults and children showed robust sensitivity to verb bias in this task. Adults' final actions, and their eye movements as sentences unfolded, also showed clear sensitivity to the number of candidate referents in the display. Children's final actions showed little influence of referential context, but analyses of their eye-gaze revealed sensitivity to the referential-context manipulation, with more consideration of the modifier interpretation in the two- than in the one-referent condition. However, this effect emerged only in a late analysis window, after the end of the sentence, suggesting that the recruitment of referential-context cues was slow in young children. The effect was also stronger in the first half of the study, perhaps because in this blocked design, children encountered trials in only one of the two referential-context conditions during the first half of the study.

A second set of results points toward a delayed effect of referential context in preschool parsing. In an experiment that explored the interpretation of sentences such as (1a) in two-referent contexts only, children performed more like adults if they were not allowed to look at the visual display until *after* hearing the entire critical sentence (Meroni & Crain, 2003), whereas children tested in the one-referent context in a separate experiment made the typical garden-path error despite not seeing the display until the sentence was complete (Meroni & Crain, 2011). Delayed access to the referential context may have prevented children from immediately committing to an interpretation of the ambiguous prepositional phrase (e.g., *on the napkin* in our example), thus granting them time to process the sentence-final disambiguating evidence (e.g., *into the box*) before they looked at the scene; this manipulation therefore may have lessened the cognitive demands of integrating top-down referential context and bottom-up lexical cues.[1]

Lastly, two studies on adjective comprehension show that young children use the referential principle to understand simpler sentences with no syntactic ambiguity. Huang and Snedeker (2013) explored 5-year-olds' online comprehension of sentences containing scalar adjectives such as *big* and *small* (e.g., *Point to the big coin*). Children, like adults, looked at the target referent more quickly in two-referent (e.g., with one big and one small coin) than in one-referent contexts; this suggests they expected scalar adjectives to be used in two-referent contexts, and used this knowledge to guide reference resolution. This referential-context effect was again smaller and later in children than in adults, suggesting that 5-year-olds were slow to integrate referential-context cues into online comprehension.[2] Another experiment examining color adjective processing (e.g., *Put the red butterfly in the paper bag*) similarly suggested delayed use of referential-context cues at age six (Sekerina & Trueswell, 2012).

In sum, the literature reviewed here suggests that 5- and 6-year-olds may have the linguistic and pragmatic knowledge required to use referential cues in online comprehension. However, the use of referential context cues at this age is fragile, often too little and too late, which makes it difficult to differentiate the cue-validity account apart from the top-down processing account. On a cue-

[1] It is worth noting that these data have not yet been successfully replicated. Weighall (2008) tested the same paradigm and did not observe the same referential context effect.

[2] Note that children presumably succeeded in this task both because the sentences contained no syntactic ambiguity and because scalar adjectives in particular are more reliably used when a modifier is needed in the referential context.

validity account we might take this pattern as evidence that referential context cues are just beginning to be identified as a parsing cue, and are not strongly weighted in the child's parser. On a top-down processing account, we might interpret the same pattern as evidence that referential-context cues, once they are available to the parser, are disadvantaged because they require slower top-down processing. In light of the hints of the fragile and late referential context effects in the literature, we aim to first reexamine the role of referential context in kindergarten syntactic parsing.

We reviewed evidence for referential context effects on online comprehension in late eye movements, both in the comprehension of adjective-noun phrases (Huang & Snedeker, 2013; Sekerina & Trueswell, 2012), and in the resolution of syntactic ambiguity of a prepositional phrase (e.g., Snedeker & Trueswell, 2004). The steps by which referential context information might have affected online comprehension are different in these two cases, but the delayed timing reported in these studies was similar. In the case of adjective-noun phrases, children retrieved the semantics and drew the pragmatic inference that an adjective implied a same-name competitor in the scene. In the case of the ambiguous postnominal prepositional phrases, children detected the referential insufficiency of *the frog* in a two-frog scene and inferred that a modifier was needed. We predict that a slower speech rate might give children more time to detect the referential ambiguity of *the frog* and therefore increase consideration of a modifier interpretation of the ambiguous phrase (e.g., *on the napkin*). Then, when children get the second prepositional phrase (e.g., *in the box*), they should be more likely to be able to confirm the modifier interpretation of the earlier phrase and assign the destination role for the second phrase.

A related possibility is that children are less likely to detect potential ambiguities spontaneously in the displays they are shown. For example, Sekerina and Trueswell (2012) reported that children were quicker to use a supportive referential context to guide interpretation of an adjective-noun phrase when the preceding *linguistic* context (rather than only the objects in view) highlighted the contrasting objects in the display. In a recent study of early language production, 3- to 5-year-olds' visual fixations suggested that they rarely monitored the display for potential ambiguity before speaking (e.g., two cars, one of which is the target to be named), though they monitored the adequacy of their own utterances as and after they spoke (e.g., looking more at the competitor object in a two- than in a one-referent display; Rabagliati & Robertson, 2017). The same may be true in the *frog on napkin* comprehension tasks. As adults, we quickly detect the potential for ambiguity in a two-frog display, and expect a modifier before the critical instruction begins; children might be slower or less likely to detect this potential ambiguity in advance. If slowness to incorporate top-down constraints is part of the reason why children fail to recruit the referential principle to resolve syntactic ambiguity, then slowing down the task should offer children a better chance to succeed. Similarly, if part of the problem is that children are less likely to spontaneously notice competing referential options as each display is revealed, then manipulating referential-context between subjects may make it quicker for children to detect the presence of two frogs (as in the first-block analyses of Snedeker & Trueswell, 2004, noted above), because the display either always contains two animals of the same kind, or never does.

In two experiments, we built on these prior findings to ask whether children can recruit referential-context cues in parsing if given time to do so. Children heard temporarily ambiguous instructions (e.g., *Put the frog on the pond into the tent*), while we monitored their eye movements and actions. We used a slow speech rate, and manipulated referential context between rather than within subjects, to give children time to bring referential-context cues into play.

Why Do Children Fail to Revise an Initial Misinterpretation?

The current literature offers a compelling explanation for the second striking deficit in preschool parsing, children's typical failure to revise garden-path misinterpretations (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Trueswell et al., 1999; Weighall, 2008; Woodard et al., 2016). Researchers have argued that revision during sentence comprehension recruits domain-general conflict resolution skills that allow us to inhibit a prepotent response to execute a subordinate response (e.g., Novick et al., 2008). Preschoolers' executive functioning skills are underdeveloped, and individually variable, as can be seen in their struggles with games such as Simon Says (Carlson, 2005; also see Davidson, Amso, Anderson, & Diamond, 2006). The maturation of the human prefrontal cortex, which is linked to core executive functions (Diamond & Doar, 1989), lags behind most of the other motor-sensory cortical areas (Chugani & Phelps, 1986; Huttenlocher & Dabholkar, 1997). Preschoolers' immature executive functioning may explain their failures to recover from garden-path misinterpretations. Evidence for this argument comes from studies of adults with focal lesions to the left prefrontal cortex. These patients showed deficits in executive functioning tasks and behaved similarly to 5-year-olds in comprehending garden-path sentences (Novick, Kan, Trueswell, & Thompson-Schill, 2009; see January, Trueswell, & Thompson-Schill, 2009, and Ye & Zhou, 2009, for converging evidence from neuroimaging data). Moreover, individual differences in conflict resolution in healthy adults predict successful ambiguity resolution in parsing tasks (Novick et al., 2008).

The relationship between executive function and language abilities is also attested in children. Individual differences in preschoolers' conflict resolution predicted their concurrent ability to take the speaker's visual perspective into account in resolving referential ambiguities (Nilsen & Graham, 2009) and their later ability to explicitly detect referential ambiguity (Nilsen & Graham, 2012). Khanna and Boland (2010) found that individual differences in older children's inhibitory control is related to their ability to use biasing linguistic context to resolve lexical ambiguity.

Recently, Woodard et al. (2016) directly explored the role of executive functioning in children's syntactic disambiguation. Five-year-olds who scored higher on tasks requiring cognitive flexibility (e.g., switching between congruent and incongruent trials in a Flanker task) were also more likely to succeed in understanding ambiguous *put* sentences (e.g., *Put the frog on the napkin onto the book*). However, mixed results have been reported regarding the relationship between different executive functioning measures and kindergarten parsing, possibly because of measurement issues in this population. As a result, there is little theoretical clarity on how domain-general executive functioning contributes to children's success in parsing temporarily ambiguous sentences. Therefore, it is important to pursue the relationship between cognitive control and sentence processing in children with new data sets. Because the syntactic parsing task used by Woodard et al. also required frequent switching between one- and two-referent trials, one might argue that the relationship between cognitive flexibility and syntactic ambiguity resolution was partly due to similar task demands in the parsing and the cognitive flexibility tasks. A secondary goal of the present research was therefore to seek further evidence of the role of executive function in children's syntactic disambiguation, where no switching is required between different types of referential context.

The Present Research

The logic of the present research is as follows: Our main goal was to test whether children would show sensitivity to referential context in online parsing if given time to do so. We aim to rule out a cue-validity-only hypothesis by providing children with what we estimate might give them a better chance to succeed in taking referential context into account. In two experiments, children moved animal cutouts on a felt board in response to instructions both with and without a temporary ambiguity, as in (2). The display (Figure 1a) included a target animal (e.g., a frog on a pond), a target destination (e.g., a tent), an incorrect destination (e.g., an empty pond) and a distractor animal. For children in the one-referent condition, the distractor animal was always a different animal (e.g., a duck on a leaf). For those in the two-referent condition, the distractor animal was of the same kind as the target animal (e.g., a frog on a leaf). We predicted that children in the two-referent condition would make fewer action errors and direct fewer fixations to the incorrect destination (e.g., the empty pond) compared with children in the one-referent condition.

(2a) *Ambiguous: Put the frog on the pond into the tent.*

(2b) *Unambiguous: Put the frog that's on the pond into the tent.*

Our secondary goal was to assess the role of executive functioning in syntactic disambiguation from an individual differences perspective. To this end, we included an executive functioning task in each experiment. If immature executive functioning influences children's ability to resolve syntactic ambiguity, then children who perform poorly on measures of executive function should show greater difficulty in resolving ambiguity compared with children who perform well on measures of executive function. Language ability is included as an additional individual measure that enables us to assess the unique role of executive functioning in syntactic parsing. We use language as the secondary measure based on prior work indicating that vocabulary is related to sentence processing. For example, vocabulary size in 5-year-olds predicted fewer errors in resolving syntactic ambiguity in a two-referent condition similar to the one tested here (Anderson et al., 2011). At the same time, children with better vocabulary might also process sentences faster (Fernald, Perfors, & Marchman, 2006; Fernald, Marchman, & Hurtado, 2008) and therefore have more time to integrate the referential cues. There is evidence that slowing down the sentence presentation rate could swamp the effect of vocabulary on children's performance in passive sentence comprehension (Messenger & Fisher, 2018). Taking these findings together, we predicted that when speech rates were slow, children's executive functioning ability would positively correlate with successful syntactic disam-

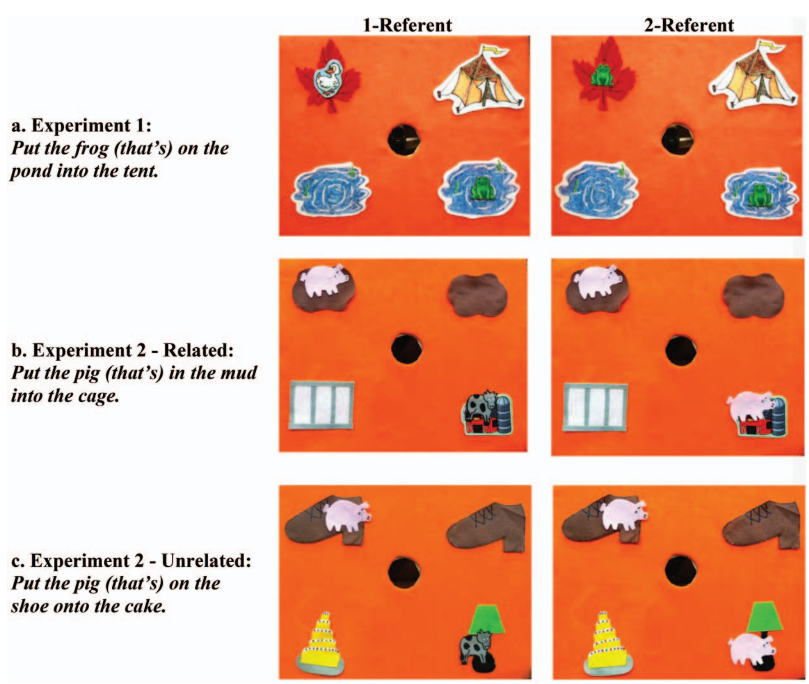

Figure 1. Examples of one-referent (left) and two-referent (right) displays for the target sentences in Experiment 1 (a) Experiment 2 (b and c). See the online article for the color version of this figure.

biguation of the *Put* sentences, above and beyond contributions from individual differences in language ability.

Our experimental design was similar to that of Trueswell et al. (1999), with the following changes: (a) All spoken instructions were prerecorded in slow, child-directed speech to accommodate children's slower processing speed. (b) Referential context (one vs. two referents) was manipulated between rather than within subjects, because switching between referential contexts may be cognitively demanding (even when implemented in a blocked design, cf. Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008), and switching may also reduce the reliability of contextual cues within the local task (Ryskin, Brown-Schmidt, Canseco-Gonzalez, Yiu, & Nguyen, 2014; Weighall, 2008). We anticipated that these changes would allow for a stronger test of whether 5-year-old children, in a supportive test paradigm, can use referential-context cues to resolve a temporary syntactic ambiguity. To preview our results, 5-year-olds showed sensitivity to the referential context in both experiments, with the clearest effects emerging in the online eye-gaze data in Experiment 1, and in the offline action data in Experiment 2. Moreover, children who performed better in the executive functioning tasks (Experiment 1: Simon Says; Experiment 2: Flanker task) showed less difficulty in resolving syntactic ambiguity.

Experiment 1

Method

Participants. Twenty-four 5-year-olds (12 boys, 12 girls; mean age = 5;0, range 4;8 to 5;4), all native English speakers from the Urbana-Champaign, IL region, participated. Twelve children were randomly assigned to the one-referent and 12 to the two-referent condition. Data from seven additional children were eliminated because of experimenter error (three), high error rates in unambiguous trials (two; see Data Coding below), or failure to follow instructions (two children looked fixedly at the camera as the sentences played). This study was approved by the Institutional Review Board at the Office of the Vice Chancellor for Research at the University of Illinois, Urbana-Champaign.

Materials. The target sentences all contained the verb *put* and appeared in one of two versions: Ambiguous sentences (2a) included a temporarily ambiguous prepositional phrase (e.g., *on the pond*) that could initially be interpreted either as a destination for the verb *put* or as a modifier of the direct object noun (e.g., *frog*). Unambiguous sentences (2b) included the phrase *that's*, and thus only permitted the modifier interpretation. The animals, locations, and destinations in the sentences were chosen to be thematically related (e.g., the frog, duck, pond, leaf, and tent of Figure 1a are all appropriate to an outdoor scene); this was done so that the same sentence materials could be embedded in a story context in a planned future experiment (see Appendix A for a full list of materials). Each child completed 12 critical trials (six ambiguous, six unambiguous), randomly interspersed among six filler trials. Each trial comprised two instructions. The first instruction of each critical trial was the target sentence; the second was a filler sentence (e.g., *Put the frog on the pond into the tent. Now put both animals on the leaf*). In filler trials, both instructions were filler sentences (e.g., *Put one goat beside the rainbow. Now put the other goat on the tree*).

The target sentence in each critical trial was either an ambiguous or an unambiguous version of the target item set. All stimuli were prerecorded by a female native English speaker in an animated child-directed style, at a slow speech rate. The average utterance length was 3.26 s (range 3.00–3.75 s) for ambiguous target sentences and 3.44 s (range 3.08–3.66 s) for unambiguous target sentences. In our stimuli the temporarily ambiguous location noun (e.g., *pond*) began on average 1.26 s (range 1.11–1.50 s) after the onset of *Put,* the disambiguating prepositional word *into* began 1.92 s (range 1.62–2.22 s) after *Put*, and the ultimate destination noun (e.g., *tent*) began 2.56 s (range 2.35–2.91 s) after *Put* in ambiguous target sentences (Supplemental Figure 1 in the online supplemental materials). In contrast, the location and destination nouns in the ambiguous sentences used by Trueswell et al. (1999; as estimated from plots of visual fixations over time) began approximately 1 s and 1.7 s after the onset of *Put*. We reasoned that this slower speech rate might give children the time needed to retrieve a modifier interpretation as the sentence unfolded.

In addition to the overall slow speech rate, a post hoc analysis of the prosody of our recorded stimuli suggested that our recordings may have helped children retrieve the modifier interpretation. Kraljic and Brennan (2005) analyzed similar temporary ambiguities in adult-to-adult speech produced in a referential communication task, and found speakers reliably produced prosodic boundary cues to disambiguate prepositional phrase attachment. In their recordings, the direct object noun in sentences like ours (e.g., *frog*) was 10.45% shorter on average than the prepositional object (*pond*) if the ambiguous prepositional phrase was a modifier, but was 17.3% longer than the prepositional object if the prepositional phrase was a destination. We calculated Kraljic and Brennan's prosodic boundary difference score for our materials, and found that the target sentences were somewhat biased toward the (correct) modifier interpretation, with the direct object noun (*frog*) on average 7.4% shorter than the prepositional object noun (*pond*), as a proportion of sentence length.

Each filler sentence also started with the verb *put* but had no prepositional-phrase modifiers, and used a variety of phrases to indicate destinations for putting (e.g., *on top of the house; near the flowers; on the scarecrow's head*). The direct objects of *put* in filler sentences were almost all definite noun phrases, varying in form (e.g., *the mouse, the green chicken*, *the little kitty, the other monkey*, *one goat*, *both animals*, *them*, or *him*).

Referential context (one-referent vs. two-referent) was manipulated between subjects. Ambiguity (ambiguous, unambiguous) was manipulated within subjects. The visual displays for the filler trials were similar to those in the critical trials. The one- versus two-referent manipulation was carried through the filler trials: Two animals of the same kind (e.g., two goats) appeared in all filler-trial displays in the two-referent condition, and animals of different kinds (e.g., one goat and a fox) appeared in all filler-trial displays in the one-referent condition.

Apparatus and procedure.

Comprehension task. Children sat in front of an inclined felt-board with a central hole for a camera. The central camera recorded children's eye movements as the sentences unfolded; a second camera positioned behind the children recorded their actions. Audio stimuli were played from an mp3 player connected to small external speakers mounted on shelves at the top of the felt-board. All the displayed objects were two-dimensional felt cutouts, rather than three-dimensional toys.

One experimenter stood to the child's left and placed objects on the board; a second experimenter sat to the child's right and monitored the video display of the child's face during the task. Children were told that they would play a game in which they followed instructions. At the start of each trial, the standing experimenter placed the felt objects on each quadrant of the board, introducing each by name as she did so. The platforms and animals were introduced as separate objects using unmodified nouns (e.g., "Here's a tent, a leaf, a duck, a pond, another pond, and a frog"). The experimenter named the objects a second time in the same order, then stepped back out of the child's line of sight and played the prerecorded sentences for that trial. Each trial began with a prerecorded instruction to "Look at the camera," followed by two instructions to move animals on the board. Upon completion of each trial, the second experimenter gave children a sticker while providing verbal feedback (e.g., "good!", "okay!"), regardless of accuracy.

Individual difference measures. Following the comprehension task, each child played a Simon Says game with the experimenter as a measure of inhibitory control (e.g., Carlson, 2005; Strommen, 1973). We chose Simon Says by analogy to the inhibitory control task used for 3- to 4-year-olds by Nilsen and Graham (2009). Nilsen and Graham found that children's performance in a Dog/Dragon task similar to Simon Says predicted their ability to overcome egocentric calculations about the referential context. Given a prior report that most 5-year-olds were at ceiling in the Dog/Dragon task, but that fewer than half of 5-year-olds reached 80% accuracy in Simon Says (Carlson, 2005), we chose Simon Says to measure children's inhibitory control. We followed the procedure described by Carlson (2005): The experimenter and child stood facing each other, and the experimenter explained that she would ask the child to do actions along with her (e.g., "Touch your head"), but that the child should imitate her actions only if she first said "Simon says" (imitation trials); otherwise, the child should remain still (inhibition trials). The experimenter practiced with the child until the child succeeded in one imitation and one inhibition trial, and then gave 10 critical commands (five with and five without "Simon says" randomly intermixed). Because children made very few errors in the imitation trials (six of 120 trials), we used children's success rate in the inhibition trials as our measure of executive function. Children's Simon-Says scores were distributed bimodally: 37.5% of children failed (to inhibit their response) in all five trials, whereas 41.7% of the participants made at most one error. Because of this bimodal distribution, we calculated nonparametric Spearman correlations for all the correlation analyses reported below involving Simon Says accuracy.

In addition to this measure of inhibitory control, measures of language ability were obtained through the Speech and Language Assessment Scale (SLAS; Hadley & Rice, 1993), which was completed by parents while the child was engaged in the main experiment. The SLAS is a 19-item checklist asking parents to rate their child's language abilities relative to other children his or her age on a 7-point scale. We chose this parental checklist measure rather than a direct assessment of the child's vocabulary (e.g., Peabody Picture Vocabulary Test [PPVT]) to avoid lengthening the behavioral testing for the children. SLAS scores correlate reasonably highly with vocabulary development, sentence production and comprehension skill (Hadley & Rice, 1993; Kooijman, Junge, Johnson, Hagoort, & Cutler, 2013; Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006).

Children's Simon-Says and SLAS scores were not significantly correlated with each other ($rho = .21$, $p = .16$, one-tailed), and Simon Says scores were not correlated with age ($rho = -0.002$, $p = .50$, one-tailed). Note that the SLAS asked parents to rate their child's language skills relative to other children his or her age; accordingly, scores on this measure also did not vary with age ($rho = 0.16$, $p = .23$, one-tailed). At least within this fairly narrow age range, the Simon-Says measure of inhibitory control was roughly independent of language skills and age.

Children assigned to the one-referent and two-referent conditions did not differ significantly in age ($M \pm SD$ = 4;11 ± 0;2 vs. 5;0 ± 0;3, $t = -0.86$), gender ratio (six boys and six girls in each condition), Simon-Says success rate ($M = 0.47$ vs. 0.45, Wilcoxon's z score = 0.20), or SLAS scores ($M = 5.19$ vs. 5.14, $t = 0.18$, all $ps > 0.40$).

Data coding. Children's actions and eye-movements in the sentence comprehension task were coded separately from the two video records described above. Children's actions were coded as correct if they picked up the target animal and moved it directly to the target destination; all other responses were coded as incorrect (see Table 1). Two participants' data were not included (as noted in the Participants section) because they made four or more errors on the six unambiguous trials, suggesting confusion about the task. Children's visual fixations were coded frame-by-frame by trained research assistants who were blind to condition, starting at target-sentence onset and ending 4.5 s later, at which point children had initiated their actions in 96% of the trials. We coded fixations to each quadrant of the board, to the camera, and away from the board. Frames were coded as missing if the child's eyes were out of camera view. Trials with fewer than one-third of the frames coded as fixations to a quadrant of the board were excluded from eye-movement analyses (22 of 288 possible trials; 7.6%). A second coder coded the eye-movements of six randomly selected children (25% of the data), and the two coders agreed on 96% of coded video frames.

Data and code sharing. The data and the analysis R codes are available on https://github.com/zhenghanQ/forestgit (Qi, 2018).

Results and Discussion

Analysis of children's actions. Table 1 shows the distribution of correct actions and errors across conditions; Figure 2a shows the proportion of critical trials with an action error. The included children made almost no errors on unambiguous trials. By contrast, on ambiguous trials children made errors both in the one- and two-referent context conditions. As in previous studies using this task, the nature of the errors (see Table 1) suggests that children initially interpreted the temporarily ambiguous prepositional phrase as a destination, and sometimes failed to revise this interpretation following the disambiguating information at the end of the sentence. For example, the most common error was a *hopping error*, in which children moved an animal first to the incorrect destination and then to the correct destination. This type of error may reflect an attempt to incorporate both prepositional phrases (as successive destinations) into the interpretation of the utterance,

Table 1
Distribution of Correct Actions and Action Errors in the One-Referent and Two-Referent Conditions, Experiment 1

		One-referent condition (n = 12)		Two-referent condition (n = 12)	
Action category	Animals moved	Ambiguous	Unambiguous	Ambiguous	Unambiguous
Correct	Target	47	72	55	71
	Distracter	—	—	—	—
Hopping	Target	22	0	2	0
	Distracter	0	0	1	0
Falling short	Target	2	0	0	0
	Distracter	1	0	3	0
One of each	Target	0	0	0	0
	Distracter	0	0	5	1
	Both	0	0	6	0
Total errors		25	0	17	1
Total trials		72	72	72	72

Note. Hopping: Children moved either animal to the incorrect destination before setting it down at the correct destination. Falling short: Children moved either animal to the incorrect destination and stopped there. One of each: Children moved one animal to the correct destination and the other to the incorrect destination, or moved only the distractor animal to the target destination.

without deriving a modifier interpretation of the first prepositional phrase.

The proportions of action errors for each child were transformed using an empirical-logit function,[3] and analyzed using the lmer function (Bates, Maechler, Bolker, & Walker, 2015) in R (Version 3.3.2; R Development Core Team). The model included fixed effects for our key experimental factors—Ambiguity, Referential Context, and their interaction—each of which was coded using mean-centered contrast coding. The model included random by-subject intercepts as well as random by-subject slopes for ambiguity condition (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). The model, shown in Table 2, revealed significant main effects of Ambiguity (more errors in ambiguous trials), but no main effect or interactions involving Referential Context.

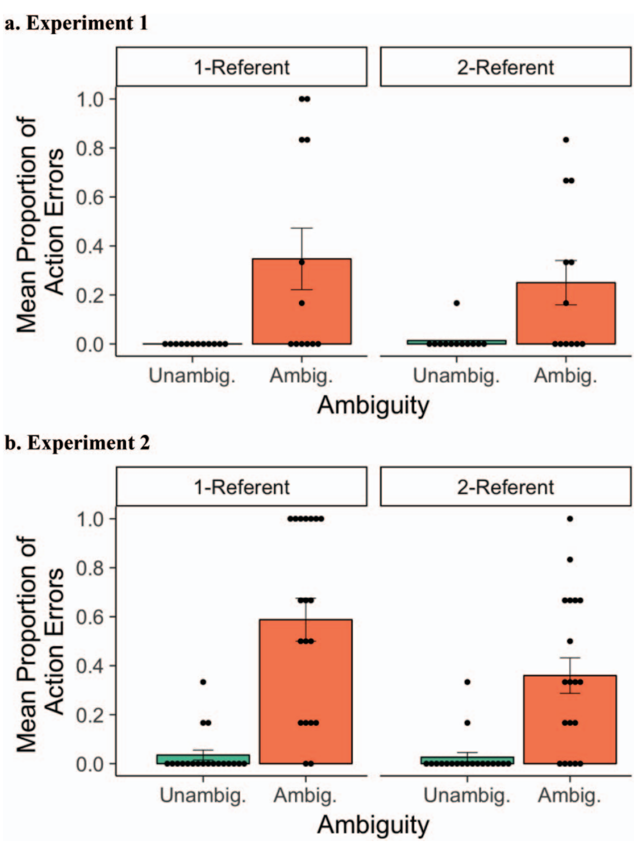

Figure 2. Mean proportion of trials with an action error, plotted separately by ambiguity (manipulated within subjects) and one- versus two-referent context (manipulated between subjects). Error bars reflect ± 1 *SE*. Ambig. = ambiguous; Unambig. = unambiguous. Individual data points represent each child's action error rate in the ambiguous and the unambiguous condition. See the online article for the color version of this figure.

Because we saw no significant effect of referential context on the likelihood of action errors, the action data did not support the hypothesis that the two-referent context helped children to infer the need for a modifier. Nevertheless, it is noteworthy that children in the two-referent condition made numerically fewer errors than those in the one-referent condition. This numerical trend is typical for adults, but rare for children (Novick et al., 2008; Trueswell et al., 1999). This more adult-like trend in our data was accompanied by a relatively low overall error rate: Children made action errors in 29% of ambiguous trials, markedly fewer than in the previous studies, in which error rates were well above 50%. This suggests that our methodological changes (slow speech rate and between-subjects manipulation of referential context) supported children's ability to generate a modifier interpretation of the temporarily ambiguous prepositional phrase.

Additional evidence that children in our task often considered a modifier interpretation of ambiguous prepositional phrases can be seen in the animals they chose to move in the two-referent condition (see Table 1). A preference to select the target (e.g., frog on pond) over the distracter (e.g., frog on leaf) indicates a (nascent, at least) modifier interpretation of the temporarily ambiguous phrase, *frog on the pond*. In ambiguous trials in the two-referent condition,

[3] We analyzed by-subject transformed error proportions rather than by-trial binomial error data because children made almost no errors in unambiguous trials, resulting in too many zero values for the binomial model to converge. The empirical logit is advantageous because, unlike a proportional scale, it is not bounded at 0 and 1 (see Barr, 2008 for a similar approach).

Table 2
Estimated Parameters for Mixed-Effects Models Predicting the Proportion of Trials With an Action Error (e-logit Transformed), Experiment 1

Predictors	Estimate	*SE*	*t* value	*p*
(Intercept)	−1.78	0.20	−9.06	7×10^{-9}
Ambiguity (amb. vs. unamb)	**1.47**	**0.37**	**4.00**	**.0006***
Referential context (2-ref vs. 1-ref)	−0.19	0.39	−0.49	.63
Referential Context × Ambiguity	−0.59	0.74	−0.80	.43

Note. Akaike information criterion (AIC) = 173.00. Model fit was based on 48 data points (24 subjects). Statistically significant main effects are listed in bold.
* $p < .05$.

children typically chose to act only on the target animal (as opposed to the distractor animal or both animals) in 78% of trials (*SE* = 8%), significantly higher than a chance level of 50% (Wilcoxon's *z* score = 2.48, $p = .006$). In contrast, 5-year-olds in previous studies were about equally likely to act on the target and distracter animals (Target animal selection: Trueswell et al., 1999: 65%; Hurewitz et al., 2000: 57%).

Analysis of children's eye-movements. Whereas the analysis of action data speaks to the children's ultimate interpretation of the temporarily ambiguous sentences, we next examined eye-movements that children made as a measure of online interpretation of the sentences as they unfolded in time. We ask whether referential context facilitates syntactic revision after children hear the second prepositional phrase (e.g., *into the tent*). The preposition *into* provides strong disambiguating evidence that the destination interpretation for the first prepositional phrase (e.g., *on the pond*) was not right. If the two-referent context facilitates syntactic revision, we expect to see fewer fixations to the incorrect destination compared with the one-referent context during this disambiguation time window. Looks to the correct destination also reflect how efficiently children recovered from a garden-path. Therefore, we report analyses of two main dependent measures: children's tendency to look at the incorrect destination (e.g., the empty pond) and their tendency to look at the correct destination (e.g., the tent) in response to the disambiguating prepositional phrase (*into the tent*). Note that eye movements are analyzed in time-regions linked to the onset of the disambiguating prepositional phrase in each sentence, and are offset by 200 ms to account for the time needed to program and launch an eye movement (Hallett, 1986) in response to the critical word.

Figure 3a plots the group mean proportion of fixations to the incorrect destination, following the onset of the disambiguating second prepositional phrase (e.g., *into the tent*), offset by 200 ms in a planned, 1.5-s analysis window, which is the average duration between the onset of *into* and the end of the utterance. We examined at each time point (each coded video frame) whether or not the child looked at the incorrect destination. We used autoregressive generalized linear mixed effects models (GLMM) to handle the strong dependency between neighboring time points (i.e., the temporal lag effects) in intensive binary time series data (Cho, Brown-Schmidt, & Lee, 2018). The first 33-ms portion was treated as the baseline to provide the lagged response at the beginning of the analysis window. Both of the experimental factors, Ambiguity and Referential context, were coded using mean-centered contrast coding. The lag 1 response reflecting the binary fixation at the previous time point was also entered as one of the fixed effects. The resulting autoregressive GLMMs used by-subject random intercepts and random slopes for ambiguity and lag 1 response, as well as by-item random intercepts and random slopes for ambiguity, referential context and lag 1 response. The analyses of looks to the correct destination followed the same procedure. Table 3 lists the parameters of the models for incorrect and correct destination looks in the analysis window.

Our analysis showed a marginal Ambiguity effect—showing more looks to the incorrect destination in ambiguous than unambiguous trials—that significantly interacted with Referential Context (Figure 4a). Children looked significantly longer at the incorrect destination in ambiguous than in unambiguous trials in the one-referent ($b = 0.77$, $SE = 0.24$, $z = 3.23$, $p = .001$) but not in the two-referent condition ($b = -0.19$, $SE = 0.24$, $z = -0.80$, $p = .42$). One interpretation of this interaction is that the two-referent

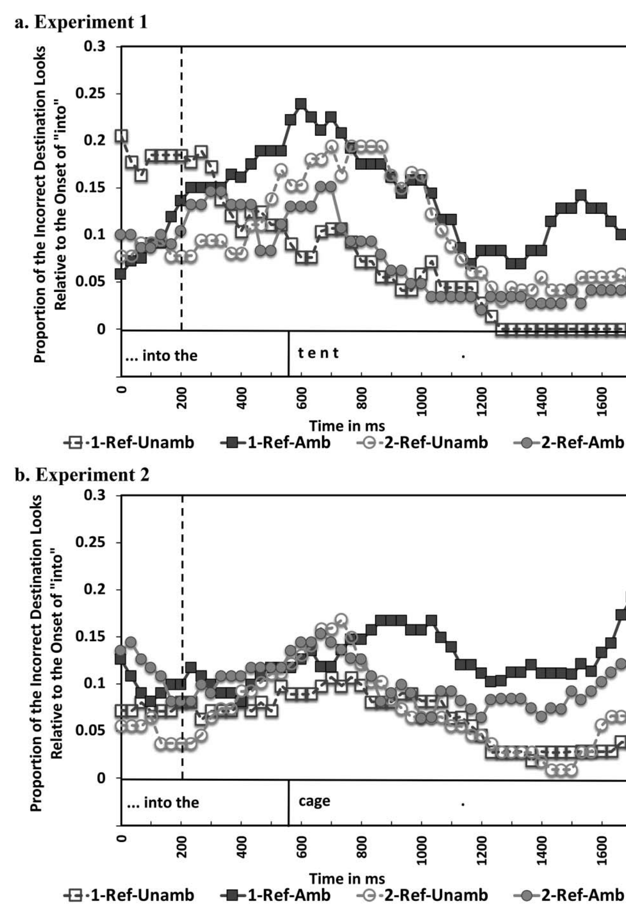

Figure 3. Proportion of fixations to the incorrect destination (e.g., the empty pond in Experiment 1 or the empty mud in Experiment 2) relative to the onset of the second prepositional phrase (e.g., *into the . . .*). The analysis window was offset by 200 ms after the onset of *into* (see text). Open symbols represent fixations in Unambiguous trials, and filled symbols in Ambiguous trials; square symbols represent fixations in the one-referent condition, and circles represent fixations in the two-referent condition.

Table 3
Summary of Fixed Effects in Autoregressive Generalized Linear Mixed-Effect Model Analyses of the Proportion of Looks to the Incorrect Destination (e.g., the Empty Pond) and the Correct Destination (e.g., the Tent) After the Onset of the Disambiguating Prepositional Phrase (e.g. Into the Tent), Experiment 1

Dependent variable	Predictor	Estimate	*SE*	*z* value	*p* value
Incorrect destination looks (AIC = 1727.6)	(Intercept)	−5.18	0.15	−34.30	$<10^{-16}$
	Lag 1 response	7.59	0.19	40.66	$<10^{-16}$
	Ambiguity (amb. vs. unamb)	0.30	0.17	1.79	.07†
	Referential context (2-ref. vs. 1-ref)	0.08	0.17	0.47	.64
	Ambiguity × Referential Context	**−0.89**	**0.31**	**−2.85**	**.004***
Correct destination looks (AIC = 3064.8)	(Intercept)	−4.00	0.14	−27.86	$<10^{-16}$
	Lag 1 response	7.31	0.14	52.87	$<10^{-16}$
	Ambiguity (amb. vs. unamb)	**−0.35**	**0.12**	**−2.90**	**.004***
	Referential context (2-ref. vs. 1-ref)	−0.003	0.17	−0.02	.99
	Ambiguity × Referential Context	0.11	0.23	0.46	.65

Note. AIC = Akaike information criterion. Window duration: 200–1,667 ms after the onset of *into*. Statistically significant main effects are listed in bold.
† $p < .1$. * $p < .05$.

context prompted children to consider a modifier interpretation of the ambiguous prepositional phrase, similar to the late effect of referential context effect found in Snedeker and Trueswell (2004). As a result, late in the trial as they prepared to act, children were less likely to consider the incorrect destination, the empty pond, in the two-referent context.

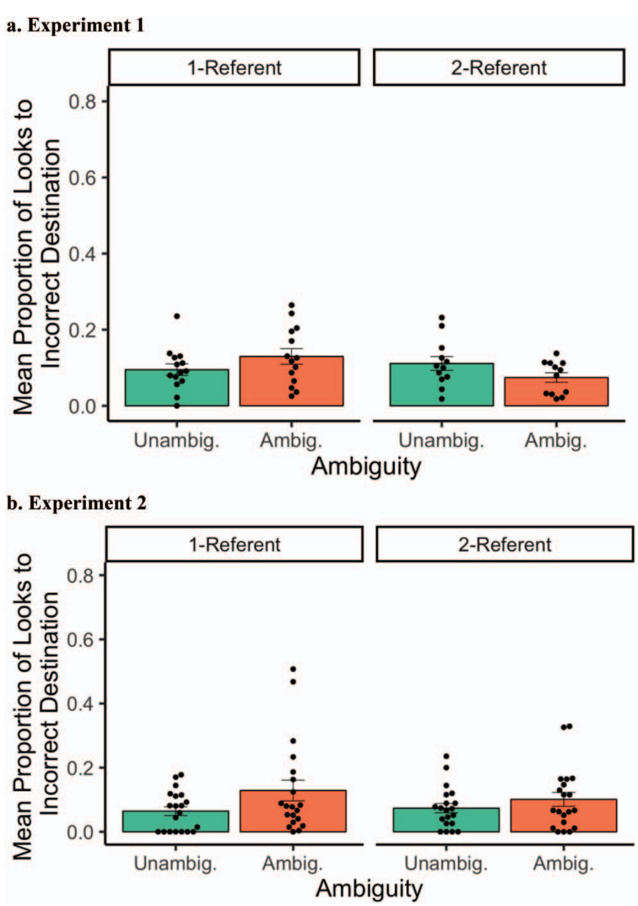

Figure 4. Mean proportion of looks to the incorrect destination after the onset of *into*. Error bars reflect ± 1 *SE*. Individual data points represent each child's proportion of fixations to the incorrect destination in the ambiguous and the unambiguous condition. See the online article for the color version of this figure.

An alternative explanation for the effect of referential context on fixations to the incorrect destination has been raised by Novick et al. (2008). Consider that in the two-referent context, if the ambiguous prepositional phrase is *not* interpreted as a modifier, the direct object noun phrase (*the frog*) is ambiguous. This ambiguity might cause children to keep alternating between the two animals (both frogs), precluding looks to the incorrect destination. However, children spent about the same amount of time looking at the animals during this window regardless of whether they were in the one-referent (M = 0.68; se = 0.03) or the two-referent condition (M = 0.61; se = 0.06). This suggests that the significant interaction of referential context with ambiguity was not attributable to the masking of incorrect-destination fixations by additional looks to animals. Instead, the two-referent context apparently helped children to compute a modifier interpretation of the ambiguous prepositional phrase.

The lack of an ambiguity effect for incorrect destination fixations within the two-referent condition is somewhat surprising, given that referential context is only one of many partial constraints on interpretation, and the verb *Put* is strongly predictive of a locative prepositional phrase argument. Following Novick et al. (2008), we explored the possibility that looks to the *correct destination* might offer a vehicle for examining lingering effects of ambiguity in the two-referent context. In an ambiguous sentence, the second preposition (e.g., *into*) was the onset of the disambiguating phrase. The time children spent considering the correct destination after *into* until the end of the sentence should reflect the efficiency of garden-path recovery, that is, how much they shifted away from the initial erroneous interpretation. As shown in Figure 5a, children's looks to the correct destination (e.g., the tent) increased gradually after the onset of the disambiguating prepositional phrase (*into the tent*). This increase in fixations to the correct destination was slower in ambiguous than in unambiguous trials, suggesting that the garden-path interpretation of the first prepositional phrase as a destination interfered with the interpretation of the second prepositional phrase. This analysis of this measure (bottom section of Table 3) revealed a main effect of Ambiguity,

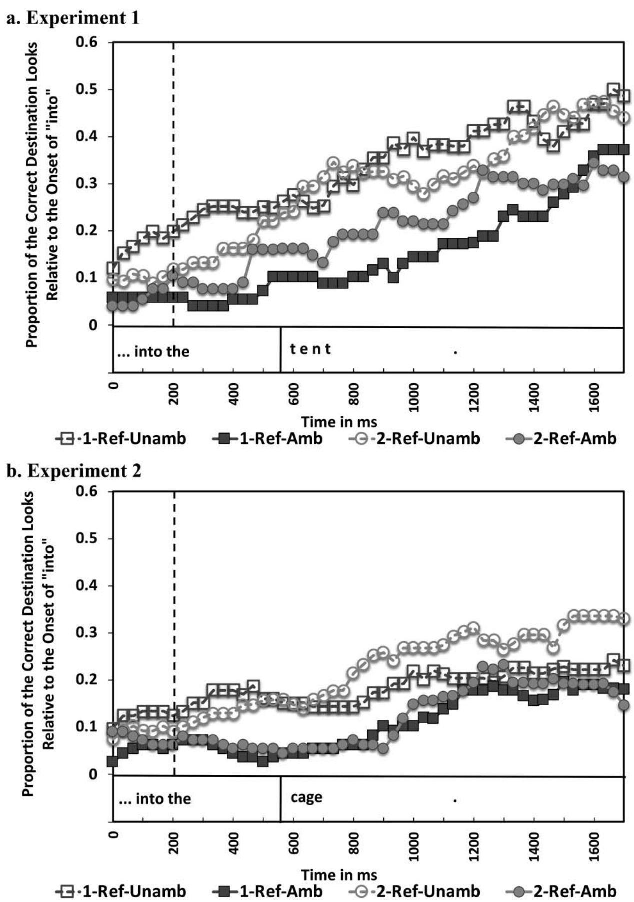

Figure 5. Proportion of fixations to the correct destination (e.g., the tent in Experiment 1 or the cage/cake in Experiment 2) relative to the onset of the second prepositional phrase (e.g., *into the . . .*). The analysis window was offset by 200 ms after the onset of *into* (see text). Open symbols show fixations in Unambiguous trials, and filled symbols in Ambiguous trials; square symbols show fixations in the one-referent condition, and circles show fixations in the two-referent condition.

with no significant interaction of Ambiguity and Referential Context (Figure 6a).

To further probe our interpretation of the effect of referential context fixations to the incorrect destination, we sought additional evidence that children in the two-referent condition considered a modifier interpretation earlier in the sentence. One measure that might reflect this is children's preference for the target animal upon hearing the ambiguous prepositional phrase (e.g., *on the pond*). We computed target animal preference as the proportion of fixations to the target animal divided by the proportion of fixations to either the target or distractor animal in an 800-msec time window anchored at the location noun (e.g., *pond*). This is the same analysis time window reported in Trueswell et al., 1999. In our study, this time window ended right before the disambiguating phrase *into the tent*. A target animal preference greater than 0.5 in the two-referent condition would reflect children's consideration of a modifier interpretation. Of course, we cannot directly compare the one- and two-referent conditions on this measure, as children in the one-referent condition saw only one possible referent for the animal name they heard (e.g., frog). As shown in Supplemental Figure 2a in the online supplemental materials, children in the two-referent condition showed a clear preference for the target animal (M = 0.61, SE = 0.05, $t[11]$ = 2.04, p = .03, compared with 0.5 chance level). This preference emerged only after *pond* and was not significant in a 1-s window extending from *frog* to *pond* (p = .14), suggesting there was no overall bias for the target animal.

Although this preference for the target animal (the frog on the pond) suggests children in the two-referent condition considered the modifier interpretation, this consideration was not sufficient to reduce their looks to the erroneous destination (the empty pond). In the same 800 ms time window, an analysis of looks to the incorrect destination showed no significant interaction between referential context and ambiguity (Supplemental Figure 3a in the online supplemental materials; b = 0.87, SE = 0.50, z = 1.75, p = .08), a null result also reported by Trueswell et al. (1999). Although this interaction neared significance, this was driven by opposing and nonsignificant differences between ambiguous and unambiguous sentences in the one-referent (b = −0.66, p = .10, ambiguous < unambiguous) and two-referent (b = 0.20, p = .60) conditions.

Overall, the analyses of actions and eye-movements at the end of the sentences revealed an ambiguity effect across both the one- and two-referent conditions: Children made more action errors and fewer fixations to the correct destination, in

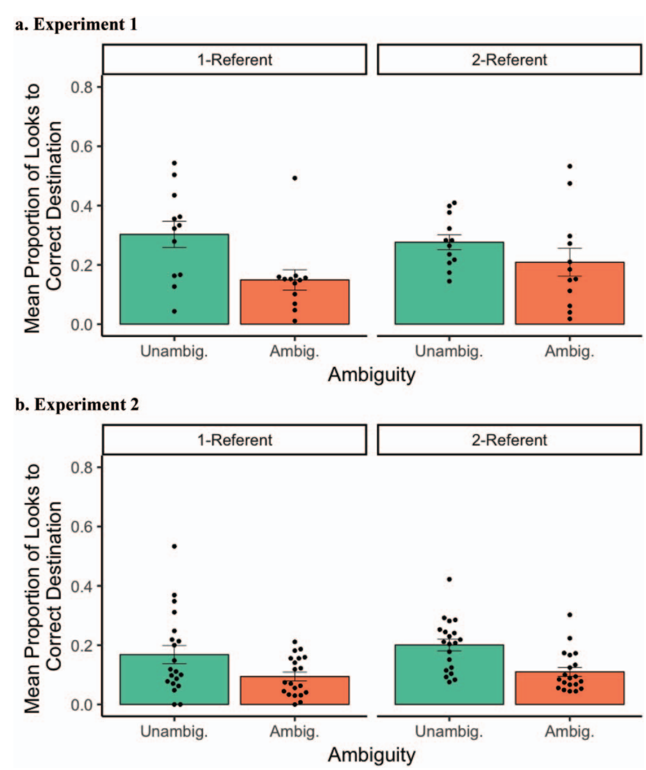

Figure 6. Mean proportion of looks to the correct destination after the onset of *into*, in Experiments 1 and 2. Error bars reflect ± 1 SE. A main effect of Ambiguity across both referential context conditions was observed in both experiments. See the online article for the color version of this figure.

ambiguous than in unambiguous trials. The evidence of a referential context effect emerged later in the sentence: Children in the two-referent condition made fewer erroneous fixations upon hearing the disambiguating prepositional phrase (e.g., *into the tent*) in ambiguous (as opposed to unambiguous) sentences compared with children in the one-referent condition.

Analyses of individual differences. We examined the relationship between Age, Language (SLAS), executive function (Simon Says) and children's performance in the *Put* sentence disambiguation task (ambiguity effect on action errors, looks to incorrect destination, and looks to correct destination). Table 4 shows descriptive statistics for the individual difference measures and Table 5 shows the correlations among these measures. Because Simon Says accuracy in the incongruent condition was not normally distributed, we conducted nonparametric Spearman correlation analyses.

Among the nine independent correlation analyses, only Simon Says accuracy was significantly associated with the ambiguity effect on children's actions ($\rho = -0.42$, $p = .02$, one-tailed). Children who successfully refrained from imitating the experimenter in the Simon Says inhibition trials were less likely to make errors in the ambiguous trials (as opposed to the unambiguous trials), compared with children who performed worse in Simon Says (Figure 7a). However, this correlation did not survive Bonferroni correction for multiple comparisons (nine comparisons). Neither Age nor SLAS was associated with measures from the sentence disambiguation task.

Summary of Experiment 1. Analysis of action and eye-movement data revealed evidence for two main conclusions.

First, we saw several indications of children's emerging sensitivity to referential context. In ambiguous trials, children in the two-referent context made numerically fewer action errors, and looked significantly less at the incorrect destination, than did children in the one-referent context, suggesting that 5-year-olds have some ability to use the referential context to avoid or recover from syntactic ambiguity.

Second, children with better executive functioning, as measured by Simon Says performance, made fewer action errors in coping with syntactic ambiguity in the sentence comprehension task. These findings, consistent with Woodard et al. (2016), suggest that children with better executive functioning were better able to override a strong initial bias toward the erroneous destination interpretation, and consider the alternative modifier interpretation. However, as noted above this correlation did not survive correction for multiple comparisons, suggesting further research is needed. Age and a vocabulary measure did not predict behavior in our task, suggesting that these effects are not simply due to differences in maturation or linguistic experience.

Table 5
Bivariate Spearman Correlations Between Individual Difference Measures and Sentence Disambiguation, Experiment 1 (n = 24)

	Individual difference measures		Sentence disambiguation		
Measure	SLAS	Simon says	Action	Incorrect destination looks	Correct destination looks
Age	0.16	−0.002	−0.04	0.003	−0.13
SLAS		0.21	−0.24	−0.19	−0.07
Simon Says		—	**−0.42**[*]	−0.24	0.21

Note. SLAS = Speech and Language Assessment Scale.
[*] Uncorrected $p < .05$, one-tailed, listed in bold.

Experiment 2

In Experiment 1, the action data showed a relatively low error rate compared with previous kindergarten-path studies, which might be due to the use of a slower speech rate or our use of a between-subjects referential context manipulation. The primary goals of Experiment 2 were to replicate the findings of Experiment 1 using (a) a different set of sentence stimuli and (b) a more sensitive executive functioning task. The key differences are as follows:

The sentence stimuli in Experiment 1 were composed of coherent scenes in which animals appeared in thematically related locations (e.g., *the frog on the pond*). This choice differed from previous studies (Trueswell et al., 1999), in which animals appeared in thematically unrelated locations (e.g., *the frog on the napkin*). To rule out the possibility that this thematic difference was the locus of children's relatively high accuracy in the two-referent ambiguous condition, in Experiment 2, we manipulated the thematic relatedness of animal and location nouns.

With respect to the executive function measure, in Experiment 1, the Simon Says measure resulted in a bimodal distribution of responses in our sample, which reduced its usefulness as an individual-difference measure. Further, children in Experiment 1 differed in their familiarity with the game, which may have com-

Table 4
Mean, Standard Deviation, and Reliability of the Individual Difference Measures of Experiment 1

	Individual difference measures			Sentence disambiguation[a]		
Measure	Age (months)	SLAS	Simon says[b]	Action	Incorrect destination looks	Correct destination looks
M	59.89	5.16	0.46	0.00	0.001	0.05
SD	2.53	0.60	0.40	0.72	0.02	0.05
Reliability	—	0.97[c]	0.82[c]	0.85[c]	—	—

Note. SLAS = Speech and Language Assessment Scale.
[a] The sentence disambiguation measures are by-subject random slopes estimated from the multilevel models. [b] Simon Says accuracy is the proportion of inhibition trials on which children correctly refrained from imitating. [c] Reliability was measured by Cronbach's alpha (Cronbach, 1951) across different items judged by the parent in SLAS and across different trials for Simon Says and the sentence task.

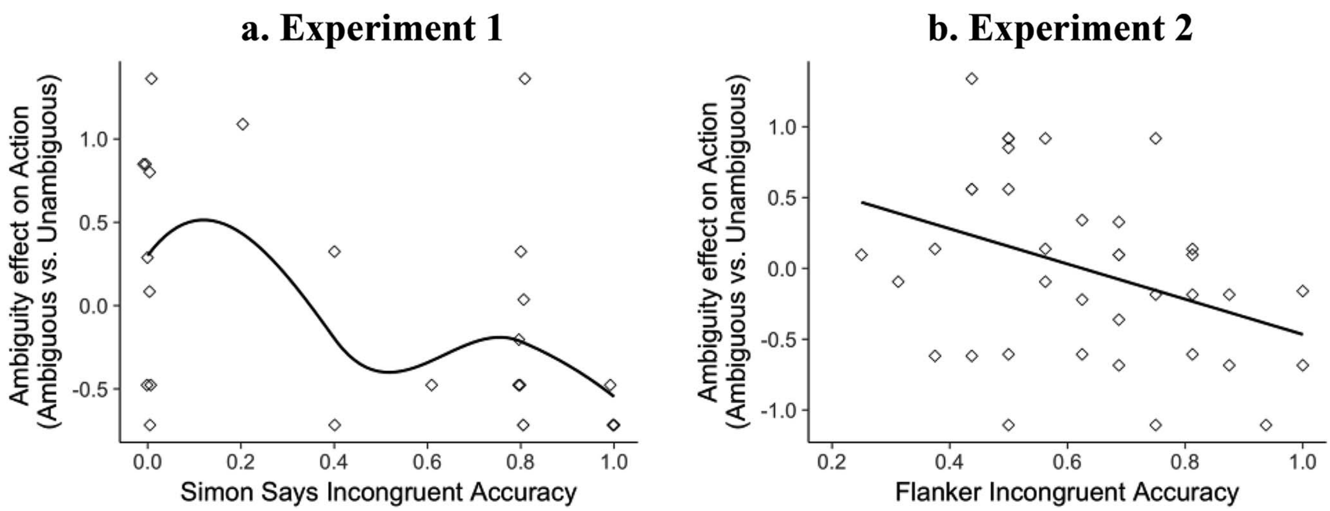

Figure 7. Significant negative correlation between the ambiguity effect (by-subject slope from the model) on action error rate and executive functioning measures (Simon Says Accuracy in Experiment 1, $\rho = -0.42$, $p = .02$; Flanker Accuracy in Experiment 2, $r = -0.37$, $p = .01$). Simon Says accuracy is the proportion of inhibition trials on which children correctly refrained from imitating. Flanker accuracy is the proportion of incongruent trials in the mixed block in which children responded correctly. The lines show trends of the correlation (a: local nonparametric regression trend line; b: linear tread line).

promised the validity of the measurement. To avoid these issues, in Experiment 2 we measured executive function using a computer-based child-appropriate Flanker task[4] which was novel to all of the 5-year-old participants. In this task, children saw five fish on the computer screen and were invited to feed either the middle or outside fish, depending on the color of the fish (i.e., feed the middle fish if the fish are blue, the outside fish if they are pink). The target and flanker fish either faced in the same direction or in different directions. Previous studies using similar tasks in both adults and children report reasonable test–retest reliability (adults: 0.52–0.77; children: 0.37–0.59; Fan, McCandliss, Sommer, Raz, & Posner, 2002; Rueda et al., 2004). Like the Simon-Says task we used in Experiment 1, the Flanker task requires children to shift their attention between opposing rules, while monitoring conflicts between visual cues or between their dominant response and game rules.

Method

Participants. Thirty-eight 5-year-olds (19 males and 19 females; mean age = 5;0, range 4;8 to 5;4), all native English speakers, participated in the experiment. Sentence ambiguity was manipulated within subjects. Referential context (one-referent vs. two-referent) and thematic relatedness (related vs. unrelated) were both manipulated between subjects. Children were randomly assigned to each of four between-subjects conditions: 10 in the one-referent Related condition, nine in the one-referent Unrelated condition, nine in the two-referent Related condition, and 10 in the two-referent Unrelated condition. The data from six additional children were eliminated because experimenter error (two), high error rates in unambiguous trials (one), or because the child declined to complete the experiment (three). This study was approved by the Institutional Review Board at the Office of the Vice Chancellor for Research at the University of Illinois, Urbana-Champaign.

Materials and procedure. As in Experiment 1, each child completed 12 critical trials (six ambiguous, six unambiguous), randomly interspersed among six filler trials. Each target sentence in the critical trials contained the verb *put* and appeared in an ambiguous or unambiguous form. The Related and Unrelated target items contained the same animal names but different location and destination nouns. Location and destination nouns were matched in syllable length across Related and Unrelated sentences (e.g., *mud* vs. *shoe* and *cage* vs. *cake*), and were also chosen to be about equally frequent, based on the 425-million-word Corpus of Contemporary American English (Davies, 2009). The Related and Unrelated prepositional phrase objects (e.g., *mud* vs. *shoe*) for the direct object (e.g., *pig*) were chosen based on our judgments of real-world plausibility and were then validated based on the Corpus of Contemporary American English. The animal and location nouns we chose for Related sentences were significantly more likely to co-occur within a 4-word window in this corpus (e.g., *pig* with *mud*; $M = 36.91$ co-occurrences, range 3–112) than were those chosen for Unrelated sentences (e.g., *pig* with *shoe*; $M = 1.25$, range 0–7; $t[11] = 9.77$; $p < .005$). Filler sentences again contained the verb *Put* (e.g., *Put the sheep beside the grass*).

Stimulus utterances were recorded as in Experiment 1 and were selected to be similar in their overall utterance length to those of Experiment 1 but with less modifier-biased prosody. Average target utterance length was 3.26 s (range 3.06–3.53 s) for ambiguous sentences and 3.45 s (range 3.25–3.69 s) for unambiguous sentences. In ambiguous target sentences, the average onset of the

[4] We thank Adele Diamond and David Abbott for generously sharing the computer scripts and materials for the Flanker Task.

location noun (e.g., *mud* or *shoe*) was 1.41 s (range 1.21–1.74 s), the disambiguating prepositional word (e.g., *into*) was 1.88 s (range 1.59–2.13 s) and the destination noun (e.g., *shoe*) 2.90 s (range 2.71–3.28 s), after the onset of *Put*. As reported in Experiment 1, we calculated Kraljic and Brennan's (2005) prosodic boundary difference score for our items. The average duration difference between the direct object noun (*pig*) and the first prepositional object (*mud*) was 0% relative to the length of the utterance. Examples of the visual stimuli in the form of 2D felt cutouts are shown in Figure 1b and 1c. See Appendix B for a full list of materials.

The apparatus and procedure were the same as in Experiment 1.

Data coding.

Action data. Children's actions and eye-movements were coded as in Experiment 1. Trials with fewer than one-third of the frames coded as fixations to a quadrant of the board were excluded from eye-movement analyses (35 of 456 possible trials, 7.7%). A second coder coded the data from 10 randomly selected children, and the two coders agreed on 92.1% of coded video frames.

Individual difference measures. We used a computer-based Flanker task (Diamond, 2013) to obtain a measure of executive functioning for each child. In each trial, children saw five fish on a computer screen and were instructed to press the left or the right arrow key on the computer's keyboard to feed the hungry fish. The target key depended on the left-right orientation of the hungry fish, and the color of the fish on the screen told children which fish to pay attention to. When the fish were blue, the middle fish was the hungry one (so if it was facing left, children should press the left arrow key); when the fish were pink, the outside fish were the hungry ones. In congruent trials all five fish faced the same direction; in incongruent trials the middle and outside fish faced opposite directions. Children completed three blocks of the game, the first blue (always feed the middle fish, four practice trials and 17 test trials), the second pink (always feed the outside fish, four practice trials and 17 test trials), and the third with blue and pink trials intermixed (8 practice trials and 45 test trials). Congruent and incongruent trials were randomly interspersed within each block. Correct responding in incongruent trials requires the child to follow the task rules, responding to the direction of the target fish and not the flanking fish; correct responding in the third block additionally requires switching flexibly between rules in response to a visual cue.

Table 6 shows the mean proportion of accurate responses and median response times in each condition and block. These group data show a typical flanker effect: children were less accurate and slower in trials with incongruent flankers than in those with congruent flankers. Consistent with previous findings (Munro, Chau, Gazarian, & Diamond, 2006), the flanker effect was most robust in the third, mixed block, in which both accuracies and response times differed significantly in congruent versus incongruent trials ($ts[37] > 3.4$, $ps < 0.002$). The mixed block posed the greatest challenge for top-down control, because children had to switch flexibly between opposing rules (feeding the middle vs. the outside fish), learned in the two previous blocks, depending on the color cue. Children's performance in this block reflected both their inhibitory control ability and their cognitive flexibility. Accuracy tends to be a more sensitive measure than response time for young children, because of their tendency to respond impulsively. For example, preschoolers' improvement in accuracy with age (from 4 to 6 years) is far more evident than their improvement in response time in a range of executive function tasks (Davidson et al., 2006). Therefore, we used each child's proportion of correct responses on incongruent trials in the mixed block as our measure of inhibitory control, a measure analogous to successful inhibition in the dog/dragon task of Nilsen and Graham (2009) and Simon Says of Experiment 1.

Table 6
Mean Proportion of Correct Responses and Mean of Median Response Time for Correct Responses, in ms, in the Flanker Task, by Block, Experiment 2

Flanker type	Target type: Block 1: Middle (blue)	Block 2: Outside (pink)	Block 3: Mixed (blue and pink)
Accuracy			
Congruent	0.75 (0.04)	0.86 (0.04)	0.80 (0.03)
Incongruent	0.73 (0.03)	0.69 (0.04)	0.62 (0.03)
Difference	0.03 (0.05)	0.17 (0.04)***	0.18 (0.03)***
Response time			
Congruent	1,565 (19)	1,457 (15)	1,498 (14)
Incongruent	1,668 (18)	1,542 (16)	1,771 (16)
Difference	−83 (15)	−84 (16)	−276 (13)**

Note. Nine trials were excluded from the analysis because children made impulsive responses within 200 ms (range: 11–137 ms). Standard errors are presented in parentheses. Asterisks indicate statistical significance of the differences between the congruent and the incongruent trials.
** $p < .01$. *** $p < .001$.

Woodard et al. (2016) used a different variant of the Flanker task, and measured children's executive functioning by computing a switch cost equal to the average of the normalized difference in accuracy and reaction time (RT) between trials preceded by a trial of the same type (congruent trials after congruent trials, or incongruent after incongruent) and those preceded by a trial of a different type (congruent trials after incongruent trials, or incongruent after congruent). Our flanker task did not permit us to use the same measure due to an insufficient number of nonswitch trials. Moreover, such a measure of switch cost does not reflect an explicit task switch. Children might essentially do the same task across the congruent and the incongruent trials. The third block of our Flanker task required children to switch frequently between the blue rule (feed the middle fish) and the pink rule (feed the outside fish). The explicit rule switch in our task enabled us to measure switch cost regardless of children's success in inhibition. Table 7 shows accuracy and RT of trials requiring a switch in rule (e.g., a pink trial following a blue trial) versus nonswitch trials (e.g., a pink trial following a pink trial); only RT showed a significant switch cost. Therefore, we also computed a switch cost score by calculating the response time difference between switch and nonswitch trials in the mixed block as a measure of cognitive flexibility.

In Experiment 1 we estimated verbal ability using a parental SLAS questionnaire. To obtain a more precise measure of children's verbal ability, in Experiment 2 children completed an abbreviated version of the Peabody Picture Vocabulary Test (Third Edition, Dunn & Dunn, 1997) at the end of the study. To keep the experimental session under one hour, all participants started from picture set six, rather than set two or four, the suggested starting

Table 7
Mean Proportion of Correct Responses and Mean of Median Response Time for Correct Responses (in ms) in Block 3 of the Flanker Task, by Switch/Nonswitch Trial Type, Experiment 2

Measure	Switch	Nonswitch	Difference
Accuracy	0.80 (0.03)	0.77 (0.03)	0.04 (0.03)
Response time	1,797 (83)	1,563 (78)	233 (45)***

Note. Standard errors are presented in parentheses. Asterisks indicate statistical significance of the differences between the switch and the nonswitch trials.
*** $p < .001$.

points for 4- and 5-year-olds. Note this change may slightly elevate PPVT scores relative to published norms. Therefore, we used PPVT raw scores as individual measures of vocabulary.

Children assigned to the one- and two-referent conditions did not differ reliably in age (mean age ± *SD* = 5;1 ± 0;3 vs. 5;0 ± 0;2), gender ratio (boys: girls = 8:6 vs. 6:7), Flanker accuracy in incongruent trials ($M = 0.62$ vs. 0.63), Flanker switch cost ($M = 240.5$ ms vs. 332.1 ms), or PPVT raw scores ($M = 104$ vs. 110, all $ps > 0.25$). Children assigned to the Related and Unrelated conditions did not differ in age ($M = 5{:}1$ vs. 5; 0) or PPVT raw scores ($M = 106$ vs. 108, all $ps > 0.29$). However, although children were randomly assigned to relatedness conditions, children in the Related condition had marginally lower Flanker accuracy in incongruent trials ($M = 0.57$ vs. 0.68, $F[1, 36] = 3.08$, $p = .09$) and a marginally smaller Flanker switch cost ($M = 210.5$ ms vs. 362.2 ms, $F[1, 36] = 3.49$, $p = .07$) than those in the Unrelated condition. Experiment 1 showed that individual differences in executive functioning were related to child's success in resolving syntactic ambiguity, as measured by children's action errors; therefore, this difference presents a confound that makes it difficult to evaluate the effect of semantic relatedness in our task. For this reason, the main analyses below collapse across the two relatedness conditions; in what follows, however, we also report descriptive statistics suggesting that there is little difference between the Related and Unrelated conditions, and that our key data patterns are replicated in both conditions.

Data and code sharing. The data and the analysis R codes are available on https://github.com/zhenghanQ/forestgit.

Results and Discussion

Analysis of children's actions. Table 8 shows the distribution of correct actions and errors, and Figure 2b shows the proportion of trials with an action error, collapsing across relatedness conditions. The data pattern closely resembled that of Experiment 1: Children made many more action errors in ambiguous (48.2%) than unambiguous trials (3.5%), and the hopping error was the most common error type. Also as in Experiment 1, the overall error rate in ambiguous trials was somewhat lower than in previous published experiments of this type, and children in the two-referent condition made fewer errors than did those in the one-referent condition, rather than the reverse (Figure 2b).

As in Experiment 1, error proportions for each child (empirical-logit transformed) were analyzed using a multilevel linear model that included fixed effects for Ambiguity, Referential Context, their interactions, and by-subject random intercepts and random slopes for ambiguity. The model, shown in Table 9, revealed significant main effects of Ambiguity (more errors in ambiguous trials), Referential Context (fewer errors in the two-referent condition), and a significant interaction of Ambiguity and Referential Context. Children in the two-referent condition showed a smaller ambiguity effect in their action errors than did those in the one-referent condition. This suggests sensitivity to the referential principle: The two-referent context helped children to infer the need for a modifier and thus to recover from the temporary syntactic ambiguity.

Note that semantic relatedness did not significantly affect error rates (unrelated error rate = 0.21; related = 0.29), and adding this

Table 8
Distribution of Correct Actions and Action Errors in the One-Referent and Two-Referent Conditions, Experiment 2

		One-referent condition ($n = 19$)		Two-referent condition ($n = 19$)	
Action category	Animals moved	Ambiguous	Unambiguous	Ambiguous	Unambiguous
Correct	Target	46	109	72	111
	Distracter	—	—	—	—
Hopping	Target	53	3	9	0
	Distracter	0	0	16	0
Falling short	Target	12	0	4	0
	Distracter	0	1	8	1
One of each	Target	0	0	1	0
	Distracter	2	0	4	1
	Both	0	1	0	1
No response		1	0	0	0
Total errors		68	5	42	3
Total trials		114	114	114	114

Note. Hopping: children moved either animal to the incorrect destination before eventually setting it down at the correct destination. Falling short: children moved either animal to the incorrect destination and stopped there. One of each: children moved one animal to the correct destination and the other animal to the incorrect destination, or moved only the distractor animal to the target destination.

Table 9
Summary of Fixed Effects in Mixed Effect Linear Model of Proportion of Errors in Action (Transformed Using an Empirical-Logit Function) in Experiment 2

Predictors	Estimate	*SE*	*t* value	*p* value
(Intercept)	−1.22	0.15	−8.07	1.3×10^{-9}
Ambiguity (amb. vs. unamb)	**2.27**	**0.27**	**8.48**	$\mathbf{4.2 \times 10^{-10}}$*
Referential context (2-ref vs. 1-ref)	**−0.61**	**0.30**	**−2.02**	**.05***
Referential Context × Ambiguity	**−1.07**	**0.54**	**−2.00**	**.05***

Note. Akaike information criterion (AIC) = 261.19. Statistically significant main effects are listed in bold.
* $p < .05$.

factor or its interaction with either Ambiguity or Referential Context into the analysis summarized in Table 9 did not significantly improve the fit of the model ($\chi^{2'}$s[1] < 0.25, *ps* > 0.6). Most importantly, the pattern of Figure 2b held regardless of relatedness condition: both in the Related and the Unrelated conditions, children in the two-referent condition made fewer errors in ambiguous trials than did those in the one-referent condition.

Additional evidence that children in the two-referent condition achieved a modifier interpretation in the ambiguous trials comes from analysis of referent choice (see Table 8). As in Experiment 1, children picked up only the target animal on 75% of ambiguous trials (*se* = 5%), significantly more often than expected by chance (Wilcoxon's *z*-score = 3.18, $p < .0001$). Despite this pattern, the two-referent group still made many more action errors in the ambiguous trials compared with the unambiguous ones, suggesting this ability to interpret the first prepositional phrase as a modifier was fragile.

Analysis of children's eye movements. Because the durations of the sentence stimuli in both experiments were similar, we used the same analysis windows as in Experiment 1. The main analysis window was 1.5 s long and was anchored on the disambiguating prepositional phrase (e.g., *into the cage*), offset by 200 ms. As in Experiment 1 we report two main dependent measures: (q) children's looks at the incorrect destination (e.g., the empty mud/shoe) and (b) children's looks at the correct destination following the disambiguating prepositional phrase *into*.

Figure 3b plots the proportion of fixations to the incorrect destination (the empty mud). As in Experiment 1, children spent the most time looking at the incorrect destination in the one-referent ambiguous condition.

A binary measure of incorrect destination fixations at each time point during this analysis window on each trial was fit with autoregressive GLMMs (Cho et al., 2018) with crossed random intercepts for subjects and items, by-subject random slopes for ambiguity and lag 1 response, and by-item random slopes for ambiguity, referential context and lag 1 response (see Table 10). The analysis revealed a significant main effect of Ambiguity. As shown in Figure 4b, children looked marginally more at the incorrect destination in the ambiguous than the unambiguous trials. Although there was no significant interaction between Ambiguity and Referential Context, the Ambiguity effect was numerically larger in the one-referent ($b = 0.46$) than the two-referent condition ($b = 0.15$), consistent with Experiment 1. As in Experiment 1, we asked whether increased looks to the two animals in the two-referent display, as children resolved the referential ambiguity, might have contributed to the numerically lower incorrect destination looks in the two-referent condition. Again, children spent a similar amount of time looking at the animals during this window in the one-referent ($M = 0.70$, $SE = 0.07$) and two-referent conditions ($M = 0.73$, $SE = 0.06$).

As in Experiment 1, additional evidence consistent with consideration of a modifier interpretation in the two-referent context comes from analysis of children's preference to fixate the target animal, as opposed to the distractor animal, after hearing the ambiguous location noun (e.g., *mud*) till 800 ms later, a window that ends before children encounter the disambiguating preposition phrase (e.g., *into the cage*). Supplemental Figure 1b in the online supplemental materials again showed a reliable preference for the target animal ($M = 0.63$, $SE = 0.04$, $t[18] = 3.45$, $p = .001$ compared with 0.5 chance level). This target-animal preference was not significant before the ambiguous location noun ($p = .72$), confirming no target animal bias in the two-referent condition.

Also as in Experiment 1, although this preference for the target animal (the pig in the mud) during the period of ambiguity suggests children in the two-referent condition considered the modifier interpretation, this did not suffice to reduce their looks to the

Table 10
Summary of Fixed Effects in Autoregressive Generalized Linear Mixed Effect Model Analyses of the Proportion of Looks to the Incorrect Destination (e.g., the Empty Mud/Shoe) and the Correct Destination (e.g., the Cage/Cake) After the Onset of the Disambiguating Prepositional Phrase (e.g. Into the Cage), Experiment 2

Dependent variable	Predictor	Estimate	*SE*	*t* value	*p* value
Incorrect destination looks (AIC = 2171.6)	(Intercept)	−5.26	0.17	−30.26	$<10^{-16}$
	Lag 1 response	7.97	0.22	36.54	$<10^{-16}$
	Ambiguity (amb. *vs*. unamb)	0.33	0.18	1.83	.07†
	Referential context (2-ref vs. 1-ref)	0.03	0.18	0.15	.88
	Referential Context × Ambiguity	−0.38	0.31	−1.22	.22
Correct destination looks (AIC = 3285.5)	(Intercept)	−4.48	0.11	−41.65	$<10^{-16}$
	Lag 1 Response	7.36	0.12	59.97	$<10^{-16}$
	Ambiguity (amb. vs. unamb)	**−0.30**	**0.13**	**−2.27**	**.02***
	Referential context (2-ref vs. 1-ref)	0.26	0.15	1.76	.08
	Referential Context × Ambiguity	−0.27	0.25	−1.11	.27

Note. AIC = Akaike information criterion. Window duration: 200–1,667 ms after the onset of *into*. Statistically significant main effects are listed in bold.
† $p < .1$. * $p < .05$.

erroneous destination (the empty cage). In the same 800-ms time window, an analysis of looks to the incorrect destination revealed no significant interaction between referential context and ambiguity (Supplemental Figure 3b in the online supplemental materials; $b = -0.08$, $SE = 0.44$, $z = -0.18$, $p = .86$).

Finally, children's fixations to the correct destination reflected the typical difficulty in interpreting ambiguous structures. Figure 5b plots children's fixations to the correct destination (e.g., *cage* or *cake*) after the onset of the disambiguating phrase (e.g., *into the cage* or *onto the cake*). The analysis of children's looks to the correct destination (the bottom section of Table 10) revealed a significant effect of Ambiguity and a marginal effect of Referential context. As in Experiment 1, children were more likely to look at the correct destination in the unambiguous than the ambiguous trials (Figure 6b), suggesting the referential context offered only limited support.

Individual differences analyses. We examined the relationship between Age, Language (PPVT), executive function (Flanker) and children's performance in the *Put* sentence disambiguation task. Table 11 shows descriptive for of the individual difference measures and Table 12 shows the bivariate Pearson correlations between the individual-difference measures and sentence disambiguation measures estimated based on by-subject random slopes estimated from the mixed-effect models. Among the eight correlational analyses, Flanker accuracy in the incongruent trials was significantly associated with the ambiguity effect on action errors ($r = -0.37$, $p = .01$, one-tailed, Figure 7b), incorrect destination looks, $r = -0.31$, $p = .03$, one-tailed, and correct destination looks, $r = .34$, $p = .02$, one-tailed. Flanker switch cost was significantly associated with the ambiguity effect on incorrect destination looks, $r = .37$, $p = .01$, one-tailed. Note that these correlations were no longer significant after Bonferroni correction for multiple comparisons. However, multiple linear regression models confirmed the significant unique contribution of Flanker accuracy to children's actions, and a marginal unique contribution of Flanker accuracy in predicting children's online fixation patterns (see Table 13). These findings replicate, with a different set of sentences and a different assay of executive function, a key result of Experiment 1, offering more support for the link between syntactic ambiguity resolution and executive function in children.

Summary of Experiment 2. The action data in Experiment 2 provide further evidence for preschoolers' use of referential-context cues, as well as an influence of individual differences in executive functioning on the ambiguity resolution process. Children in the two-referent condition made significantly fewer action errors than did their peers in the one-referent condition when interpreting ambiguous sentences, demonstrating for the first time that they were able to incorporate visual context into syntactic processing. Children's likelihood of making errors in their interpretation of the ambiguous sentences was associated with their success in solving conflicts and shifting between opposing game rules in the flanker task.

As in Experiment 1, the eye-movement results also suggested an attenuation of the ambiguity effect in the two-referent compared with the one-referent condition during the second prepositional phrase, though this effect was not significant in Experiment 2. Children's eye-movement patterns during online syntactic disambiguation were marginally associated with their success in the flanker task.

General Discussion

The results of two experiments examining young children's parsing of temporary syntactic ambiguities yielded two major findings. First, our results provide new evidence that under some circumstances (we return to how those circumstances might best be characterized below), children use information from the referential context to guide syntactic processing, broadly consistent with constraint-based theories of adult sentence processing (MacDonald et al., 1994; Trueswell & Tanenhaus, 1994). Second, these findings point to the maturation of executive function as a key factor in young children's syntactic disambiguation (Woodard et al., 2016). We next discuss each main finding in turn.

In the classic studies with which we began, children failed to show sensitivity to the visual referential context in syntactic ambiguity resolution (e.g., Trueswell et al., 1999; Weighall, 2008), suggesting that they were unable to use the referential principle to influence parsing decisions. As we noted at the outset, this has long struck us as surprising, given universal agreement that young children use likely speaker meanings inferred from real-world contexts to guide language acquisition (Fisher, Gertner, Scott, & Yuan, 2010; Gleitman, 1990; Pinker, 1984; Tomasello, 1995; Tomasello, 1999). We built on hints in the literature that 5-year-olds showed delayed rather than absent sensitivity to the referential principle in online comprehension of simpler sentences, but also in parsing globally ambiguous sentences such as *Feel the frog with*

Table 11
Mean, Standard Deviation, and Reliability of the Individual Difference Measures of Experiment 2 (N = 38)

	Individual difference measures				Sentence disambiguation[a]		
Measure	Age (months)	PPVT Raw	Flanker incong.[b]	Flanker switch cost (ms)[c]	Action	Incorrect destination looks	Correct destination looks
M	60.20	106.95	0.63	233.17	0.00	0.03	−0.005
SD	2.60	16.69	0.19	45.26	0.64	0.26	0.18
Reliability	—	—	0.88[d]	—	0.89[d]	—	—

Note. PPVT = Peabody Picture Vocabulary Test.
[a] The sentence disambiguation measures are by-subject random slopes estimated from the multilevel models. [b] Flanker incong. is the proportion of incongruent trials in the mixed Block 3 on which children responded correctly. [c] Flanker switch cost is the response time difference between switch trials and nonswitch trials in the mixed Block 3. [d] Reliability was measured by Cronbach's alpha (Cronbach, 1951) across different trials in the mixed Block 3 of the Flanker task and across different trials in the sentence task.

Table 12
Bivariate Correlations Between Individual Difference Measures and Sentence Disambiguation Measures, Experiment 2 (n = 38)

	Individual difference measures		Sentence disambiguation			
Measure	Flanker incong.	Flanker switch cost	PPVT	Action	Incorrect destination looks	Correct destination looks
Age	0.12	0.10	0.25	0.06	0.18	0.14
Flanker incong.		−0.10	−0.03	**−0.37***	**−0.31***	**0.34***
Flanker switch cost			0.28	**0.14**	**0.37***	**−0.08**
PPVT			—	0.14	0.26	−0.08

Note. PPVT = Peabody Picture Vocabulary Test.
* Uncorrected $p < .05$, one-tailed listed in bold.

the feather (Snedeker & Trueswell, 2004). In our task, with modifications intended to give children more time to recruit referential-context cues, we found new evidence of 5-year-olds' sensitivity to the referential context in online parsing. We observed notably fewer action errors in our study than previous kindergarten-path studies (e.g., Anderson et al., 2011; Hurewitz et al., 2000; Trueswell et al., 1999; Weighall, 2008).

Both experiments showed the same patterns: First, online eye-gaze data revealed that children spent more time looking at the incorrect destination (e.g., the empty pond) in ambiguous sentence trials in the one- than in the two-referent condition (although the critical ambiguity by referential context interaction was significant only in Experiment 1). This effect appeared in a late time window as the sentence ended and children prepared to act; this finding replicates key prior evidence for emerging sensitivity to the referential context reported by Snedeker and Trueswell (2004). Second, we found the first evidence that children's offline interpretations of temporarily ambiguous sentences showed sensitivity to the referential context. Children made more action errors in comprehending temporarily ambiguous sentences in the one- than in the two-referent condition (though the ambiguity by referential context interaction was significant only in Experiment 2). Moreover, we found, also for the first time in this task, that children chose to act on the target animal (e.g., the frog on the pond) reliably more often than the distractor animal in two-referent contexts, suggesting that our task modifications permitted children to entertain the (correct) modifier interpretation of the temporarily ambiguous prepositional phrase. The emergence of these effects in children's actions, as well as in their late eye movements, tells us that 5-year-olds' sensitivity to the referential context need not always be too little and too late to influence their final interpretations.

Table 13
Results of Multiple Regression Models Showing the Relationship Between Individual Difference Measures and Sentence Disambiguation Measures

Model	Estimate	*SE*	*t* value	*p* value
Ambiguity effect on action				
Age	0.02	0.04	0.49	.63
Flanker incong.	**−1.25**	**0.54**	**−2.28**	**.03***
Flanker switch cost	0.0002	0.0004	0.432	.67
PPVT	0.003	0.007	0.49	.63
Ambiguity effect on incorrect destination looks				
Age	0.002	0.02	0.12	.90
Flanker incong.	−0.38	0.21	−1.87	.06†
Flanker switch cost	0.0002	0.0001	1.68	.07†
PPVT	0.004	0.002	1.52	.14
Ambiguity effect on correct destination looks				
Age	0.009	0.07	0.79	.44
Flanker incong.	0.03	0.01	1.97	.06†
Flanker switch cost	−0.00002	0.0001	−0.18	.86
PPVT	−0.001	0.002	−0.57	.57

Note. PPVT = Peabody Picture Vocabulary Test. Significant independent variables are listed in bold.
† $p < .1$. * $p < .05$.

We found these effects even though the *put* sentences in the classic *frog on napkin* task arguably represent a worst-case scenario for the integration of top-down referential context cues into parsing decisions. Because the verb *put* strongly predicts a destination phrase, it biases listeners toward the destination interpretation of the temporarily ambiguous prepositional phrase. As noted in the Introduction, such strong bottom-up constraints might limit the influence of top-down constraints, simply because top-down processing takes time (e.g., Britt, 1994; see discussion in Snedeker & Trueswell, 2004). Interestingly, the (late) effects of referential context that Snedeker and Trueswell (2004) found in children's eye-gaze patterns were numerically strongest in sentences with equibiased verbs that did not permit strong predictions about prepositional phrase attachment. In the present data, we found evidence for an emerging adult-like capability to integrate referential context into their parsing decisions. Moreover, despite the facilitation from the two-referent context, our analyses of correct-destination fixations revealed a robust ambiguity effect similar to that found in adults (Novick et al., 2008). This suggests that the strong bias of the verb *put* for a destination phrase creates a garden-path in test sentences like ours that is difficult for referential-context cues to eliminate entirely.

These results provide new evidence for developmental continuity in the use of referential-context cues in syntactic parsing and thus strong new evidence that sentence processing in children, like adults, continuously integrates multiple constraints. The difference between our findings with children, and previous findings with adults in similar tasks, lies primarily in the time course and stability of referential-context effects, rather than in whether such effects appear at all. Adults quickly and strongly show evidence of referential-context effects in parsing temporary ambiguities (Novick et al., 2008; Trueswell et al., 1999), whereas 5-year-olds appear to incorporate referential-context cues more slowly, and therefore fail to show effects of these cues in their offline sentence interpretation, unless task features give them more time to do so. In contrast, slightly older children more often manage to recruit referential-context cues in time to affect offline sentence interpretation (Weighall, 2008; Weighall & Altmann, 2011).

By the same token, our results provide new evidence that referential-context cues are employed in much the same way in the online processing of different kinds of modifiers and different kinds of linguistic ambiguity. For example, Huang and Snedeker (2013) proposed that 5-year-olds showed sensitivity to the referential principle in interpreting scalar adjectives (*the big coin*), because scalar adjectives are more strongly linked with multiple-referent contexts in the child's language experience than are other types of modifiers. This interpretation is in line with a cue-validity account, which suggests that children might recruit different combinations of cues to resolve different linguistic ambiguities, depending on separate calculations of cue validity for each type of ambiguity (see Rabagliati, Pylkkänen, & Marcus, 2013, for an analogous proposal regarding the use of top-down cues in lexical ambiguity resolution). In contrast, the present evidence that 5-year-olds also recruit referential-context effects in parsing syntactically ambiguous postnominal modifiers points toward a more general sensitivity to the referential principle in language comprehension.

Interestingly, similar arguments have recently been made regarding preschoolers' sensitivity to pragmatic constraints in other aspects of language processing. When adults hear a sentence such as *I ate some of the cookies*, they infer that *not all* of the cookies were consumed. This inference reflects the hearer's knowledge of alternative terms that the speaker could have chosen (e.g., *all the cookies*). Children famously struggle with such inferences (e.g., Huang & Snedeker, 2009; Noveck, 2001), and these failures, along with children's failures to recruit the referential principle in the typical 'frog on napkin' task, might be seen as evidence that children are generally poor at pragmatic inference. However, recent evidence suggests that young children's failures in these tasks reflect a failure to retrieve (or perhaps to retrieve in time) the alternative terms the speaker could have used, which are needed to drive the pragmatic inference. For example, 5-year-olds are more likely to interpret *some* as implying *not all* if the alternative term *all* is made readily accessible by its earlier use within the same task (Skordos & Papafragou, 2016; see also Barner, Brooks, & Bale, 2011), and 3- to 4-year-olds draw ad hoc pragmatic implicatures when the alternatives available to the speaker are clearly shown in the visual context (Stiller, Goodman, & Frank, 2015). We see our data, like these recent findings, as evidence that youngsters' language comprehension may be more sensitive to pragmatic considerations than sometimes meets the eye.

It might seem puzzling that the significant effects of referential context appeared in different measures across the two experiments. Although we saw the same data patterns in both experiments, as noted throughout, we saw a significant interaction of referential context and ambiguity in the online eye-tracking data in Experiment 1, but in the offline action data in Experiment 2. These differences in which effects emerged as significant in the two experiments could have been caused by multiple differences. For example, we noted that the prosody of sentence stimuli was less modifier-biased in Experiment 2; also in Experiment 2, half of the participants heard sentences with pairs of nouns (e.g., *the pig on the lamp*) that were chosen to be unrelated, and that were highly implausible. Either or both of these changes might have increased the overall difficulty of understanding the sentences, resulting in higher offline error rates (see Figure 2) and increased looks to the incorrect destination (see Figure 3) in Experiment 2 relative to Experiment 1. Higher offline error rates, reflecting a failure to revise an initial misinterpretation before launching an action, might have provided a better opportunity for us to measure reliable facilitation from the two-referent context. In addition, different samples of participants might simply have differed in within-group variability. However, as noted earlier, the overall pattern of numerically fewer erroneous looks and actions in the two-referent than in the one-referent condition had not been observed in previous studies. We interpret this robust similarity between the two experiments as evidence for children's ability to use referential information, when given time to do so.

What aspects of our task made it possible for us to measure 5-year-olds' sensitivity to the referential context, in contrast to previous studies? We used a slow speech rate and a between-subjects manipulation of referential context. The referential context effect in both experiments suggested that children were more likely to succeed in incorporating the top-down referential-context information during online parsing, when offered more time and repeated opportunities to notice the competing referential options. These findings provided support for the *top-down processing account*. In addition, our recordings included modest versions of the prosodic pattern that adults use in producing prepositional-phrase modifiers, which might serve as another helpful cue for children. Kindergarteners successfully integrate exaggerated versions of similar prosodic boundary cues during comprehension of syntactically ambiguous sentences (Snedeker & Yuan, 2008), and even toddlers can use prosodic boundary cues to constrain syntactic analysis in other contexts (e.g., de Carvalho, Dautriche, Lin, & Christophe, 2017). In natural speech, adults sometimes produce prosodic boundaries to disambiguate sentences similar to our stimuli (Kraljic & Brennan, 2005; Snedeker & Trueswell, 2004), including in speech to children (Snedeker & Yuan, 2008). However, as far as we know there is no evidence that children can recruit prosodic boundary cues as modest as those we measured in our stimuli to resolve syntactic ambiguity (and see Snedeker & Yuan, 2008, for a report that 5-year-olds could not do so). In sum, we assumed some combination of these three features (slow speech rate, between-subjects manipulation, and prosodic cues) might have increased children's chances of considering the correct modifier interpretation of the ambiguous prepositional phrase. However, because we did not separately manipulate these factors, we cannot make inferences about which of our task modifications were helpful. Future studies may speak to which of these factors most strongly supported young children's ability to recover from these temporary ambiguities.

Our study, following Woodard et al. (2016), also examined the link between domain-general executive functioning and the developing ability to recover from garden-path misinterpretations, and yielded further evidence for the proposed link. In both experiments, children's performance in an executive function task (Simon Says in Experiment 1, and Flanker in Experiment 2) predicted how many action errors children made in processing garden-path sentences. In Experiment 2, children with better executive functioning (higher accuracy and smaller switch cost) also tended to exhibit fewer erroneous looks to the incorrect destination, although the unique contribution of these measures in predicting children's eye gaze was only marginally significant in a multiple regression analysis. Moreover, children's language abilities (measured using SLAS in Experiment 1, and PPVT in Experiment 2) were not associated with syntactic disambiguation measures. These findings highlight the specificity of the relationship between executive functioning and syntactic disambiguation, especially when children are given sufficient time to process these sentences.

These results therefore provide additional evidence that domain-general executive functioning plays a role in children's developing ability to recover from temporary syntactic ambiguity (Khanna & Boland, 2010; Mazuka, Jincho, & Oishi, 2009; Novick, Trueswell, & Thompson-Schill, 2010). Children with better executive functioning may have been better able to inhibit their own initial interpretation long enough to detect the ambiguity (given the visual context) of sequences such as *the frog on the pond*. In addition, children with stronger executive functioning may have been better able to correct their initial garden-path misinterpretation after hearing the disambiguating cue (e.g., *into the tent*). These correlational results suggest that domain-general cognitive processes are engaged in achieving adult-like parsing behavior, but further data are required to address how cognitive control plays a role in the incremental processing of linguistic information. Recent research has begun to investigate the dynamic interplay between executive functioning and syntactic parsing. Hsu and Novick (2016) reported that a Stroop task preceding the syntactic disambiguation task helped adult listeners to recover from garden-paths in understanding *Put* sentences like those tested here. Studies have also started to investigate how cognitive-control engagement affected sentence processing in 5-year-olds (Ovans, Novick, & Huang, 2018).

A limitation of the individual differences component of this study is that we chose only a single executive function task, albeit with two measures in Experiment 2, rather than a battery of tasks. This choice was made for practical reasons, including a desire to keep the tasks short for kindergartners. However, these tasks were chosen based on prior results and theoretical considerations. Previous work measured executive function using similar tasks, in predicting individual differences in children's ability to take the speaker's perspective during sentence comprehension (e.g., Nilsen & Graham, 2009) and in successful interpretation of garden-path sentences (e.g., Woodard et al., 2016). Simon Says and the Flanker task have also been recommended for assessing switching and inhibition, two of the three aspects of a dominant executive functioning model (Miyake et al., 2000; Miyake & Friedman, 2012) in children (Carlson, 2005; Diamond, 2013; Lee, Bull, & Ho, 2013). To play Simon Says, children must inhibit a prepotent response to imitate the experimenter's action whenever they do not hear "Simon Says"; they must also switch flexibly between inhibition and imitation trials. Therefore, both switching and inhibition contribute to accuracy in the Simon Says task. In contrast, the switch cost in the final mixed block of the Flanker task is a relatively independent measure of switching. Children must switch flexibly between two familiar rules, attending to either the middle or the surrounding stimuli. Thus, our findings provided some preliminary evidence for the role of executive functioning, in particular switching and inhibition, in syntactic disambiguation. However, a fuller understanding of how distinct executive constructs (inhibition, switching, working memory, and etc.) influence sentence processing will certainly require much further investigation. For example, better working memory has been found to be related to more successful linguistic ambiguity resolution in both children (Khanna & Boland, 2010) and adults (Novick et al., 2008; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014). However, whether working memory is an independent construct distinct from inhibition and switching in young children is still debated (Hughes, Ensor, Wilson, & Graham, 2010; Lee et al., 2013; St. Clair-Thompson & Gathercole, 2006; Wiebe, Espy, & Charak, 2008).

We acknowledge that mixed results have been reported in the kindergarten parsing literature regarding the relationship between executive functioning and parsing. As a result, there is little theoretical clarity regarding how general-purpose cognitive control mechanisms participate in sentence processing. One issue is measurement problems in this population. The reliability of individual-difference tasks is of paramount importance for investigations of individual differences. In the current study, the reliability of the two executive functioning tasks and the action measure of the sentence task were all above 0.8, measured by Cronbach's alpha, suggesting relatively high reliability. Another issue could be that this relationship has been wobbly in the literature because a third variable, such as motivation, task engagement, or social economic status, plays a more important role (e.g., Huang & Hollister, 2019). All of these considerations point to the necessity of more thorough assessment of the role of domain-general cognitive control in syntactic parsing. All we would suggest, based on our findings, is that we should keep these executive functioning measures in the mix as we continue to assess the proposal that domain-general cognitive control participates in kindergarten parsing.

In conclusion, we found evidence that 5-year-olds can use pragmatic inferences about what it makes sense to say in the current referential context to guide online processing of temporary syntactic ambiguity—at least if given a little more time to do so. This effect emerged somewhat differently in two experiments that had different materials and different overall error rates: In Experiment 1, we saw the strongest evidence of use of the referential principle (the critical interaction between referential context and ambiguity) in measures of children's eye-gaze; in Experiment 2, this interaction effect emerged in analyses of the accuracy of children's offline interpretations. These findings reinforce several previous hints of children's sensitivity to the referential context (Huang & Snedeker, 2013; Sekerina & Trueswell, 2012; Snedeker & Trueswell, 2004; Weighall & Altmann, 2011), and yield new evidence of 5-year-olds' sensitivity to the referential principle in parsing under ambiguity. This sensitivity appears later in children's than in adults' sentence processing, and is fragile, but can nonetheless be seen in the right contexts. Our study also provided new evidence from two independent samples that executive functioning predicts children's syntactic ambiguity resolution. However, the effects were relatively modest and did not survive corrections for multiple comparisons.

References

Altmann, G. T. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences, 2,* 146–152. http://dx.doi.org/10.1016/S1364-6613(98)01153-X

Anderson, S. E., Farmer, T. A., Goldstein, M., Schwade, J., & Spivey, M. (2011). Individual differences in measures of linguistic experience account for variability in the sentence processing skill of five-year-olds. In I. Arnon & E. V. Clark (Eds.), *Experience, variation, and generalization: Learning a first language* (pp. 203–221). Amsterdam, the Netherlands: H. John Benjamins.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition, 118,* 84–93. http://dx.doi.org/10.1016/j.cognition.2010.10.010

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59,* 457–474. http://dx.doi.org/10.1016/j.jml.2007.09.002

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Lme4: linear mixed-effects models using eigen and S4. *Journal of Statistical Software, 67,* 1–48.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33,* 251–283. http://dx.doi.org/10.1006/jmla.1994.1013

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science, 32,* 643–684. http://dx.doi.org/10.1080/03640210802066816

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28,* 595–616. http://dx.doi.org/10.1207/s15326942dn2802_3

Cho, S. J., Brown-Schmidt, S., & Lee, W. Y. (2018). Autoregressive generalized linear mixed effect models with crossed random effects: An application to intensive binary time series eye-tracking data. *Psychometrika, 83,* 751–771. http://dx.doi.org/10.1007/s11336-018-9604-2

Chugani, H. T., & Phelps, M. E. (1986). Maturational changes in cerebral function in infants determined by 18FDG positron emission tomography. *Science, 231,* 840–843. http://dx.doi.org/10.1126/science.3945811

Cimpian, A., Meltzer, T. J., & Markman, E. M. (2011). Preschoolers' use of morphosyntactic cues to identify generic sentences: Indefinite singular noun phrases, tense, and aspect. *Child Development, 82,* 1561–1578. http://dx.doi.org/10.1111/j.1467-8624.2011.01615.x

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives (Studies in Natural Language Processing,* pp.320–358). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511597855.011

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334. http://dx.doi.org/10.1007/BF02310555

Dagerman, K. S., Macdonald, M. C., & Harm, M. W. (2006). Aging and the use of context in ambiguity resolution: Complex changes from simple slowing. *Cognitive Science, 30,* 311–345. http://dx.doi.org/10.1207/s15516709cog0000_46

Dale, P. S., Loftus, E. F., & Rathbun, L. (1978). The influence of the form of the question on the eyewitness testimony of preschool children. *Journal of Psycholinguistic Research, 7,* 269–277. http://dx.doi.org/10.1007/BF01068110

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44,* 2037–2078. http://dx.doi.org/10.1016/j.neuropsychologia.2006.02.006

Davies, M. (2009). The 385+ million word corpus of contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14,* 159–190. http://dx.doi.org/10.1075/ijcl.14.2.02dav

de Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition, 163,* 67–79. http://dx.doi.org/10.1016/j.cognition.2017.02.018

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321. http://dx.doi.org/10.1037/0033-295X.93.3.283

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition, 11,* 159–184. http://dx.doi.org/10.1016/0010-0277(82)90024-5

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168. http://dx.doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed response task. *Developmental Psychobiology, 22,* 271–294. http://dx.doi.org/10.1002/dev.420220307

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Engelhardt, P., Bailey, K., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of quantity? *Journal of Memory and Language, 54,* 554–573. http://dx.doi.org/10.1016/j.jml.2005.12.009

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14,* 340–347. http://dx.doi.org/10.1162/089892902317361886

Fernald, A., Marchman, V. A., & Hurtado, N. (2008). Input affects uptake: How early language experience influences processing efficiency and vocabulary learning. *2008 7th IEEE International Conference on Development and Learning* (pp. 37–42). Monterey, CA: IEEE. http://dx.doi.org/10.1109/DEVLRN.2008.4640802

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42,* 98–116. http://dx.doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9,* 228–231. http://dx.doi.org/10.1111/1467-9280.00044

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *WIREs Cognitive Science, 1,* 143–149. http://dx.doi.org/10.1002/wcs.17

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1,* 3–55. http://dx.doi.org/10.1207/s15327817la0101_2

Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering, and referential overspecification. *Topics in Cognitive Science, 4,* 269–289. http://dx.doi.org/10.1111/j.1756-8765.2012.01186.x

Hadley, P., & Rice, M. (1993). Parental judgments of preschoolers' speech and language development: A resource for assessment and IEP planning. *Seminars in Speech and Language, 14,* 278–288. http://dx.doi.org/10.1055/s-2008-1064177

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York, NY: Wiley.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science, 27,* 572–582. http://dx.doi.org/10.1177/0956797615625223

Huang, Y. T., & Hollister, E. (2019). Developmental parsing and linguistic knowledge: Reexamining the role of cognitive control in the kindergarten path effect. *Journal of Experimental Child Psychology, 184,* 210–219. http://dx.doi.org/10.1016/j.jecp.2019.04.005

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology, 45,* 1723–1739. http://dx.doi.org/10.1037/a0016704

Huang, Y. T., & Snedeker, J. (2013). The use of lexical and referential cues in children's online interpretation of adjectives. *Developmental Psychology, 49,* 1090–1102. http://dx.doi.org/10.1037/a0029477

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35,* 20–36. http://dx.doi.org/10.1080/87565640903325691

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research, 29,* 597–626. http://dx.doi.org/10.1023/A:1026468209238

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *The Journal of Comparative Neurology, 387,* 167–178. http://dx.doi.org/10.1002/(SICI)1096-9861(19971020)387:2<167::AID-CNE1>3.0.CO;2-Z

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience, 21,* 2434–2444. http://dx.doi.org/10.1162/jocn.2008.21179

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109,* 490–501. http://dx.doi.org/10.1037/0033-2909.109.3.490

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86,* 199–225. http://dx.doi.org/10.1016/0001-6918(94)90003-5

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 63,* 160–193. http://dx.doi.org/10.1080/17470210902866664

Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language, 38,* 222–234. http://dx.doi.org/10.1017/S0305000909990316

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology, 4,* 25.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50,* 194–231. http://dx.doi.org/10.1016/j.cogpsych.2004.08.002

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development, 84,* 1933–1953. http://dx.doi.org/10.1111/cdev.12096

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703. http://dx.doi.org/10.1037/0033-295X.101.4.676

Mazuka, R., Jincho, N., & Oishi, H. (2009). Development of executive control and language processing. *Language and Linguistics Compass, 3,* 59–89. http://dx.doi.org/10.1111/j.1749-818X.2008.00102.x

Meroni, L., & Crain, S. (2003). On not being led down the kindergarten path. *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 531–544). Somerville, MA: Cascadilla Press.

Meroni, L., & Crain, S. (2011). Children's use of context in ambiguity resolution. In E. Gibson & N. J. Pearlmutter (Eds.), *The processing and acquisition of reference* (pp. 43–64). Cambridge, MA: MIT Press. http://dx.doi.org/10.7551/mitpress/9780262015127.003.0003

Messenger, K., & Fisher, C. (2018). Mistakes weren't made: Three-year-olds' comprehension of novel-verb passives provides evidence for early abstract syntax. *Cognition, 178,* 118–132. http://dx.doi.org/10.1016/j.cognition.2018.05.002

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21,* 8–14. http://dx.doi.org/10.1177/0963721411429458

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. http://dx.doi.org/10.1006/cogp.1999.0734

Morisseau, T., Davies, C., & Matthews, D. (2013). How do 3-and 5-year-olds respond to under-and over-informative utterances? *Journal of Pragmatics, 59,* 26–39. http://dx.doi.org/10.1016/j.pragma.2013.03.007

Munro, S., Chau, C., Gazarian, K., & Diamond, A. (2006, April). *Dramatically larger flanker effects*. Poster presented at the Cognitive Neuroscience Society Annual Meeting, San Francisco, CA.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13,* 329–336. http://dx.doi.org/10.1111/j.0956-7976.2002.00460.x

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology, 42,* 643–655. http://dx.doi.org/10.1037/0012-1649.42.4.643

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58,* 220–249. http://dx.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. http://dx.doi.org/10.1111/j.1467-8624.2012.01762.x

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

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78,* 165–188. http://dx.doi.org/10.1016/S0010-0277(00)00114-1

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience, 29,* 186–217. http://dx.doi.org/10.1080/01690965.2012.758297

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology, 26,* 527–567. http://dx.doi.org/10.1080/02643290903519367

Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition, 107,* 850–903. http://dx.doi.org/10.1016/j.cognition.2007.12.011

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass, 4,* 906–924. http://dx.doi.org/10.1111/j.1749-818X.2010.00244.x

Ovans, Z., Novick, J., & Huang, Y. T. (2018). *Better to be reliable than early: Cognitive-control effects on developmental parsing*. Boston University Conference on Language Development. Boston, MA.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Qi, Z. (2018). *Referential context and executive functioning influence children's resolution of syntactic ambiguity*. Retrieved from https://github.com/zhenghanQ/forestgit

Rabagliati, H., Pylkkänen, L., & Marcus, G. F. (2013). Top-down influence in young children's linguistic ambiguity resolution. *Developmental Psychology, 49,* 1076–1089. http://dx.doi.org/10.1037/a0026918

Rabagliati, H., & Robertson, A. (2017). How do children learn to avoid referential ambiguity? Insights from eye-tracking. *Journal of Memory and Language, 94,* 15–27. http://dx.doi.org/10.1016/j.jml.2016.09.007

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia, 42,* 1029–1040. http://dx.doi.org/10.1016/j.neuropsychologia.2003.12.012

Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language, 74,* 46–76. http://dx.doi.org/10.1016/j.jml.2014.04.003

Sekerina, I. A., & Trueswell, J. C. (2012). Interactive processing of contrastive expressions by Russian children. *First Language, 32,* 63–87. http://dx.doi.org/10.1177/0142723711403981

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition, 153,* 6–18. http://dx.doi.org/10.1016/j.cognition.2016.04.006

Snedeker, J. (2013). Children's sentence processing. In R. van Gompel (Ed.), *Sentence processing* (pp. 189–220). New York, NY: Psychology Press.

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. http://dx.doi.org/10.1016/j.cogpsych.2004.03.001

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58,* 574–608. http://dx.doi.org/10.1016/j.jml.2007.08.001

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45,* 447–481. http://dx.doi.org/10.1016/S0010-0285(02)00503-0

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 59,* 745–759. http://dx.doi.org/10.1080/17470210500162854

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development, 11,* 176–190. http://dx.doi.org/10.1080/15475441.2014.927328

Strommen, E. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon Says." *Child Development, 44,* 849–853. http://dx.doi.org/10.2307/1127737

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. http://dx.doi.org/10.1126/science.7777863

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role* (pp. 103–130). Hillsdale, NJ: Erlbaum.

Tomasello, M. (1999). Having intentions, understanding intentions, and understanding communicative intentions. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 63–75). Mahwah, NJ: Erlbaum.

Trueswell, J., & Gleitman, L. (2004). Children's eye movements during listening: Developmental evidence for a constraint-based theory of sentence processing. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 319–346). London, UK: Psychology Press.

Trueswell, J. C., Papafragou, A., & Choi, Y. (2011). The development of referential and syntactic processes: What develops. In E. A. Gibson and N. Pearlmutter (Eds.), *The processing and acquisition of reference* (pp. 1–47). Cambridge, MA: MIT Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73,* 89–134. http://dx.doi.org/10.1016/S0010-0277(99)00032-3

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In L. Frazier & K. Ranyer (Eds.), *Perspectives on sentence processing* (pp. 155–179). Mahwah, NJ: Erlbaum.

van Hout, A., Harrigan, K., & de Villiers, J. (2010). Asymmetries in the acquisition of definite and indefinite NPs. *Lingua, 120,* 1973–1990. http://dx.doi.org/10.1016/j.lingua.2010.02.006

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology, 99,* 75–95. http://dx.doi.org/10.1016/j.jecp.2007.10.004

Weighall, A. R., & Altmann, G. T. (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *Journal of Child Language, 38,* 579–605. http://dx.doi.org/10.1017/S0305000910000267

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44,* 575–587. http://dx.doi.org/10.1037/0012-1649.44.2.575

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology, 141,* 187–209. http://dx.doi.org/10.1016/j.jecp.2015.08.005

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience and Biobehavioral Reviews, 33,* 1168–1177. http://dx.doi.org/10.1016/j.neubiorev.2009.03.003

Appendix A

Complete Materials for Experiment 1

Stimuli type	Auditory stimuli: Sentence 1	Auditory stimuli: Sentence 2 (1-ref/2-ref)	Board setting: Target animal	Distracter animal (1-ref/2-ref)	Target animal location	Distracter animal location	Destination	Incorrect destination
Target	Put the duck (that's) on the pond into the tent.	Now put the frog/other duck on the leaf.	duck	frog/duck	pond	leaf	tent	pond
Target	Put the duck (that's) on the leaf into the basket.	Now put him onto the pond.	duck	frog/duck	leaf	pond	basket	leaf
Target	Put the frog (that's) on the leaf into the basket.	Now put him back to the pond.	frog	duck/frog	leaf	pond	basket	leaf
Target	Put the frog (that's) on the pond into the tent.	Now put both animals on the leaf.	frog	duck/frog	pond	leaf	tent	pond
Target	Put the cow (that's) on the grass into the house.	Now put the pig/other cow on the grass.	cow	pig/cow	grass	road	house	grass
Target	Put the cow (that's) on the road into the barn.	Now put the pig/other cow over on the road.	cow	pig/cow	road	grass	barn	road
Target	Put the pig (that's) on the grass into the house.	Now put the cow/other pig on top of the house.	pig	cow/pig	grass	road	house	grass
Target	Put the pig (that's) on the road into the barn.	Now put both animals in the middle of the grass.	pig	cow/pig	road	grass	barn	road
Target	Put the monkey (that's) on the rock into the cave.	Now put him on the tree.	monkey	squirrel/monkey	rock	tree	cave	rock
Target	Put the monkey (that's) on the tree into the boat.	Now put the squirrel/other monkey on the tree.	monkey	squirrel/monkey	tree	rock	boat	tree
Target	Put the squirrel (that's) on the rock into the cave.	Now put the monkey/other squirrel on top of the tree.	squirrel	monkey/squirrel	rock	tree	cave	rock
Target	Put the squirrel (that's) on the tree into the boat.	Now put him back on the tree.	squirrel	monkey/squirrel	tree	rock	boat	tree
Filler	Put one butterfly onto the pond.	Now put the kitty and the butterfly/both butterflies near the flowers.	butterfly	kitty/butterfly	flowers	grass	tent	pond
Filler	Put the little kitty on the leaf.	Now put the butterfly/other kitty with the flowers.	kitty	butterfly/kitty	grass	flowers	basket	leaf
Filler	Put the mouse on the scarecrow's head.	Now put the chicken/other mouse on the barn.	mouse	chicken/mouse	scarecrow	fence	barn	road
Filler	Put the green chicken on the grass.	Now put him on the fence.	chicken	mouse/chicken	fence	scarecrow	house	grass
Filler	Put one goat beside the rainbow.	Now put the fox/other goat on the tree.	goat	fox/goat	rainbow	hill	boat	tree
Filler	Put the fox at the bottom of the hill.	Now put the other animal into the cave.	fox	goat/fox	hill	rainbow	cave	rock

(*Appendices continue*)

Appendix B

Complete Materials for Experiment 2

Stimuli type	Auditory stimuli: Sentence 1	Auditory stimuli: Sentence 2 (1-ref/2-ref)	Board setting: Target animal	Board setting: Distracter animal (1-ref/2-ref)	Board setting: Target animal location	Board setting: Distracter animal location	Board setting: Destination	Board setting: Incorrect destination
Related condition								
Target	Put the bear (that's) on the mountain into the stream.	Now put the sheep/other bear on a mountain.	bear	sheep/bear	mountain	grass	stream	mountain
Target	Put the bear (that's) on the grass into the cave.	Now put him back.	bear	sheep/bear	grass	mountain	cave	grass
Target	Put the sheep (that's) on the mountain into the stream.	Now put both animals in the stream.	sheep	bear/sheep	mountain	grass	stream	mountain
Target	Put the sheep (that's) on the grass into the cave.	Now put the bear/other sheep in the cave.	sheep	bear/sheep	grass	mountain	cave	grass
Target	Put the pig (that's) in the barn into the truck.	Now put him with the cow.	pig	cow/pig	pen/barn	mud	truck	barn
Target	Put the pig (that's) in the mud into the cage.	Now put the cow/other pig in some mud.	pig	cow/pig	mud	barn	cage	mud
Target	Put the cow (that's) in the barn into the truck.	Now put the pig/other cow in the truck.	cow	pig/cow	barn	mud	truck	pen/barn
Target	Put the cow (that's) in the mud into the cage.	Now put him back.	cow	pig/cow	mud	barn	cage	mud
Target	Put the snake (that's) on the rock onto the beach.	Now put him back.	snake	eagle/snake	rock	tree	beach	rock
Target	Put the snake (that's) on the tree onto the field.	Now put the eagle/other snake on the field.	snake	eagle/snake	tree	rock	field	tree
Target	Put the eagle (that's) on the rock onto the beach.	Now put the other animal on a rock.	eagle	snake/eagle	rock	tree	beach	rock
Target	Put the eagle (that's) on the tree onto the field.	Now put the snake/the other eagle up in a tree.	eagle	snake/eagle	tree	rock	field	tree
Filler	Put the bear next to the mountain.	Now put the sheep/the other bear next to the bear/him.	bear	sheep/bear	cave	stream	mountain	grass
Filler	Put the sheep beside the grass.	Now put the bear/other sheep next to the cave.	sheep	bear/sheep	mountain	stream	grass	cave
Filler	Put the pig in the center of the pen/barn.	Now put the cow/other pig near the cage.	pig	cow/pig	mud	truck	pen/barn	cage
Filler	Put the cow by the mud.	Now put the pig/other cow on top of the cage.	cow	pig/cow	barn	truck	mud	cage
Filler	Put the eagle above the rock.	Now put the snake/other eagle below the eagle/him.	eagle	snake/eagle	beach	field	rock	tree
Filler	Put the snake at the bottom of the tree.	Now put the snake back.	snake	eagle/snake	beach	field	tree	rock

(*Appendices continue*)

Appendix B (*continued*)

Stimuli type	Auditory stimuli: Sentence 1	Auditory stimuli: Sentence 2 (1-ref/2-ref)	Board setting: Target animal	Distracter animal (1-ref/2-ref)	Target animal location	Distracter animal location	Destination	Incorrect destination
Unrelated condition								
Target	Put the bear (that's) on the paper onto the street.	Now put the sheep/other bear on some paper.	bear	sheep	paper	bread	street	paper
Target	Put the bear (that's) on the bread onto the cape.	Now put him back.	bear	sheep	bread	paper	cape	bread
Target	Put the sheep (that's) on the paper onto the street.	Now put both animals on the street.	sheep	bear	paper	bread	street	paper
Target	Put the sheep (that's) on the bread onto the cape.	Now put the bear/other sheep on the cape.	sheep	bear	bread	paper	cape	bread
Target	Put the pig (that's) on the lamp onto the brick.	Now put him with the cow.	pig	cow	lamp	shoe	brick	lamp
Target	Put the pig (that's) on the shoe onto the cake.	Now put the cow/other pig on a shoe.	pig	cow	shoe	lamp	cake	shoe
Target	Put the cow (that's) on the lamp onto the brick.	Now put the pig/other cow on the brick.	cow	pig	lamp	shoe	brick	lamp
Target	Put the cow (that's) on the shoe onto the cake.	Now put him back.	cow	pig	shoe	lamp	cake	shoe
Target	Put the snake (that's) on the spoon onto the wheel.	Now put him back.	snake	eagle	spoon	chair	wheel	spoon
Target	Put the snake (that's) on the chair onto the feet.	Now put the eagle/other snake on the feet.	snake	eagle	chair	spoon	feet	chair
Target	Put the eagle (that's) on the spoon onto the wheel.	Now put the other animal on a spoon.	eagle	snake	spoon	chair	wheel	spoon
Target	Put the eagle (that's) on the chair onto the feet.	Now put the snake/other eagle up on a chair.	eagle	snake	chair	spoon	feet	chair
Filler	Put the bear next to the paper.	Now put the sheep/other bear next to the bear/him.	bear	sheep	cape	street	paper	bread
Filler	Put the sheep beside the bread.	Now put the bear/other sheep next to the cape.	sheep	bear	paper	street	bread	cape
Filler	Put the pig on the center of the lamp.	Now put the cow/other pig near the cake.	pig	cow	shoe	brick	lamp	cake
Filler	Put the cow by the shoe.	Now put the pig/other cow on top of the cake.	cow	pig	lamp	brick	shoe	cake
Filler	Put the eagle above the spoon.	Now put the snake/other eagle below the eagle/him.	eagle	snake	wheel	feet	spoon	chair
Filler	Put the snake at the bottom of the chair.	Now put the snake back.	snake	eagle	wheel	feet	chair	spoon

Received August 5, 2018
Revision received May 4, 2020
Accepted May 5, 2020 ■