



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Taking your own path: Individual differences in executive function and language processing skills in child learners



Kristina Woodard*, Lucia Pozzan, John C. Trueswell

Department of Psychology and Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:

Received 11 May 2015

Revised 14 August 2015

Keywords:

Executive function
Cognitive development
Individual differences
Sentence processing
Garden path recovery
Cognitive Flexibility

ABSTRACT

Children as old as 5 or 6 years display selective difficulties in revising initial interpretive commitments, as indicated by both online and offline measures of sentence comprehension. It is likely, however, that individual children differ in how well they can recover from misinterpretations and in the age at which they become adult-like in these abilities. To better understand the cognitive functions that support sentence processing and revision, the current work investigated how individual differences in children's ability to interpret temporarily ambiguous sentences relate to individual differences in other linguistic and domain-general cognitive abilities. Children were tested over 2 days on a battery of executive function, working memory, and language comprehension tasks. Performance on these tasks was then used to predict online and offline measures of children's ability to revise initial misinterpretations of temporarily ambiguous sentences. We found two measures of children's cognitive flexibility to be related to their ambiguity resolution abilities. These results provide converging evidence for the hypothesis that the ability to revise initial interpretive commitments is supported by domain-general executive function abilities, which are highly variable and not fully developed in children.

© 2015 Elsevier Inc. All rights reserved.

* Corresponding author. Fax: +1 215 573 9247.

E-mail address: woodardk@sas.upenn.edu (K. Woodard).

Introduction

Real-time sentence parsing and revision

To interpret spoken language, listeners must rapidly categorize the linguistic input into candidate phonemes, syllables, words, and phrases and assign it a provisional structural analysis and interpretation based on the currently available linguistic and nonlinguistic evidence (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). However, because listeners accomplish much of their interpretation in real time as sentences unfold, they are forced to deal with temporary ambiguity that frequently arises in the input. They must rapidly resolve ambiguities associated with words that have multiple possible meanings and phrases that have multiple possible parses. A natural consequence of real-time interpretation is that initial interpretive commitments can turn out to be incorrect, resulting in the need for processing revision when late-arriving evidence supports a different interpretation. For example, listeners typically experience a “garden path” when hearing a sentence like (1) below. Even though the prepositional phrase (PP) “on the napkin” could serve as a modifier for the preceding noun phrase “the frog,” listeners tend to initially interpret it as a goal of the action (i.e., where to put the frog). Upon hearing the actual goal phrase (“onto the book”), listeners are forced to revise their initial interpretation of “on the napkin,” reanalyzing it as a noun phrase modifier.

(1) Put the frog on the napkin onto the book

Evidence for this processing pattern comes from studies within the visual world paradigm, in which participants’ eye movements and actions are recorded while they hear speech about a visually co-present referent world (e.g., Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). When listeners hear “the frog” in sentence (1) within the context shown in Fig. 1, they typically first shift their gaze to the target referent (the frog sitting on a napkin). After hearing the PP “on the napkin,” they shift their gaze toward the so-called “incorrect goal” (a second “empty” napkin); this suggests that this phrase is being interpreted as the goal of the action. Finally, when hearing “onto the book,” listeners tend to look around the scene, showing some signs of confusion, before carrying out the correct action of moving the frog onto the book. This pattern of looks toward the incorrect goal, and the resulting delay integrating the “correct goal” within listeners’ current interpretation of the sentence, do not occur when listeners instead hear sentence (2), an analogous sentence that contains the same lexical items but does not present a temporary ambiguity. This difference suggests that the processing difficulty associated with sentence (1) is due to an initial mis-parse of “on the napkin” and the resulting revision needed when hearing “onto the book”.

(2) Put the frog that’s on the napkin onto the book

Strikingly, children as old as 5 or 6 years display selective difficulties in revising initial interpretive commitments during sentence comprehension (see Trueswell, Sekerina, Hill, & Logrip, 1999, for the first report of the so-called *kindergarten-path* effect, subsequently replicated in a number of studies, e.g., Anderson, Farmer, Goldstein, Schwade, & Spivey, 2011; Choi & Trueswell, 2010; Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Weighall, 2008). In response to temporarily ambiguous sentences like (1) above, children perform an incorrect action on approximately 50% of the trials. These errors almost exclusively involve moving the frog onto the empty napkin, suggesting a failure to revise an initial goal interpretation for the PP “on the napkin” even after hearing unambiguous evidence against this interpretation (i.e., once the second PP has been heard). Children’s virtually error-free performance in response to the corresponding unambiguous sentence (2) indicates that act-out errors associated with ambiguous sentences like (1) do not stem from generalized difficulties with complex sentences, but from selective difficulties in revising initial misinterpretations. Importantly, children’s difficulties in recovering from garden path sentences have also been documented in other languages and other structures (Choi & Trueswell, 2010; Omaki, Davidson White, Goro, Lidz, & Phillips, 2014), suggesting that it is a fundamental characteristic of the developing parser.

It is important to note, however, that individual children likely differ from each other in how well they can recover from initial misinterpretations and might also differ regarding when their sentence

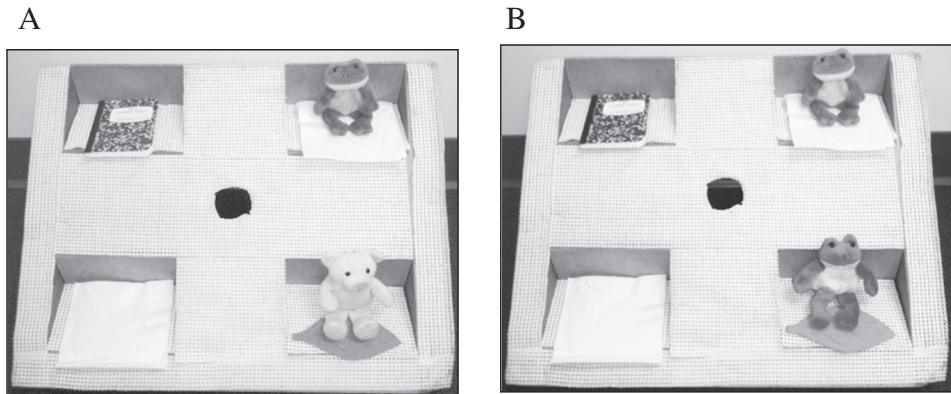


Fig. 1. Referential context for “Put the frog [that’s] on the napkin onto the book” in the 1-referent (A) and 2-referent (B) contexts.

parsing abilities become adult-like. The purpose of the current study was to explore these differences in some detail and relate them to other linguistic and cognitive abilities in hopes of gaining a better understanding of the cognitive functions that support garden path recovery. We take as our starting point the proposal found in [Novick, Trueswell, and Thompson-Schill \(2005\)](#), who suggested that children’s parsing difficulties may stem from their well-documented difficulties with domain-general executive function (EF) abilities, including mental set shifting, information updating and monitoring, and prepotent response inhibition (e.g., [Mazuka, Jincho, & Onishi, 2009](#); [Miyake et al., 2000](#)).

One aspect of EF that is particularly relevant to garden path recovery is the ability to successfully handle situations of “representational conflict,” that is, situations in which conflict arises between a habitual interpretation of a stimulus and current task demands. Under the EF account of garden path recovery (e.g., [Novick et al., 2005](#)), in a sentence like (1), late-arriving linguistic evidence (i.e., the prepositional phrase “onto the book”) provides a way of categorizing earlier linguistic material (i.e., the prepositional phrase “on the napkin” as a nominal modifier) that is in conflict with the initially preferred way of categorizing the same material (i.e., the prepositional phrase “on the napkin” as a verbal modifier). Thus, inhibitory control and conflict monitoring/resolution components of EF would be engaged when the sentence processing systems need to abandon a preferred way of analyzing a sentence in favor of an initially dispreferred one.

Below we report evidence that children’s difficulty in revising interpretive commitments in response to garden path sentences like (1) as compared to (2) is related to independent nonverbal measures of EF abilities and, in particular, their ability to rapidly and efficiently adapt to changing circumstances and demands, suggesting a connection between domain-general cognitive abilities and garden path recovery. This hypothesized link between EF skills and language processing has potentially important implications for grammar learning, which we expand on in the Discussion. Before turning to the results of our investigation, we review what is currently known about the role of EF during garden path recovery from a wide-ranging body of research that includes (a) correlational and training studies in adult native speakers, (b) adult neuroimaging and neuropsychological studies, and (c) studies investigating garden path resolution in populations with underdeveloped EF skills (children).

EF skills and garden path recovery: Correlational evidence in adult native speakers

Studies of individual differences in healthy adult participants have shown that nonsyntactic conflict monitoring EF abilities, as measured by (verbal) Stroop tasks, correlate with individuals’ (varying) abilities to revise misinterpretations in sentence processing reading tasks ([Mendelsohn, 2002](#); [Vuong & Martin, 2014](#)). In addition, consistent with the existence of a causal relation between EF skills and garden path recovery, training of EF skills has been found to be selectively linked to improvements in

garden path recovery in adult native speakers. A recent study by Novick, Hussey, Teubner-Rhodes, Harbison, and Bunting (2014) found that training-related improvements on an *N*-back task with lures (an EF measure targeting conflict resolution) positively correlated with improvements in processing garden path sentences, as measured by eye-tracking and offline accuracy scores.

EF skills and garden path recovery: Neuroimaging and neuropsychological evidence

Support for the engagement of aspects of EF during garden path recovery in adults also comes from neuroimaging and neuropsychological case studies. The left inferior frontal gyrus (LIFG) is believed to be the locus of some aspects of cognitive control and conflict resolution, in that it has been found to be engaged during conflict trials of the Stroop task (e.g., Milham et al., 2001) and other conflict resolution tasks such as flanker (Ye & Zhou, 2009), proactive interference (e.g., Jonides, Smith, Marshuetz, Koeppe, & Reuter-Lorenz, 1998), and verbal fluency (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997).

Similar to the nonsyntactic conflict-monitoring tasks above, the resolution of syntactic conflict during garden path processing has also been found to engage the LIFG in functional magnetic resonance imaging (fMRI) studies (e.g., January, Trueswell, & Thompson-Schill, 2009; Mason, Just, Keller, & Carpenter, 2003). For instance, voxel-by-voxel activity of the LIFG within individuals during the conflict trials of the Stroop task was found to co-localize with syntactic ambiguity resolution during spoken sentence comprehension (January et al., 2009). Consistent with these findings, focal damage to the LIFG has been shown to be associated with a striking inability to recover from garden paths in a patient with a focal lesion to LIFG (Novick, Kan, Trueswell, & Thompson-Schill, 2009). In response to ambiguous sentences like (1), this patient made substantial act-out errors, e.g., moving the frog onto the napkin (the incorrect goal) and then onto the book (the correct goal). These actions suggest that, similarly to healthy adults and children, the patient initially interpreted “on the napkin” as the goal of the action but, differently from adults and similarly to children, he was never able to revise this initial processing commitment. In line with the hypothesis that this patient's difficulties were related to revision of initial parsing commitments, his performance was error free in response to unambiguous sentences like (2).

EF skills and garden path recovery: Evidence from children

The existence of a link between EF and garden path recovery has been hypothesized based on children's documented difficulties in both domains (e.g., Choi & Trueswell, 2010; Mazuka et al., 2009; Novick et al., 2005) and the finding that for adults both EF and garden path recovery tasks engage cortical areas within the prefrontal cortex (e.g., the LIFG)—cortical areas that are still developing in children. It is in fact well known that typically developing children between 3 and 6 years of age show high levels of interference and processing difficulty on a range of EF tasks such as the Day/Night task, in which children are asked to refer to an object by its antonym (i.e., calling the day “night” and vice versa; Gerstadt, Hong, & Diamond, 1994) and Stroop tasks (e.g., Carter, Mintun, & Cohen, 1995). Because the neurodevelopment of the frontal lobes is a prolonged process that continues throughout childhood, children's difficulties on these tasks have been hypothesized to stem from the underdevelopment of the frontal lobe systems. Thus, children's delays and subsequent age-related improvements in many EF tasks (e.g., Anderson, 2002; Diamond, Kirkham, & Amso, 2002) might reflect the naturally occurring protracted development of the frontal lobes (e.g., Huttenlocher & Dabholkar, 1997). Children's delays in EF abilities have been connected to some language abilities, most notably homophone processing (Khanna & Boland, 2010) and perspective taking (Nilsen & Graham, 2009).

To date, however, little research has examined individual differences in children's ability to recover from garden path sentences and how this ability relates to cognitive and linguistic measures.¹ Because of the temporal coincidence between children's protracted development of the frontal

¹ In preliminary unpublished work, Qi, Fisher, and Brown-Schmidt (2011) reported that children's executive function abilities as measured by performance on the Simon Says task correlated with individual differences in garden path recovery. Anderson and colleagues (2011) found that children's linguistic abilities, as measured by the Peabody Picture Vocabulary Test (PPVT), correlated with children's understanding of temporarily ambiguous sentences in pragmatically supportive contexts.

lobe and delays and subsequent age-related improvements in the processing of temporarily ambiguous sentences, the current study aims to test behaviorally the proposed link between individual differences in EF skills and garden path recovery in young children. Finding such a connection is not a guaranteed outcome because children and adults differ in many other ways, including vocabulary knowledge, working memory, general linguistic experience, and literacy, raising the possibility that children's difficulty with garden path recovery arises for reasons other than developmental delays in EF. Finding a clear absence of such a relationship between EF and garden path recovery would require a rethinking of our current understanding of child garden path recovery and child language processing abilities more generally.

Experimental prospectus

The current study aimed to document the relationship between EF skills and children's (in)ability to revise garden path sentences by examining the extent to which children's performance on temporarily ambiguous and unambiguous structures is correlated with different components of EF skills (e.g., conflict monitoring, inhibitory control skills). To assess EF skills, we used (a) a Flexible Item Selection Task or "Card Sort" game (Jacques & Zelazo, 2001), which is hypothesized to be an index of cognitive flexibility, because children need to sort cards along varying dimensions, and of inhibitory control, because children are asked to sort the same cards twice, thereby needing to inhibit an initial sorting rule in favor of an additional one; (b) a variation of the Day/Night task, which is thought to tap into children's ability to inhibit a prepotent verbal response; and (c) a combined Flanker/No-Go task, which is a measure of children's ability to disregard irrelevant information (Flanker), inhibit a prepotent response (No-Go), and flexibly switch between different rules of responding (i.e., different responses to Flanker and No-Go trials). We hypothesized that increased costs associated with the conflict conditions in these EF tasks (as compared with no-conflict conditions) should be related to increased processing costs for temporarily ambiguous garden path sentences like (1) as compared with their unambiguous counterparts like (2).

Nevertheless, differences in EF skills are unlikely to explain all of the differences between children regarding their processing difficulties associated with garden path sentences. For example, verbal working memory resources (e.g., Just & Carpenter, 1992; Miyake, Carpenter, & Just, 1994) and general language abilities may also contribute. For this reason, we collected two additional individual difference measures designed to examine these factors. We used an object memory task modeled on Nilsen and Graham (2009) to assess verbal working memory resources and a standardized receptive grammar task (Test for Reception of Grammar-2) to assess language abilities more generally.

Finally, it is also possible that children's inability to revise in the "Put" task is related to their linguistic understanding of the definite determiner "the" (see Trueswell, Papafragou, & Choi, 2011; Wexler, 2011). This is because in the Put task children typically perform quite poorly in response to temporarily ambiguous sentences even when they are presented in referentially supportive contexts that are known to help adults arrive at the correct interpretation more quickly. Specifically, in a visual context containing two frogs—one on a napkin and the other on a plate (see Fig. 1B)—the intended modifier interpretation of "on the napkin" for sentence (1) becomes more plausible because a modifier phrase is needed to individuate the target frog. In line with this, in 2-referent contexts adults show fewer looks to the incorrect goal and fewer signs of processing difficulty (e.g., Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). However, children show just as many errors in 2-referent and 1-referent contexts (Trueswell et al., 1999), as if they failed to understand that a definite noun phrase ("the frog") requires a unique referent and, thus, did not realize that the prepositional phrase ("on the napkin") is likely to be a noun phrase modifier in 2-referent contexts. For these reasons, we included both 1-referent and 2-referent contexts in our "Put" task; in addition, to determine whether children's performance on 2-referent context is modulated by their knowledge of definite determiners, we attempted to independently assess each child's understanding of definite determiners using a novel (never before used) determiner "discrimination task" described below.

Method

Participants

A total of 40 monolingual, typically developing children participated from area preschools (18 female; mean age = 4;10 [years;months], range = 4;0–5;9). Parents gave informed consent in accordance with institutional review board protocol. Children were compensated with a small toy.

Procedure

Over two sessions occurring within 2 weeks of each other, children completed seven tasks (each described below): the “Put” task, the Test for Reception of Grammar (TROG-2; Bishop, 2003), three EF tasks (Flexible Item Selection Task or Card Sort, Day/Night, and Flanker/No-Go), one Verbal Working Memory task, and the Determiner task. The “Put” task always occurred at the beginning of Session 1, and TROG-2 always occurred at the beginning of Session 2. The order of the remaining tasks was counterbalanced across participants and sessions. Sessions were video- and audio-recorded for offline scoring.

Materials and design

Put task and materials

Each child was tested individually in a quiet room. The child sat in front of a table with an inclined podium. At the center of the podium was a hole for a camera focused on the child's eyes to record direction of gaze. At the beginning of each trial, the experimenter placed objects in each of the quadrants of the podium. The child was asked to name the objects, with the experimenter correcting “incorrect” names. The child was then asked to look at the center of the podium and carry out prerecorded instructions played from a computer. A second camera, pointed at the podium, recorded the child's actions.

Target trials consisted of sentences with and without a temporary ambiguity that concerned a prepositional phrase. The target structure for the experimental sentences always required the PP to modify the noun phrase object (see sentence (1) repeated below for ease of reference). Unambiguous sentences were the same except that the ambiguity was avoided due to the presence of the relativizer “that's” (see sentence (2) below). The referential context of each target sentence was also manipulated. In half of the trials a modifier interpretation was supported by presenting two animals of the same type (e.g., two frogs: 2-referent condition; see Fig. 1B), whereas the other half did not encourage a modifier interpretation (e.g., one frog and one pig: 1-referent condition; see Fig. 1A):

- (1) Put the frog on the napkin onto the book.
- (2) Put the frog *that's* on the napkin onto the book.

In the example scenes shown in Fig. 1, the *target* animal is the frog on a napkin, the *goal* is the book, the *incorrect goal* is the “empty” napkin, and the *competitor* is the other animal on a platform (a pig on a leaf for the 1-referent condition [Fig. 1A] or a frog on a leaf for the 2-referent condition [Fig. 1B]). Position of the target, competitor, goal, and incorrect goal were counterbalanced within and across lists.

Each trial consisted of two sentences: a target or filler sentence followed by a continuation that was always a filler sentence (e.g., “Now make the animals hug each other”). There were 16 target and four filler trials. An experimental list was created such that the 16 target trials were divided equally among the four conditions: 1-referent ambiguous, 2-referent ambiguous, 1-referent unambiguous, and 2-referent unambiguous. Target and filler trials appeared in a fixed random order. Three additional lists were created by rotating the items through the conditions so that each target would be seen in each of the four conditions by different participants. Four additional lists were created in which trial order was reversed. Each participant was assigned to one of the eight lists.

EF tasks

Three age-appropriate tasks were selected to investigate children's EF skills. One task was selected to examine children's cognitive flexibility (Flexible Item Selection Task/Card Sort), one to examine children's ability to override a prepotent response (Day/Night task), and one to examine children's ability to focus on relevant information while ignoring irrelevant information (Flanker portion of the Flanker/No-Go task), inhibit prepotent responses (No-Go portion of Flanker/No-Go task), and flexibly switch between different trial types and responses (Flanker/No-Go task).

Flexible Item Selection Task/Card Sort. In this task (Jacques & Zelazo, 2001) children were given sets of three cards (from the game "Set") that could be sorted along two dimensions (e.g., color and shape). At the start of each trial, the experimenter asked children to "show two cards that go together." The experimenter followed up on children's responses by asking them what the selected cards had in common. Next, children were asked to "show two cards that could go together in a different way" and again were prompted to state what the two new cards had in common. One card in each set could always be sorted along two different dimensions, whereas the other two cards could be sorted along only one dimension. Overall, cards could be sorted along four different dimensions: color (green, red, or purple), shape (oval, squiggle, or diamond), number (one, two, or three), and filling pattern (solid, empty, or striped). Children completed two practice trials with feedback and six test trials. In the practice trials, children sorted along each of the four dimensions once.

Children received one point for selecting the correct pair of cards and another point for correctly stating why the two cards were paired together, for a total of four possible points per trial (two points for each sort). To correctly sort the cards a first time, children need to be able to notice abstract commonalities among the different objects (e.g., that two objects are both green even if they have different shapes and filling patterns). To correctly sort the cards a second time, children need to consider an additional abstract dimension of commonality, thereby possibly inhibiting a more salient dimension (e.g., color) in favor of a less salient one (e.g., number). Because we were interested in measuring children's cognitive flexibility, children's responses were scored based on their performance on the second sort, relative to their performance on the first sort.

Day/Night task. To investigate children's ability to override a prepotent response (here defined as the ability to override an automatic response), children completed the Day/Night task (Gerstadt et al., 1994). Children saw and were asked to name pairs of images (e.g., "boy," "girl," "day," "night"). Next, children were told that the objects "went to sleep and woke up on *opposite day*." On opposite day, objects were to be called by their antonyms (e.g., "boy" for girl, "day" for night, etc.). At the start of opposite day, the experimenter asked children to name each object by its antonym. If children failed to do so, the experimenter provided the antonym and children were asked to repeat it. For testing, the objects were randomized and children needed to name each of the objects by its antonym. Children's accuracy was based on their first intelligible response, ignoring any self-corrections offered by children.

Our version of the task included four Stroop-like pairs: day versus night, big versus small, up versus down, and girl versus boy (Livesey, Keen, Rouse, & White, 2006). Originally, the task contained only eight trials (one response for each object) and objects were printed out on cards; however, because of a ceiling effect, the task was modified to include more trials, so that each object appeared three times, each presented individually on a computer screen via E-Prime 2.0. Presentation order was randomized for each child. For the analyses involving this measure, we either (a) excluded the 15 children who participated in the earlier simpler version of this task or (b) included all participants for the analyses. Exclusion/inclusion did not affect the overall results.

Flanker/No-Go task. To examine children's inhibitory control, their ability to disregard irrelevant information, and their ability to switch between different trial types and demands, children were tested on a variant of the Flanker task (Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012; Rueda et al., 2004) that included a No-Go task. Children were told that they were going to play a game where they would "feed hungry fish." Fish appeared in a row in the center of the screen (Fig. 2), and children were told to feed the middle fish by pressing a left or a right key on the keyboard, depending on the

direction of the middle fish. Stickers with fish swimming left and right appeared on the “A” and “K” keys of the keyboard. When fishbowls appeared on the screen, children were not to press any keys (No-Go condition).

Children completed 12 practice trials. In the first six, only one fish/fishbowl was present in the middle of the screen; this block was used to teach children to respond using the keyboard. In the second block of six practice trials, multiple fish/fishbowls were present on the screen at the same time and children were taught to respond to the direction of the middle fish; to draw children's attention to the middle fish, the middle fish was highlighted using a red line under it. Feedback was provided throughout practice. Experimental trials were identical to the second block of practice trials except that the middle fish was not underlined. There were 36 trials involving three conditions. In the congruent condition (12 trials), the flanking and middle fish faced the same direction. In the incongruent condition (12 trials), the middle fish faced the direction opposite of the flanker fish. In the No-Go condition (12 trials), the center fish was surrounded by fishbowls and children were instructed to refrain from responding. Across trials, the fish faced left and right in equal proportions. The three conditions were randomly intermixed, with each child receiving a different randomization.

Each trial was preceded by a 1000-ms fixation cross in the center of the screen. Each trial lasted until children responded, up to a maximum display time of 5000 ms. Reaction times (RTs) and accuracy were recorded. The task was administered via E-Prime 2.0.

Verbal Working Memory task

Children completed a variation of a Verbal Working Memory task from [Nilsen and Graham \(2009\)](#). Children named pictures in a row on the computer screen (e.g., pig–ball–shoe) and were told to remember the last picture in each row (e.g., shoe). Each row contained three pictures, with the final picture in each row highlighted with a yellow box. Rows appeared incrementally so that children initially saw one row, then two rows, and so forth up to three or four rows. After children named all of the images on the screen, the pictures disappeared and children needed to recall the last item(s) in the row(s) from memory. Children completed four sets of three rows (12 trials) and one set of four rows (four trials) for a total of 16 trials. Accuracy was computed based on how many items children could recall out of the 16 total test trials.

TROG-2: Test for Reception of Grammar

Children were administered the TROG-2 ([Bishop, 2003](#)). Children were read sentences and instructed to point to the picture that the experimenter was describing by choosing among four different pictures in the TROG-2 picture book. Children completed a subset of the TROG-2: two practice trials and 13 experimental blocks composed of four trials each. The grammatical constructions tested in the 13 blocks included negation (e.g., “The pen is not only red but also long”), conjunction (e.g., “The horse sees the cup and the book,” “The man is looking at the horse and is running”), singular and plural morphology (e.g., “The boy is picking flowers,” “The cows are under the tree”), pronoun gender and number (e.g., “They are carrying him”), reversible passives (e.g., “The cow is chased by the girl”), subject and object relative clauses (e.g., “The man that is eating looks at the cat”; right branching: “The girl chases the dog that is jumping”; center-embedded: “The dog the girl chases is jumping”), and reduced relative clauses (e.g., “The elephant pushing the boy is big”). Children's first response was coded for accuracy, ignoring self-corrections. Accuracy was computed based on children's performance on the selected TROG-2 subset. For consistency with other tasks used in this study, TROG-2 scores were not age-corrected.

Determiner task

To assess children's ability to discriminate between definite and indefinite determiners, we designed a task in which children saw two types of imaginary creatures (e.g., three yellow triangles with red hair and three blue circles with green shoes) paired with two novel labels (“Look, these are heefs and glorps!”). Children were not told which was which, that is, how the creatures and labels were paired together. After the creatures had been introduced, children were presented with only one creature of one type and more than one creature of the other type and were asked to choose the creature that the experimenter was talking about (e.g., “the glorp,” “a glorp”). The rationale was that

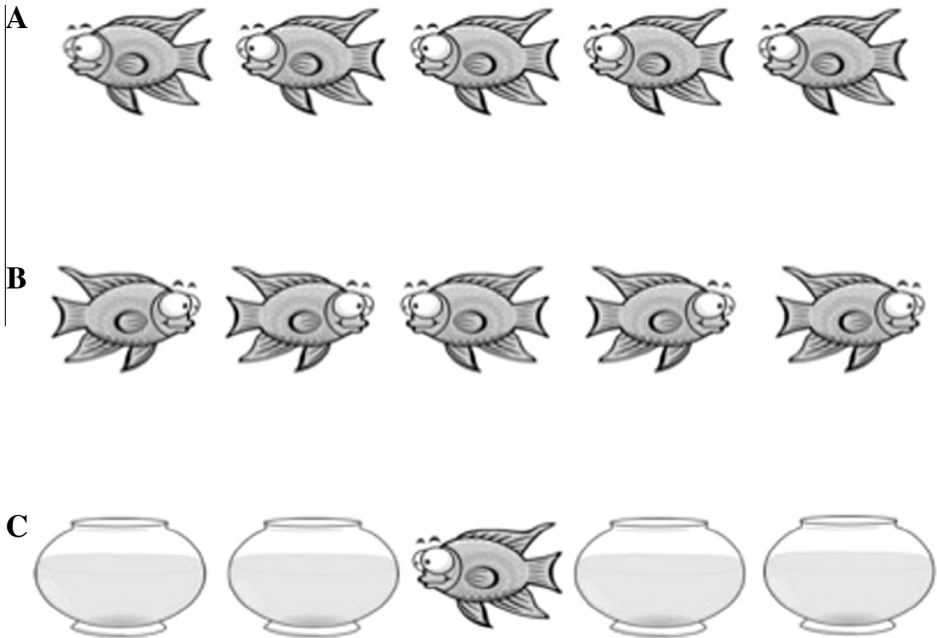


Fig. 2. Example of the three trial types in the Flanker/No-Go task: (A) congruent trial; (B) incongruent trial; (C) No-Go trial.

children would use the determiner to infer the target creature (i.e., the unique creature if the experimenter used the definite determiner; either creature, possibly with a preference for the nonunique creature, if the experimenter used the indefinite determiner).

Each trial (eight in total) was manipulated so that the test picture always included one unique creature of one type and more than one unique creature for the other type. Half of the trials contained the definite determiner “the” and half contained the indefinite determiner “a”. Each item was associated with both determiners across two lists, each with the same fixed random order. As a measure of discrimination between determiners, we took the proportion of unique responses given a definite determiner minus the proportion of unique responses given an indefinite determiner.²

Results

We first summarize children’s performance on each of the seven tasks and then present the central findings on the relation between children’s garden path recovery and EF, working memory and linguistic abilities.

Performance on “Put” task

Here children often failed to revise their initial parsing commitments, replicating key findings of Trueswell and colleagues (1999). Evidence comes from our analyses of both the actions carried out by children and their eye movement patterns.

Actions

Children’s actions were coded following Trueswell and colleagues’ (1999) coding scheme. Actions were coded as “correct” (adult-like) if the target (i.e., the frog sitting on the napkin in Fig. 1) was

² The Determiner task was piloted on adult participants ($N = 21$) and found to be an effective measure of determiner discrimination: average difference = .36, $SD = .33$, $t(20) = 4.99$, $p < .001$, $d = 2.23$.

moved directly to the correct goal (the book). Similar to previous findings, incorrect actions were relatively common and almost always involved moving an animal to the incorrect goal (the empty napkin), with these errors disproportionately occurring in the ambiguous rather than the unambiguous condition (see Fig. 3).³ Such a pattern suggests that children frequently misinterpreted the first PP (“on the napkin”) as the goal argument of the verb “put” and failed to revise their initial interpretation, even after evidence incompatible with the initial analysis became available (e.g., “onto the book”). Also in line with prior results, error rates did not differ across the two referential contexts, suggesting that children did not use referential context to refrain from a goal interpretation of the post-nominal PP (see Table 1).

These observations found support in a statistical analysis of the non-adult-like (“incorrect”) action patterns; test trials were coded as either 1 (*incorrect*) or 0 (*correct*) and were analyzed using a mixed-effects logistic regression model with ambiguity (ambiguous vs. unambiguous), referential context (1-referent vs. 2-referent), and their interaction as fixed effects; the random effect structure for the model included by-item and by-participant random intercepts and by-item random slopes for the effect of referential context.⁴ A summary of the converging maximal model for these data is presented in Table 2; a significant main effect of ambiguity emerged, but no main effect of referential context and no interaction.

In addition to replicating the overall pattern of performance observed by Trueswell and colleagues (1999), we also replicated their original findings with respect to children’s referent selection rates in 1-referent and 2-referent ambiguous trials. Similar to the original study, children were well above chance in selecting the correct referent (e.g., the frog on the napkin) in 1-referent ambiguous trials (94% accuracy) but were at chance in 2-referent ambiguous trials (46% accuracy).

Eye gaze

Eye gaze was analyzed for 37 participants (data for three participants were not available due to equipment error). Trials in which track loss exceeded 40% were excluded from analysis ($n = 47$); in addition, one experimental item that had visually confusable platforms in the 2-referent condition (a lion on a napkin and a lion on a towel, $n = 33$)⁵ was excluded. Altogether, 12% of trials were dropped. Eye gaze was coded manually from the start of the sentence (“Put...”) until the child started to perform an action using E-Lan 4.0 Linguistic Annotator.

Eye movements provided further evidence that children struggled with temporarily ambiguous sentences. Fig. 4 plots by condition the proportions of time spent looking at the correct and incorrect goals from the onset of the correct goal phrase (e.g., “book”) until the child began an action. Because we were interested in revision, we chose this late time window to compare children who, on hearing “onto the book,” are able to disengage from considering the incorrect goal (the empty napkin) and then look toward the correct goal (the book) with children who instead continue to look at the incorrect goal after hearing this phrase. As seen in the figure, children spent more time looking at the correct goal in the unambiguous condition, as compared with the ambiguous condition in both referential contexts. A multi-level linear model performed on e-logit transformed values supported these conclusions (see Table 3); there was a reliable main effect of ambiguity, no main effect of referential context, and no interaction on children’s looks to the correct goal.

³ The current study raises a coding issue with the original Trueswell and colleagues (1999) study concerning the actions involving movement of both the target and its platform (e.g., the frog and the napkin under it) to the correct goal (e.g., the book). Although this coding issue does not affect the pattern of results of the original kindergarten path effect, it influences the magnitude of the effect in this study. The current study codes this action as an error, whereas the original study coded it as correct. The decision to code it as an error was based on the fact that the first prepositional phrase was not unambiguously interpreted as a modifier in this case; in line with this coding decision, this type of action occurred numerically more often in the ambiguous condition than in the unambiguous condition, although this difference was not significant: average difference = .20, $SD = .21$, $t(39) = 0.97$, $p = .34$, $d = 0.31$. In addition, there were eight children who performed this action in the ambiguous condition but not in the unambiguous condition, whereas there was only one child on one trial who performed the action in the unambiguous condition but not in the ambiguous condition. This action is labeled as a “together hop” in Table 1.

⁴ By-participant random slopes for the effects of ambiguity and referential context together with by-item random slopes for the effect of ambiguity were dropped due to convergence issues.

⁵ Fewer than 37 because some trials were already excluded for track loss or equipment error.

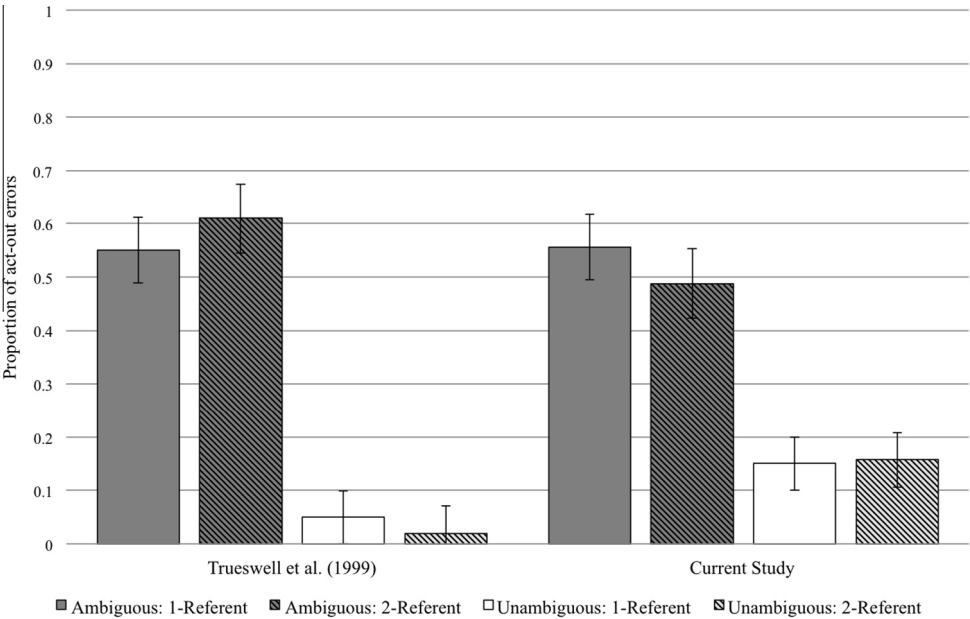


Fig. 3. Proportions of act-out errors as a function of ambiguity and referential context compared with Trueswell and colleagues' (1999) study. Error bars represent standard errors.

Table 1
Number of actions on ambiguous trials per action type.

Action type	Object that was moved to the		Type of context		Total
	Incorrect goal	Goal	1-Ref	2-Ref	
Correct	–	Target	70	77	147
Incorrect			93	87	180
Short	Target	–	10	0	10
Short	Other	–	0	7	7
Mixed	Other	Target	2	1	3
Mixed	Target	Other	0	4	4
Long	Target	Target	47	27	74
Long	Other	Other	3	21	24
Together	–	Target (w/platform)	25	15	40
Together	–	Other (w/platform)	0	0	0
Other	–	–	6	12	18

Table 2
Logistic mixed effects model of the effects of ambiguity and referential context on incorrect actions in the Put task.

Fixed effects	Estimate	SE	z-Value	p-Value
Intercept	–1.62	0.53	–3.06	<.01**
Ambiguity	3.86	0.40	9.53	<.001***
Referential context	–.19	.34	–0.56	.57
Ambiguity/referential context	–.75	.59	–1.27	.20

Note. The model above includes the fixed effects of ambiguity, referential context, and their interaction. The maximum random effects structure that converged included by-item and by-participant random intercepts and by-item random slopes for the effect of referential context.

** $p < .01$.
*** $p < .001$.

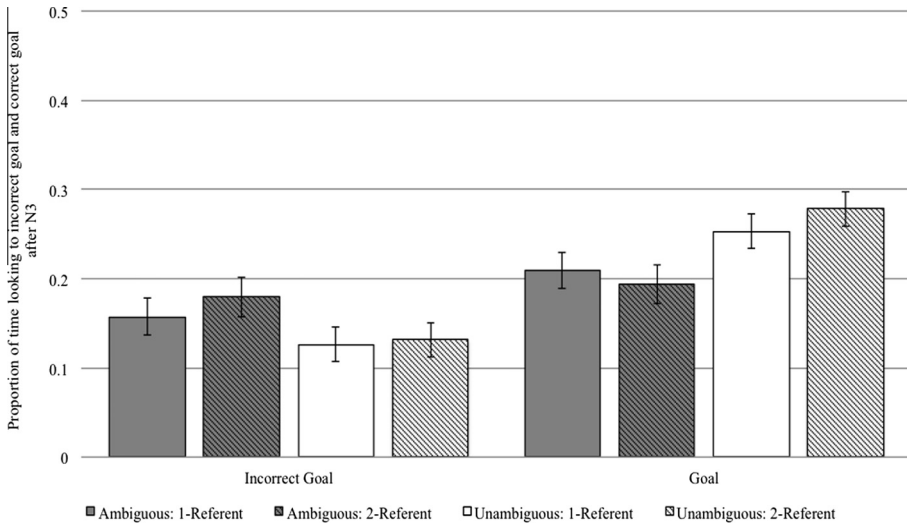


Fig. 4. Average proportions of time spent looking at the incorrect goal (left bars) and the correct goal (right bars) after the onset of the correct goal (e.g., “book”) as a function of ambiguity and referential context. Error bars represent standard errors.

Also as shown in Fig. 4, children spent less time looking at the incorrect goal (e.g., the empty napkin) in the unambiguous condition as compared with the ambiguous condition, although this effect was not significant; a corresponding multi-level model (Table 3) with looks to the incorrect goal as the dependent variable yielded no main effects or interactions.

Assessments used to predict garden path performance

Table A1 in Appendix A presents the mean, standard deviation, and range for the key measures of interest for each task. Fig. A1 graphs histograms of standardized scores for each of these measures. Note that all scores reflect *error/difficulty* such that higher values indicate poorer performance. Explanations of each measure appear below, but only briefly. Average proportions of errors on all target trials are reported for the TROG-2, Working Memory task, and Day/Night task. Second Card Sort Cost reports the difference between the proportion of errors for the first and second sorts in the Flexible Item Selection Task. No-Go Cost is the difference between omission errors on Go trials (i.e., no response) and hits on No-Go trials (i.e., no response). Determiner “Confusion” is the difference between the proportion of unique singleton character choices in indefinite determiner and definite determiner trials. For the Flanker task, a difference score in reaction times and accuracy between the congruent and incongruent conditions for each individual was calculated as a means to measure children’s ability to focus on relevant information and suppress irrelevant information. In line with similar findings from the adult literature, where sequential congruency effects (the so-called Gratton effect; Gratton, Coles, & Donchin, 1992; see also Botvinick, Braver, Barch, Carter, & Cohen, 2001; Weissman, Jiang, & Egner, 2014) are consistently observed, observation during testing suggested that children might react faster and respond more accurately when the current trial was of the same type as the preceding trial (“no switch,” e.g., a congruent trial following a congruent trial) than when it appeared right after a different trial type (“switch,” e.g., a congruent trial following an incongruent trial). For this reason, both the Flanker Congruency Cost and Flanker Switch Cost measures were included as predictors, where Flanker Congruency is the combined z-score for the RT and error difference between incongruent and congruent trials and Flanker Switch is the combined z-score for the RT and error difference between switch and no-switch trials. Composite z-scores were used for the Flanker measures in order to reduce the amount of predictors associated with a single task.

Table 3

Linear mixed effects model for the effects of ambiguity and referential context on e-logit transformed looks to incorrect goal (IG) and correct goal (CG) in the Put task.

Fixed effects	Looks to IG after hearing CG ^a				Looks to CG after hearing CG ^b			
	Estimate	SE	t-Value	p-Value	Estimate	SE	t-Value	p-Value
Intercept	−2.62	0.18	−14.13	<.001***	−1.71	0.14	−12.49	<.001***
Ambiguity	.26	.21	1.28	.21	−.54	.19	−2.87	<.01**
Referential context	.18	.22	0.85	.41	.09	.24	0.36	.72
Ambiguity/referential context	.24	.36	0.66	.51	−.14	.37	−0.38	.70

^a The model for looks to the IG after CG included the fixed effects of ambiguity, referential context, and their interaction. The maximum random effects structure that converged included by-item and by-participant random intercepts, together with by-participant random slopes for the main effects of ambiguity and referential context, and their interaction, and by-item random slopes for the main effects of ambiguity and referential context.

^b The model for looks to CG after CG included fixed effects of ambiguity, referential context, and their interaction. The maximum effects structure that converged included by-participant and by-item random intercepts and by-participant and by-item random slopes for all main effects and the interaction between ambiguity and referential context.

** $p < .01$.

*** $p < .001$.

Within our relatively small age window, children ranged widely in their performance on many of the cognitive tasks that they completed, although some tasks did not reveal large individual differences in performance. Children made a substantial amount of errors on the Working Memory task ($M = .24$),⁶ the Day/Night task ($M = .20$), and the TROG-2 ($M = .37$), indicating that the tasks were appropriately challenging for this age group. There were costs associated with the Second Card Sort (M [sort cost] = .26) and Flanker Congruency (M [RT cost] = 129.03 ms; M [error cost] = .16),⁷ although there were no obvious *mean* costs associated with No-Go (M [No-Go cost] = −.69) or Flanker Switch (M [error cost] = −.03; M [RT cost] = −10.62), despite substantial costs for some children at the individual level. Children struggled with the Determiner task and did not seem to base their referential choices on the determiner used by the experimenter ($M = −.02$, where a score of zero indicates no discrimination).

Relating assessments to garden path recovery

Here we analyze the relationship between individual differences on the assessments reported above and children's performance on the "Put" task. For each child, we calculated an average "error" score for each task, together with "cost" scores for the Card Sort task, the No-Go task, the Determiner task, and the Flanker task (as described above). These measures, in addition to age and performance for the unambiguous sentences in the "Put" task (measured in log odds), were used as predictor variables in a series of multiple regressions; all predictors were standardized as z-scores and were entered at the same time in the regression model.⁸ To ease concerns of multicollinearity, the variance inflation factor (VIF) was calculated for each regression analysis, and the standard threshold of 5 applied (Stine, 1995). The first set of analyses used act-out errors in garden path sentences (measured in log odds converted to z-scores) as the dependent variable (see "Predictors of act-out performance" section below). The second set of analyses used eye gaze during garden path sentences (measured in e-logits converted to z-scores) as the dependent variable (see "Predictors of eye gaze" section below). In both cases, we first modeled overall performance in garden path sentences and then the effect of referential context for garden path sentences only.

We included unambiguous "put" errors as a predictor in these analyses in order to better target children's ability to revise initial interpretive commitments in the ambiguous garden path condition. Errors in the unambiguous condition could result from a variety of factors that do not involve garden path recovery (e.g., language difficulties, perseveration errors, inattention). Therefore, including

⁶ Because one participant refused to complete the task, his score was replaced with the overall mean.

⁷ RTs less than 500 ms ($n = 5$) were excluded from the analyses.

⁸ See Appendix B for a correlation matrix of performance on all tasks.

unambiguous performance allowed us to target errors that are specific to garden path processing and recovery, while accounting for individual differences in performance in the unambiguous condition.

Predictors of act-out performance

We analyzed children's performance in the ambiguous condition of the "Put" task using all task predictors, together with age and performance in the unambiguous condition, in a multiple linear regression analysis (see Table 4 for model summary). Children's errors in the ambiguous condition were reliably predicted by their performance on unambiguous "put" sentences and their Flanker Switch Cost measure. In particular, higher error rates in the unambiguous condition were associated with higher error rates in the ambiguous condition (unambiguous "put": $\beta = .57$, $SE = .17$, $t = 3.37$, $p < .01$), and larger performance differences between switch and non-switch trials in the Flanker task were also associated with more errors on ambiguous sentences (Flanker Switch: $\beta = .52$, $SE = .15$, $t = 3.41$, $p < .01$); this suggests that children who were less affected by the difficulty of the "switching" between trial types in the Flanker task also performed better on garden path sentences (see Fig. 5). In line with this finding, performance on the Card Sort task was also related to performance on ambiguous sentences, although this effect did not quite reach significance (Card Sort Cost: $\beta = .27$, $SE = .14$, $t = 1.98$, $p = .06$). This marginal effect of the Card Sort Cost, although promising, should be interpreted with caution. Overall, these results suggest that a subset of EF abilities, cognitive flexibility in particular, play a significant role in garden path recovery, above and beyond the processing abilities necessary to correctly interpret complex unambiguous sentences.⁹

To better understand the effects of referential context in the ambiguous condition, we analyzed children's performance in the 2-referent ambiguous condition of the Put task using all task predictors, age, and performance on 1-referent ambiguous sentences in a multiple linear regression. This analysis allowed us to see which (if any) abilities would selectively predict performance in the 2-referent condition. Performance in the 1-referent ambiguous condition was the only significant predictor of children's performance in the 2-referent ambiguous condition (see Table 5 for model summary). This result is not particularly surprising given that children's act-out performance was not modulated by referential context (see "Performance on Put task" section above).

Predictors of eye gaze

Parallel analyses were also conducted on visual inspection patterns for garden path sentences. The dependent variables were looks toward the incorrect and correct goals after the onset of the correct goal phrase (e.g., "onto the book").

Children's gaze toward the *incorrect goal* in the ambiguous condition were predicted by their fixation patterns in the unambiguous condition; increased fixations to the incorrect goal in ambiguous sentences were associated with a parallel increase in fixations to the incorrect goal in unambiguous sentences (estimate = .46, $SE = .17$, $t = 2.66$, $p = .01$). None of the other measures contributed to explaining children's varying degrees of consideration of the incorrect goal in ambiguous sentences (all $ps > .10$; see Table 6).

Children's gaze toward the *correct goal* in the ambiguous condition were marginally predicted by the cost of switching between different trial types in the Flanker task (Flanker Switch Cost: estimate = $-.42$, $SE = .22$, $t = -0.19$, $p = .06$). None of the other measures, including gaze toward the correct goal in the unambiguous condition, contributed to predicting fixation patterns toward the correct goal in the ambiguous condition (see Table 7). The observed marginal effect of Flanker Switch Cost on children's eye gaze patterns is in line with our findings for their act-out performance; children who experience a greater cognitive cost of "switching" between trial types and demands in the Flanker task

⁹ Given that some of our predictors are partially correlated with each other (see Appendix B), one might be concerned that multicollinearity is affecting our key results. To address this issue, we used the common practice of examining the variance inflation factor (VIF) and found it to be very low across all predictors in all multiple regressions reported in this article ($M = 1.66$, range = 1.19–2.66). The highest VIF (2.66) is still well below 5, suggesting the absence of multicollinearity (but see O'Brien, 2007, for discussions of when a high VIF is acceptable). Moreover, given the fact that the Flanker Congruency Cost correlates negatively with the Flanker Switch Cost, and the Unambiguous "Put" Errors correlate with some of our other measures, we ran an additional multiple regression like that reported in Table 4 but removing Flanker Congruency Cost and Unambiguous "Put" Errors; this did not change the significant finding of the Flanker Switch Cost, suggesting that this relationship is not an artifact of multicollinearity.

Table 4

Results of a multiple regression showing the relationship between performance on ambiguous sentences and performance on the other cognitive and linguistic measures.

Fixed effects	β	SE	t-Value	p-Value
Intercept	.00	.12	0.00	1.00
Age	–.21	.18	–1.20	.24
Second Card Sort Cost	.27	.14	1.98	.06*
Day/Night Errors	–.10	.16	–0.58	.56
Determiner Confusion	–.24	.14	–1.72	.10
Flanker Congruency Cost	.13	.15	0.88	.38
Flanker Switch Cost	.52	.15	3.41	<.01**
No-Go Cost	–.20	.15	–1.29	.21
TROG-2 Errors	–.04	.17	–0.25	.80
Working Memory Errors	–.13	.16	–0.82	.42
Unambiguous “Put” Errors	.57	.17	3.37	<.01**

Note. Adjusted $R^2 = .42$, $p < .01$. Significant and marginally significant predictors are bolded. Variance inflation factor (VIF) <5 for all measures.

* $p < .10$.

** $p < .01$.

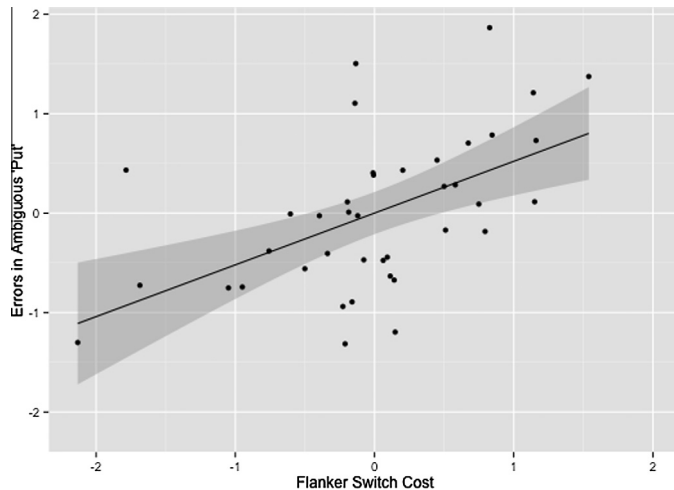


Fig. 5. Relation between errors in temporarily ambiguous sentences (residuals) and flanker switch costs (residuals). To plot the relationship between errors in temporarily ambiguous sentences and flanker switch costs in the regression model presented in Table 4, all predictors in the model (except performance on ambiguous sentences) were used to predict flanker switch costs and the residuals plotted on the x-axis; all predictors in the model (except flanker switch costs) were used to predict errors in temporarily ambiguous sentences and the residuals plotted on the y-axis.

seem to experience greater processing costs in garden path sentences, resulting in a decreased consideration of the correct goal (see Fig. 6).

Discussion

Summary and key observations

Children between 4 and 6 years of age showed difficulties in revising their initial misinterpretation of temporarily ambiguous sentences. When hearing a sentence containing a temporary ambiguity (e.g., “Put the frog on the napkin onto the book”), children often carried out an incorrect action,

Table 5

Results of a multiple regression showing the relationship between performance on 2-referent ambiguous sentences and performance on the other cognitive and linguistic measures.

Fixed effects	β	SE	t-Value	p-Value
Intercept	.00	.10	0.00	1.00
Age	-.05	.14	-0.34	.73
Second Card Sort Cost	-.05	.12	-0.38	.71
Day/Night Errors	.06	.12	0.51	.61
Determiner Confusion	-.12	.11	-1.04	.31
Flanker Congruency Cost	.03	.13	0.26	.79
Flanker Switch Cost	.21	.15	1.42	.16
No-Go Cost	.11	.18	0.93	.36
TROG-2 Errors	.05	.14	0.36	.72
Working Memory Errors	.07	.13	0.53	.60
Ambiguous 1-Referent "Put" Errors	.69	.13	5.46	< .001***

Note. Adjusted $R^2 = .59$, $p < .001$. Significant and marginally significant predictors are bolded. Variance inflation factor (VIF) <5 for all measures.

*** $p < .001$.

Table 6

Results of a multiple regression showing the relationship between looks to the incorrect goal during ambiguous sentences and the other cognitive and linguistic measures.

Fixed effects	β	SE	t-Value	p-Value
Intercept	.02	.16	0.13	.90
Age	-.08	.22	-0.36	.72
Second Card Sort Cost	.20	.20	1.01	.32
Day/Night Errors	.17	.23	0.73	.47
Determiner Confusion	-.17	.17	-0.94	.36
Flanker Congruency Cost	.06	.20	0.32	.75
Flanker Switch Cost	.28	.21	1.33	.19
No-Go Cost	-.26	.18	-1.42	.17
TROG-2 Errors	.01	.22	0.05	.96
Working Memory Errors	-.13	.20	-0.63	.53
Looks to IG in Unambiguous "Put"	.46	.17	2.66	.01*

Note. Adjusted $R^2 = .11$, $p = .22$. Significant and marginally significant predictors are bolded. Variance inflation factor (VIF) <5 for all measures.

* $p < .05$.

Table 7

Results of a multiple regression showing the relationship between looks to the correct goal during ambiguous sentences and the other cognitive and linguistic measures

Fixed effects	β	SE	t-Value	p-Value
Intercept	-.01	.16	-0.07	.95
Age	.24	.24	0.97	.34
Second Card Sort Cost	-.22	.20	-1.15	.26
Day/Night Errors	.04	.24	0.20	.85
Determiner Confusion	-.08	.18	-0.44	.66
Flanker Congruency Cost	.06	.20	0.33	.75
Flanker Switch Cost	-.42	.22	-1.95	.06*
No-Go Cost	-.05	.20	-0.24	.82
TROG-2 Errors	.11	.26	0.43	.67
Working Memory Errors	.25	.21	1.14	.26
Looks to CG in Unambiguous "Put"	.24	.21	1.10	.28

Note. Adjusted $R^2 = .04$, $p = .38$. Significant and marginally significant predictors are bolded. Variance inflation factor (VIF) <5 for all measures.

* $p < .10$.

moving the frog to an empty napkin. Very seldom did children make similar errors for unambiguous variants like “Put the frog that’s on the napkin onto the book.” This suggests that children in this age range often fail to recover from an initial incorrect analysis of the prepositional phrase “on the napkin” as the goal of the action described by the verb. Whereas children as a group often failed to recover from so-called garden paths, we also observed considerable individual differences; although the mean error rate for temporarily ambiguous sentences was approximately 60%, a roughly normal distribution was observed across children, ranging from 0% to 100% errors (see Fig. A2 in Appendix A).

Perhaps unsurprisingly, our results suggest that some of the variability in how well children respond to temporarily ambiguous sentences is not specific to garden path recovery, but instead reflects variability among children in how well they can process and act out similar unambiguous sentences containing an embedded relative clause—as evidenced by the fact that variance in performance on the ambiguous trials was partially accounted for by variance in performance on the unambiguous trials. This variability, in turn, might stem from a number of sources (e.g., differences in speed of processing/lexical access, vocabulary knowledge, general attention, and task engagement).

More importantly, our results suggest that, above and beyond the contribution of individual differences in performance on unambiguous sentences, individual differences in garden path recovery are associated with differences in general cognitive abilities related to executive function and cognitive control. In particular, individual children’s ability to overcome inertial tendencies favoring staying in the “groove” (see Davidson, Amso, Anderson, & Diamond, 2006; Kirkham, Cruess, & Diamond, 2003), flexibly adapting to different stimuli (e.g., congruent, incongruent, No-Go) in a nonlinguistic (Flanker) task, reliably predicted their degree of garden path recovery in an entirely different linguistic task (see Table 4 and Fig. 5). Additional converging evidence for a role of cognitive flexibility in garden path recovery comes from the existence of a marginally significant positive relationship between children’s performance on temporarily ambiguous sentences and their ability to flexibly and accurately switch from one rule to another for categorizing the same stimulus set in the Card Sort game. A relationship between garden path performance and cognitive flexibility similarly emerged in one of our online measures of garden path recovery; children who displayed lower switching costs in the Flanker task were also (marginally) more prone to “switch” their attention to the correct goal as soon as this was heard.

Although a considerable debate exists in the literature as to the exact causes and cognitive mechanisms that underlie sequential congruency (“switch”) effects (see Weissman et al., 2014, for an in-depth review), it seems reasonable to propose that underlying Flanker Switch, Second Card Sort, and Sentence Revision Costs is a generalized difficulty to quickly and flexibly adapt one’s behavior to changing situations, for example, a stimulus for which the current target response differs from the one just given either to the preceding stimulus (as in the case of the Flanker Switch and Second Card Sort) or to the linguistic input processed up to that point (as in the case of revision of a partial sentence structure).

Consistent with this interpretation, children’s garden path recovery abilities did not correlate with EF measures that did not include a switching component (e.g., the Flanker Congruency measure, the Day/Night task). To the extent that cognitive flexibility and inhibition are indeed two separate components of EF (Hill, 2004; Mazuka et al., 2009), evidence exists in the literature that the former, rather than the latter, might be particularly compromised in young children (see Davidson et al., 2006; Ozonoff, Strayer, McMahon, & Filloux, 1994) and, thus, might be linked to children’s difficulties with parsing revision. In line with our findings of an important link between linguistic and task-switching/set-shifting abilities, recent work on EF abilities and language learning has shown that task-switching abilities predict children’s abilities to learn an artificial language (Kapa & Colombo, 2013) and that, at least for young adults, a bilingual advantage in EF abilities emerges only when sequential congruency (“switch”) effects are taken into account (Grundy & Bialystok, 2015).

Children’s performance on the working memory task also failed to predict their garden path resolution abilities; this null result is surprising given a related line of findings in the adult literature that highlight the importance of (verbal) working memory resources on garden path resolution. According to this view, working memory capacity plays a central role in garden path resolution because when structural analyses and interpretations need to be updated in light of additional information, already-processed material needs to be brought back in the focus of attention, re-accessed, and

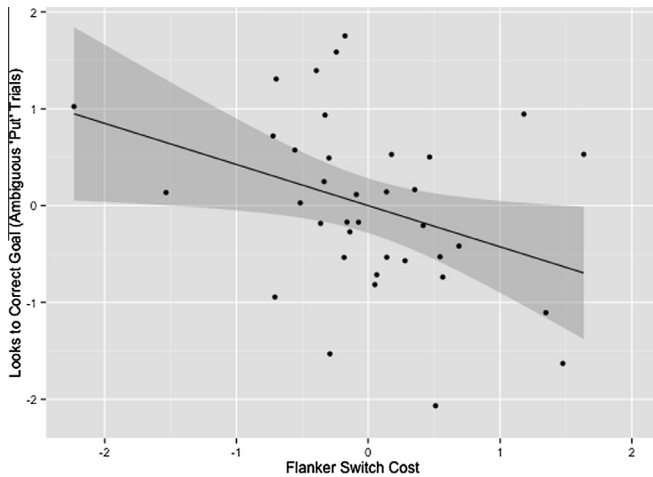


Fig. 6. Relation between looks to the correct goal and flanker switch cost. To plot the relationship between looks to the correct goal (in temporarily ambiguous sentences) and flanker switch cost in the regression model presented in Table 5, all predictors in the model (except looks to the correct goal in ambiguous sentences) were used to predict flanker switch costs and the residuals plotted on the x-axis; all predictors in the model (except flanker switch costs) were used to predict looks to the correct goal in temporarily ambiguous sentences and the residuals plotted on the y-axis.

re-analyzed. Indeed, individual differences in adults' working memory capacity have been found to correlate with their ability to process complex (e.g., Just & Carpenter, 1992; Miyake et al., 1994) and temporarily ambiguous sentences (e.g., Christianson, Williams, Zacks, & Ferreira, 2006; MacDonald, Just, & Carpenter, 1992; but see, e.g., Caplan & Waters, 1999). One possible explanation for the discrepancies between these studies and our own is that, in the former, the working memory and sentence processing tasks were quite similar in nature; both tapped, to some extent, into participants' ability to process written sentences because working memory was assessed by asking participants to read sentences and remember the last word in a sentence (for a similar criticism, see MacDonald & Christiansen, 2002). In contrast, our Working Memory task did not involve sentence processing given that children were simply asked to name pictures, and remember the last picture(s) in an array.

Evaluation of findings and implications for language learning

Our findings are broadly consistent with the EF account of garden path recovery (Novick et al., 2005) and similar statements about the role of EF in language processing (Mazuka et al., 2009). They fit nicely within the growing evidence in the adult literature—including individual differences studies (Mendelsohn, 2002; Vuong & Martin, 2014), training studies (Novick et al., 2014), neuroimaging (January et al., 2009), and neuropsychological case studies (Novick et al., 2009)—that EF abilities support revision of sentence interpretation.

The role of EF in garden path recovery and language processing more generally, suggested by our work and that of others, also has potentially important implications for grammar learning and, possibly, language change and language typology. For example, a straightforward prediction of the finding that children (and perhaps language learners in general) experience difficulties in revising initial interpretations is that language-specific properties that are likely associated with revision (e.g., they arrive late in a sentence after the parser has already committed to an interpretation) will be harder for children to learn and, for this reason, more subject to change (see Pozzan & Trueswell, 2015; Trueswell, Kaufman, Hafri, & Lidz, 2012).

Additional contributors to garden path recovery

General cognitive abilities are, of course, not the only factor that plays a role in garden path recovery. Linguistic knowledge is expected to play a role in that, in order to successfully process both ambiguous and unambiguous sentences, children need to parse the speech stream, access and retrieve word meanings, combine words into phrases, etc. as sentences unfold. In line with this, performance on unambiguous “put” sentences significantly predicted children’s performance on temporarily ambiguous ones above and beyond the effect of nonlinguistic EF abilities (see [Table 4](#)), suggesting that children’s varying lexical and syntactic abilities also play an important role in the processing of complex sentences. Although our work did not fully examine the contribution of lexical and syntactic knowledge to children’s sentence processing, our data suggest that individual differences in grammatical abilities (as measured by the TROG-2) are associated with children’s comprehension of unambiguous sentences ($r = .37, p < .01$; see [Appendix B](#)), but that these abilities do not contribute to additional variability specific to garden path recovery.

Likewise, another line of research on garden path recovery highlights another key difference between adults and children in the parsing of ambiguous sentences: Adults can use referential scene information to parse ambiguous sentences, whereas children cannot (e.g., [Snedeker & Trueswell, 2004](#); [Trueswell et al., 1999](#); [Weighall, 2008](#)). That is, adults are much less likely to be led down the garden path when two referents of the same kind (e.g., two frogs) are present in the visual scene when hearing a temporarily ambiguous sentence like “Put the frog on the napkin onto the book.” Children’s lack of sensitivity to this referential information might stem from a number of interconnected sources. For example, it seems reasonable to hypothesize that children’s varying abilities to use contextual information to correctly interpret 2-referent ambiguous sentences might be connected with differences in language abilities. In line with this, [Anderson and colleagues \(2011\)](#) found that children’s performance on 2-referent, but not 1-referent, ambiguous sentences correlated positively with vocabulary size.

More specifically, we had hypothesized that children’s inability to use referential context to avoid garden paths might stem from an insensitivity to the presuppositional properties of definite determiners. We explored whether children who do not yet master the distinction between definite and indefinite determiners might fail to notice that a definite determiner (“the frog”) in a context with multiple potential referents (e.g., two frogs) is pragmatically inappropriate and, thus, might be less likely to use it as a cue for a modifier interpretation of the following PP (“on the napkin”). Thus, we constructed a “determiner discrimination” task and used it to predict children’s accuracy on ambiguous 2-referent contexts. Contrary to our predictions, however, performance on this task did not predict performance on 2-referent ambiguous contexts (see “Predictors of act-out performance” section in Results).

It might be then that children’s failure to use referential context as a cue for assigning a “modifier” interpretation to the post-nominal PP is not due to a lack of the relevant knowledge of the semantics of definite and indefinite determiners, but rather to an inability to anticipate when their interlocutor is likely to produce a restrictive modifier. For instance, adult speakers are known to use a wide range of speaker and addressee knowledge to tailor their referential descriptions, often based on the current discourse, in ways that make them appear to be under- or over-informative in their descriptions of visually co-present objects (e.g., adult speakers are likely to utter sentences like “Give me the square” in the presence of multiple squares if they believe that the listener knows which square is relevant; [Brown-Schmidt, Campana, & Tanenhaus, 2002](#)). Given these facts, children might simply fail to compute their interlocutor’s referential domain and assume that their own referential domain is the same as their interlocutor’s (i.e., they might think that “the frog” that their interlocutor is referring to is the frog that they are currently looking at; see [Trueswell et al., 2011](#)).

Limitations and further questions

Our study used a cross-sectional design in which children were observed and tested at a single point in time. Because of this, the role of EF abilities in syntactic revision and language learning more

generally deserves further research. An intervention study whereby children are trained on executive function tasks and tested for improvements in sentence processing is necessary to better explore the proposed relationship. Our group has begun looking at this relationship in second-language learners, and the results look promising (see [Pozzan, Woodard, & Trueswell, 2014](#)).

Conclusions

We aimed to understand individual differences in syntactic ambiguity resolution in children and relate them to measures of linguistic and cognitive abilities in hopes of gaining a better understanding of the causes underlying children’s difficulties with revision and the cognitive mechanisms that support this process. We provide converging support for the hypothesis that children’s parsing difficulties

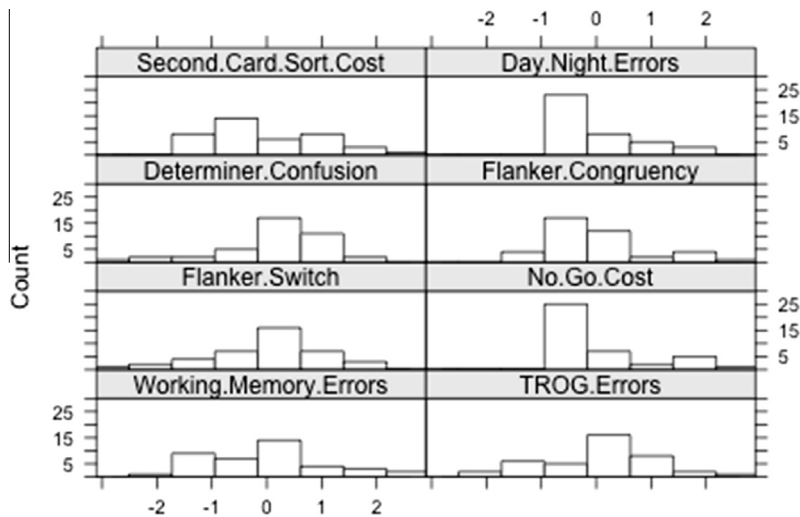


Fig. A1. Children’s performance on language, Working Memory, and EF tasks (standardized).

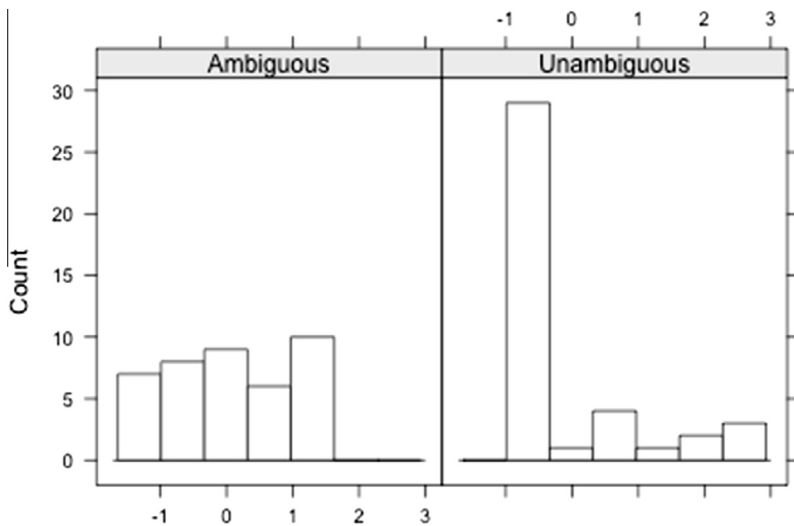


Fig. A2. Children’s act-out errors in the Put task by ambiguity condition (z-scores).

Table A1

Summary of performance on additional tasks (unstandardized).

	<i>M</i>	<i>SD</i>	Range (low:high)
Second Card Sort Cost	.26	.25	–.08:.83
Day/Night Errors	.20	.23	.00:1.00
Determiner Confusion	–.02	.32	–1.00:.50
Flanker Congruency Errors (I minus C)	.16	.23	–.16:.75
Flanker Congruency RTs (I minus C)	129.03	463.73	–975.58:1297.63
Flanker Switch Errors (Switch minus No Switch)	–.03	.21	–.52:.34
Flanker Switch RTs (Switch minus No Switch)	–10.62	369.65	–612.64:1005.35
No-Go Cost (Omission Errors minus Hits)	–.69	.35	–1.00:.08
TROG-2 Errors	.37	.13	.10:.69
Working Memory Errors	.24	.12	.03:.50

Note. For the Day/Night task, TROG-2, and Working Memory task, we report average proportion of errors on all target trials. Second Card Sort Cost is the difference between the proportion incorrect for the first and second sorts. Flanker Congruency RTs is the average reaction time difference between Incongruent (I) and Congruent (C) trials for correctly responded trials only. Flanker Congruency Errors is the average error difference between Incongruent and Congruent trials. Flanker Switch RT is the average reaction time difference between Switch and No-Switch trials for correctly responded trials only. Flanker Switch Errors is the average error difference between Switch and No-Switch trials. No-Go Cost is the average difference between omission errors on Go trials (i.e., no response) and hits on No-Go trials (no response). Determiner Confusion indicates the difference between singleton character choices for indefinite determiner trials and definite determiner trials.

Table B1

Zero-order correlation matrix of EF and language tasks.

	2	3	4	5	6	7	8	9	10	11
1. Age	–.28 ⁺	–.36 ⁺	–.31 ⁺	–.11	.12	.06	–.59 ^{***}	–.54 ^{***}	–.08	–.09
2. Second Card Sort Cost		.31 ⁺	.20	.15	–.24	.01	.24	.28 ⁺	.12	.17
3. Day/Night Errors			.09	.13	–.13	–.30 ⁺	.28 ⁺	.15	.33 ⁺	.21
4. Determiner Confusion				–.03	–.03	.06	.36 ⁺	.31 ⁺	.27 ⁺	–.06
5. Flanker Congruency					–.47 ^{**}	.10	.20	–.01	.21	.03
6. Flanker Switch						.21	–.07	–.12	.13	.44 ^{**}
7. No-Go Cost							.02	–.19	.34 ⁺	.14
8. TROG-2 Errors								.47 ^{**}	.37 ⁺	.17
9. Working Memory Errors									.00	–.07
10. Unambiguous “Put” Errors										.54 ^{***}
11. Ambiguous “Put” Errors										

⁺ $p < .10$.^{*} $p < .05$.^{**} $p < .01$.^{***} $p < .001$.

are associated with their immature EF abilities (Novick et al., 2005) and that the ability to quickly use linguistic evidence to abandon an initial structural analysis in favor of a novel one might be linked to the ability to flexibly adapt one's behavior to rapidly changing situations, overcoming inertial tendencies.

Taken together with recent findings that language-specific properties are easier to learn if they arrive early in a sentence and, thus, prevent the parser from engaging in costly revision, our findings highlight the interconnectedness of cognitive development, language processing, and language acquisition. It is our contention that processing preferences and limitations, which stem from the architecture of the human cognitive system, have a profound effect on both vocabulary and grammar acquisition.

Acknowledgements

We would like to thank Max Kaplan, Michal Snyder, Joseph Levasseur, Sarah Hendry, and Yi Lin for their assistance in data collection and scoring. We would also like to thank Vladimir Sloutsky for allowing us to use some of his images for our determiner task (Sloutsky, Kloos, & Fisher, 2007). This work was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant 1-R01-HD-37507.

Appendix A

See Figs. A1, A2 and Table A1

Appendix B

See Table B1.

References

- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8, 71–82.
- 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 (Trends in Language Acquisition Research (Vol. 7, pp. 203–221). Philadelphia: John Benjamins.*
- Bishop, D. (2003). *Test for Reception of Grammar-Version 2 (TROG-2)*. Oxford, UK: Pearson Assessment.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.
- Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2002, August). *Reference resolution in the wild: How addresses circumscribe referential domains in a natural, interactive problem-solving task*. Paper presented at the annual meeting of the Cognitive Science Society, Fairfax, VA.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77–126.
- Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An H2150 PET study of Stroop task performance. *NeuroImage*, 2, 264–272.
- Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, 106, 41–61.
- Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, 42, 205–238.
- 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.
- Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, 38, 352–362.
- Engel de Abreu, P. M., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor enhanced cognitive control in low-income minority children. *Psychological Science*, 23, 1364–1371.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like Day-Night test. *Cognition*, 53, 129–153.
- Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121, 480–506.
- Grundy, J. G., & Bialystok, E. (2015). *It's not that simple: Sequential congruency effects reveal a bilingual disengagement advantage*. Poster presented at the Workshop on Bilingualism and Executive Function: An Interdisciplinary Approach, CUNY Graduate Center, New York, NY.
- Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, 8, 26–32.
- 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.
- Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology*, 387, 167–178.
- Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, 20, 573–591.
- 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.
- Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 8410–8413.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kapa, L. L., & Colombo, J. (2013). Attentional control in early and later bilingual children. *Cognitive Development*, 28, 233–246.
- Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *Quarterly Journal of Experimental Psychology*, 63, 160–193.
- Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, 6, 449–467.
- Livesey, D., Keen, J., Rouse, J., & White, F. (2006). The relationship between measures of executive function, motor performance, and externalising behaviour in 5- and 6-year-old children. *Human Movement Science*, 25, 50–64.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109, 35–54.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56–98.

- Mason, R. A., Just, M. A., Keller, T. A., & Carpenter, P. A. (2003). Ambiguity in the brain: What brain imaging reveals about the processing of syntactically ambiguous sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1319–1338.
- Mazuka, R., Jincho, N., & Onishi, H. (2009). Development of executive control and language processing. *Language and Linguistics Compass*, 3, 59–89.
- Mendelsohn, A. (2002, March). *The role of inhibition in sentence processing as demonstrated by a new task*. Paper presented at the CUNY Conference on Human Sentence Processing, Cambridge, MA.
- Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., et al (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, 12, 467–473.
- Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, 11, 671–717.
- 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.
- Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, 58, 220–249.
- 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.
- 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.
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5, 263–281.
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41, 673–690.
- Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in English and Japanese wh-questions. *Language Learning and Development*, 10, 206–233.
- Ozonoff, S., Strayer, D. L., McMahon, W. M., & Filloux, F. (1994). Executive function abilities in autism and Tourette syndrome: An information processing approach. *Journal of Child Psychology and Psychiatry*, 35, 1015–1032.
- Pozzan, L., & Trueswell, J. C. (2015). Revise and resubmit: How real-time parsing limitations influence grammar acquisition. *Cognitive Psychology*, 80, 73–108.
- Pozzan, L., Woodard K., & Trueswell, J. C. (2014, November). *Executive function predictors of learners' language processing abilities: A training study*. Paper presented at the 39th annual Boston University Conference on Language Development, Boston, MA.
- Qi, Z., Fisher, C., & Brown-Schmidt, S. (2011). *On recovering from “kindergarten path” errors: Referential context and executive functioning influence children's online ambiguity resolution*. Paper presented at AMLaP (Architectures and Mechanisms for Language Processing) 2011, Paris, France.
- Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., et al (2004). Development of attentional networks in childhood. *Neuropsychologia*, 42, 1029–1040.
- Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything Appearance similarity versus kind information in early induction. *Psychological Science*, 18(2), 179–185.
- 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.
- 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.
- Stine, R. A. (1995). Graphical interpretation of variance inflation factors. *American Statistician*, 49, 53–56.
- 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.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 14792–14797.
- Trueswell, J. C., Kaufman, D., Hafri, A., & Lidz, J. (2012). Development of parsing abilities interacts with grammar learning: Evidence from Tagalog and Kannada. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th annual Boston University Conference on Language Development* (pp. 620–632). Somerville, MA: Cascadia Press.
- Trueswell, J. C., Papafragou, A., & Choi, Y. (2011). Syntactic and referential processes: What develops. In E. Gibson & N. Pearlmuter (Eds.), *The processing and acquisition of reference* (pp. 65–108). 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.
- Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience*, 29, 312–325.
- 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.
- Weissman, D. H., Jiang, J., & Egner, T. (2014). Determinants of congruency sequence effects without learning and memory confounds. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 2022–2037.
- Wexler, K. (2011). Cues don't explain learning: Maximal trouble in the determiner system. In E. Gibson & N. Pearlmuter (Eds.), *The processing and acquisition of reference* (pp. 15–42). Cambridge, MA: MIT Press.
- Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*, 48, 280–290.