Cognitive Psychology 131 (2021) 101442

Contents lists available at ScienceDirect

Cognitive Psychology

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

Word vs. World Knowledge: A developmental shift from bottom-up lexical cues to top-down plausibility

Anthony Yacovone *, Carissa L. Shafto, Amanda Worek, Jesse Snedeker

Harvard University, United States

ARTICLE INFO

Keywords:
Ambiguity resolution
Comprehension
Cue validity
Lexical bias
Plausibility
Syntax

ABSTRACT

Both 5-year-old children and adults infer the structure of a sentence as they are hearing it. Prior work, however, has found that children do not always make use of the same information that adults do to guide these inferences. Specifically, when hearing ambiguous sentences like "You can tickle the frog with the feather," children seem to ignore the aspects of the referential context that adults rely on to resolve the ambiguity—e.g., are there two frogs in the scene, one with a feather and one without? Or is there only one frog to be tickled by using a feather? The present study explored two hypotheses about children's failure to use high-level, top-down context cues to infer the structure of these ambiguous sentences: First, children may be less likely to use *any* top-down cue during comprehension. Second, children may only have difficulties with top-down cues that are unreliable predictors of which syntactic structure is being used. Thus, to disentangle these hypotheses, we conducted a visual world study of adults' and children's ambiguity resolution, manipulating a more reliable top-down cue (the plausibility of the interpretation) and pitting it against a robust bottom-up cue (lexical biases). We find that adults' and children's final interpretations are influenced by both sources of information: adults, however, give greater weight to the top-down cue, whereas children primarily rely on the bottom-up cue. Thus, children's tendency to make minimal use of top-down information persists even when this information is highly valid and dominates adult comprehension. We propose that children have a systematic propensity to rely on bottom-up processing to a greater degree than adults, which could reflect differences in the architecture of the adult and child language comprehension systems or differences in processing speed.

1. Introduction

By their fourth birthday, children seem to understand most (if not all) of what is being said to them. To do this, they have to identify the words they are hearing, link them together into phrases, and figure out why the speaker said that particular utterance. Understanding how each of these steps is carried out, and how they interact with one another during comprehension, has been a central goal of cognitive science for over 50 years. Language comprehension involves the coordination of both low-level and high-level information—the lowest levels of representation are defined as those closest to perception, i.e., perceiving speech sounds, recognizing individual words (lexical access), and characterizing the intonation of an utterance (prosodic analysis). In contrast, in comprehension,

* Corresponding author at: Department of Psychology, Harvard University, William James Hall, 33 Kirkland St., Cambridge, MA 02138, United States.

E-mail address: anthony_yacovone@g.harvard.edu (A. Yacovone).

https://doi.org/10.1016/j.cogpsych.2021.101442
Received 17 August 2020; Received in revised form 3 October 2021; Accepted 9 October 2021
Available online 24 November 2021
0010-0285/

the highest levels of representation are those linked to the speaker's intended message and goals. This includes information about how the message relates to the world around you (referential context) or what is the most likely interpretation of the utterance given your knowledge of the world (plausibility).

Contemporary theories emphasize two features of adult language comprehension: *incrementality* and *interactivity* (e.g. MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; see MacWhinney, 2001 for an overview). An incremental language system is one that passes up information from lower levels of representation to higher ones without waiting for the lower-level processes to finish (Allopenna, Magnuson, & Tanenhaus, 1998; Altmann & Steedman, 1988; Boland, Tanenhaus, Garnsey, & Carlson, 1995; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Incrementality allows us to start constructing the meaning of a sentence after hearing the first few words. For example, after hearing "Tim started to…" we know that there is a person named Tim who started a new action, which will soon be described. Incremental processing is largely beneficial, as it distributes the work of comprehension across the entire utterance, rather than leaving it all to the end. Incrementality, however, has a cost—sometimes we initially select the wrong interpretation of an ambiguous word or phrase, and we have to go back and revise it. For example, in sentence (1a) below, there is a prepositional phrase (PP)-attachment ambiguity, as it is initially unclear whether the prepositional phrase (e.g. *with the feather*) is describing how the action was completed (e.g. Tim brushed the cat *by using* the feather) or describing which cat was being brushed (e.g.

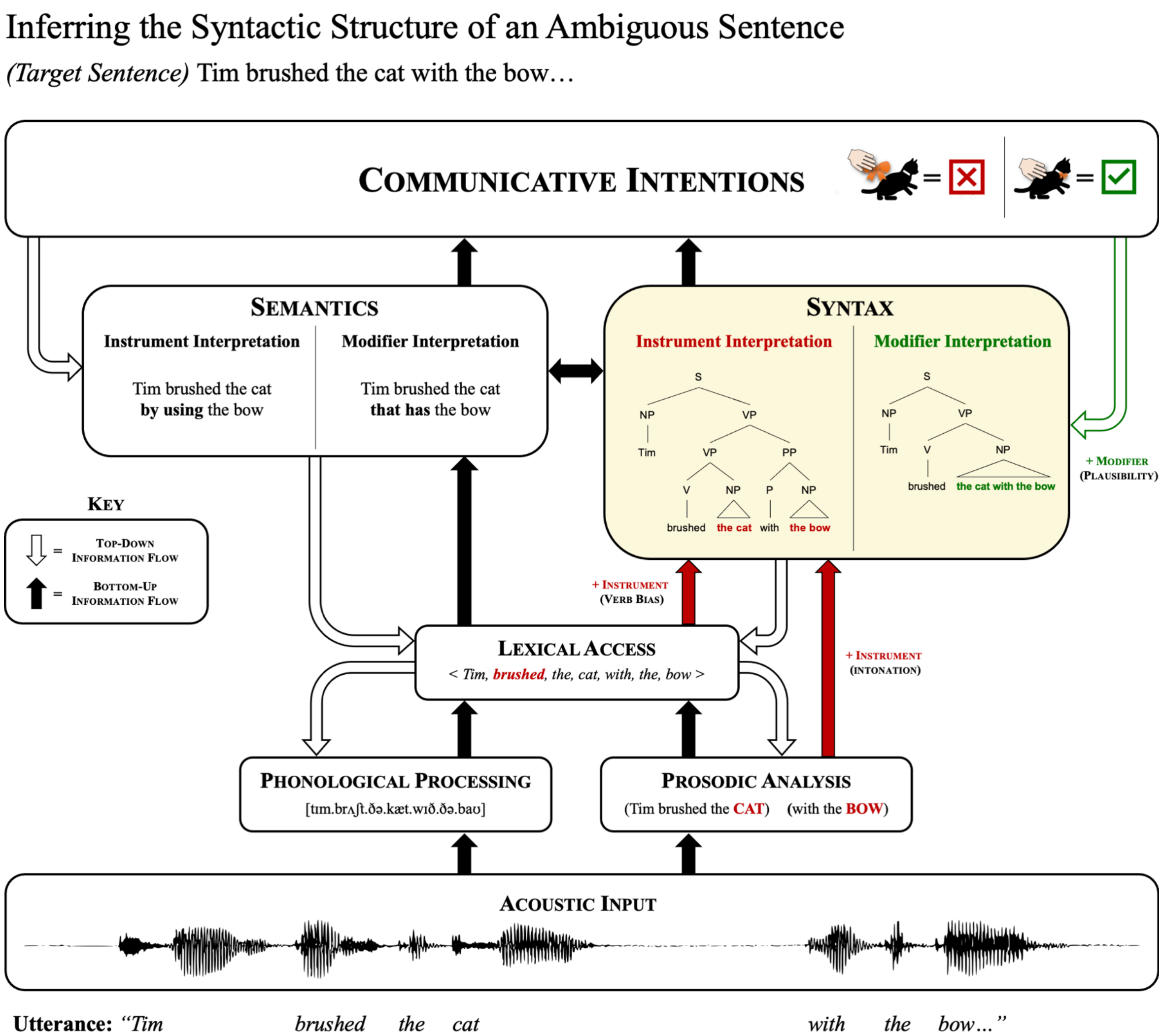

Fig. 1. *Illustration of how levels of representation interact when inferring the structure of "Tim brushed the cat with the bow…"*. This target sentence is ambiguous at the syntactic level (highlighted in yellow), and thus at higher levels as well (e.g. Semantics, Communicative Intentions). A prosodic cue (the pause after *cat* in the speech stream) and a lexical cue (the verb) provide bottom-up information favoring the instrument interpretation (e.g. brush the cat *by using the bow*, see red arrows). Our knowledge of likely communicative intentions, however, provides top-down information (e.g. the plausibility of wearing a bow and the implausibility of brushing with a bow) that supports the modifier interpretation (e.g. brush the cat *that is wearing the bow*, see green arrow). Note, this figure illustrates some, but not all, of the relevant cues that could be used to infer syntactic structure when a sentence is ambiguous. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Tim brushed the cat *that was holding the feather*, and not some other cat). This initial ambiguity is ultimately resolved by the end of the sentence after people learn that Tim used a comb to brush the cat—however, since this information arrived late in the sentence, people may have first interpreted (1a) incorrectly as *Tim used the feather to brush the cat*, forcing them to revise that interpretation after hearing about the comb (see Bever, 1970; Frazier & Fodor, 1978).

(1) a. Tim brushed the cat with the *feather* using the comb.
b. Tim brushed the cat with the *bow* using the comb.

An interactive language system is one in which many distinct sources of information are used to arrive at an interpretation. In such a system, a process that is ongoing at one level of representation can be influenced by information coming from levels below and/or above it. For example, when inferring the syntactic structure of an utterance, adults might rely on low-level information like prosody or the words themselves, as well as high-level information like the broader discourse context. When information flows from a lower level to a higher one, this process is called *bottom-up processing*. Alternatively, when information flows from a higher level to a lower one, this process is called *top-down processing* (see Fig. 1 for an illustration of these processes).

Interactivity allows listeners to avoid, or quickly revise, any misinterpretations that can arise as a result of incremental processing (e.g. MacWhinney, 1987; MacDonald et al., 1994; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell & Gleitman, 2004). For example, as we noted above, when a listener first encounters "with" in the sentences in (1), they probably assume that it is introducing an instrument phrase because the verb is one that frequently appears with an instrument (e.g. brushing *with* some object). In (1a), the next noun (feather) reinforces this hypothesis because it is a reasonable instrument for brushing. In contrast, in (1b), the next noun (bow) refers to an object that is not a good instrument for brushing. Adults may use this high-level information (e.g. the affordances of a bow) to revise their hypothesis, ultimately interpreting the *with*-phrase as a modifier instead, e.g., Tim brushed the cat *that had a bow* (see Rayner, Carlson, & Frazier, 1983; Taraban & McClelland, 1988; Garnsey et al., 1997; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Kamide, Altmann, & Haywood, 2003).

Comprehension in young children (4–7 years old) is similar to adults in some ways and different in others. Like adults, children incrementally interpret words and sentences (e.g. Allopenna et al., 1998; Trueswell, Sekerina, Hill, & Logrip, 1999; Yacovone, Rigby, & Omaki, 2020). Children's language comprehension can also be interactive. For example, when interpreting PP-attachment ambiguities like "Tickle the frog with the feather," children use the intonation of the utterance and information about the verb to constrain their syntactic analyses in a bottom-up fashion (Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008). However, interactivity seems to be more limited in children's comprehension for reasons that we do not yet fully understand. For example, when adults hear PP-attachment ambiguities, their interpretations are often constrained by top-down information about the number of referents (e.g. frogs) in the context in which the utterance occurred. If there are two or more frogs, most adults interpret the sentence as meaning "Tickle the frog *that has the feather.*" If there is just one frog, adults generally interpret the sentence as "Tickle the frog *by using the feather*" (e.g. Tanenhaus et al., 1995). In contrast, young children make little to no use of this top-down referential context information (e.g. Trueswell et al., 1999; Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Snedeker & Trueswell, 2004; Anderson, Farmer, Goldstein, Schwade, & Spivey, 2011; see Omaki & Lidz, 2015; Snedeker & Huang, 2015 for reviews).

Children's failures to make use of top-down cues like the referential context during syntactic processing is an interesting puzzle.[1] In the remainder of this Introduction, we review what we know about how children infer the syntactic structure of an utterance in real time. Then, we discuss two hypotheses for why children might fail to use top-down cues like the referential context when interpreting structurally ambiguous phrases. One hypothesis is that children have a broad deficit in using all kinds of top-down information during syntactic processing. A second hypothesis is that some top-down cues, like referential contexts, may simply be poor predictors of a sentence's intended structure, and thus may be acquired late. Finally, we describe how the present study tests these two hypotheses by exploring the degree to which children use a top-down cue that should be highly predictive of a sentence's intended structure in child-directed speech. Specifically, we manipulate the plausibility of the instrument interpretation of an ambiguous *with*-phrase by varying the affordances of the object that appears in it—similar to the examples in (1), where the feather was a highly plausible instrument for brushing but the bow was not.

[1] We want to briefly clarify the scope of the present investigation. Prior work has shown that children are less likely to use higher levels of information to disambiguate lower-level representations for a variety of ambiguities at multiple levels of representation (see Snedeker & Huang, 2015 for a review). In this paper, however, we are specifically interested in how children and adults use different bottom-up and top-down sources of information to guide inferences about sentence structure. Our discussion of children's prior failures to make use of top-down information during syntactic processing is *not* an argument that children are unable (or even less able than adults) to construct high-level representations (e.g. referential intent) and then deploy them in any task. For example, there are ample data showing that children rely on the referential context for processes like word learning (see Akhtar, 2002; Fennell & Waxman, 2010; Matthews, Butcher, Lieven, & Tomasello, 2012). The hypotheses that we are exploring do not make any predictions about tasks like word learning in which there is no clear competition between bottom-up and top-down information. Although, our hypotheses could be quite naturally extended to other phenomena that *do* involve competition of this kind like disambiguation of homophones, phonological processing, and even non-linguistic tasks like object categorization—but exploring these predictions is beyond the scope of the present paper.

1.1. Children's moment-to-moment comprehension of structurally ambiguous sentences

Our understanding of children's moment-to-moment language comprehension largely comes from studies using the visual world paradigm (Allopenna et al., 1998). In this paradigm, participants look at scenes containing various objects while they listen to simple instructions or descriptions (see Huettig, Rommers, & Meyer, 2011 for paradigm review). For example, Trueswell et al. (1999) presented children and adults with sentences containing PP-attachment ambiguities like "Put the frog on the napkin in the box." While listening to these sentences, participants saw one of two types of scene contexts. One scene type had an empty napkin, an empty box, a horse (sitting on nothing), and a single frog sitting on a napkin. Thus, it is clear that this frog is the one being mentioned in the sentence above. In contrast, the other scene type had the same objects, but instead of having a horse sitting on nothing, they had a second frog. In the contexts with two frogs, it is now unclear which frog is being referenced in the early part of the sentence above.

Trueswell et al. (1999) found that, in contexts with just one frog, adults initially pursued Verb-Phrase (VP)-attachment, looking to the empty napkin in anticipation of moving the frog there (e.g. "Put the frog *onto* the napkin…). But as the utterance progressed, adults revised their misanalysis and provided final actions that matched the intended Noun Phrase (NP)-attached structure (e.g. "Put the frog *that's on* the napkin into the box). In contexts with two frogs, adults rarely looked to the empty napkin, suggesting they made rapid use of the referential context to arrive at the intended structure. Children performed differently than adults in two ways: First, they were not influenced by the referential manipulation. They looked to the empty napkin equally often in both the one frog and two frog contexts, indicating their pursuit of VP-attachment in both cases (e.g. "Put…on the napkin"). Second, they did not revise their initial misinterpretations upon hearing the rest of the utterance. In both contexts, they would typically pick up a frog, move it to the empty napkin, and then move that frog (or the other one, if present) to the box. This pattern persists until about 8 years of age (e.g. Weighall, 2008; see also Hurewitz et al., 2000; Choi & Trueswell, 2010; Anderson et al., 2011).

The current study focuses on the first part of this pattern: children's failure to use the referential context. Early on, researchers considered the possibility that this failure was due to a shift over development from a modular comprehension system to an interactive one (e.g. Traxler, 2002; Felser, Marinis, & Clahsen, 2003; Joseph et al., 2008). However, this hypothesis turned out to be unsupported, as children between the ages of 4 and 7 use a variety of cues to inform syntactic analyses (see Omaki & Lidz, 2015; Snedeker & Huang, 2015 for reviews).

One of the cues available to young children is how frequently certain verbs occur in particular syntactic or semantic environments. For example, verbs like *tickle* or *brush* appear in many different environments, but they often occur in sentences that explicitly mention an instrument, and thus we come to expect an instrument after hearing these instrument-biased verbs. Similarly, modifier-biased verbs like *hug* or *find* occur more often with modified NPs than with instruments (e.g. hug the *bear with spots*, find the *horse with stripes*), which generates the expectation of a modifier, rather than an instrument, after the noun. Cues like these are often referred to as a lexical or verb biases, and they can have a strong influence on how comprehenders interpret ambiguous syntactic structures (e.g. Connine, Ferreira, Jones, Clifton, & Frazier, 1984; Garnsey et al., 1997; Roland & Jurafsky, 2002; Gahl, Jurafsky, & Roland, 2004; Snedeker & Trueswell, 2004). For example, Snedeker and Trueswell (2004) found that both adults and children used the bias of the verb to interpret syntactically ambiguous sentences like "You can tickle the frog with the feather."

But verb biases are not the only cues that children use. Snedeker and Yuan (2008) found that children (and adults) were also able to use prosody (i.e. the intonation of a sentence) to infer the intended phrase structure. If the speaker paused briefly after the verb, e.g., "You can tickle…the frog with the feather," listeners adopted more modifier interpretations such as *tickle the frog that is holding the feather*. If the pause occurred after the noun, e.g., "You can tickle the frog…with the feather," they adopted more instrument interpretations such as *tickle the frog by using the feather* (cf. Choi & Mazuka, 2003). Taken together, these studies demonstrate that children's syntactic processing is interactive, as it relies (at least) on verb biases and intonation.

These studies, however, do not explain why children consistently fail to use referential context cues when inferring phrase structure. Are these failures related to using contextual information more broadly or to using the specific referential cue manipulated in these studies? Understanding the scope of this deficit could help us understand the development of interactive systems. Thus, we consider two hypotheses: First, children may have a broad deficit in using higher-level, top-down cues during syntactic processing regardless of how robust these cues are. Second, children may have difficulties using cues that are weak predictors of phrase structure, regardless of whether they come from higher or lower levels of representation, perhaps because weaker cues are more difficult to acquire (see further discussion below). Under both of these hypotheses, we would predict that sensitivity to top-down cues like the referential context will develop later in life, after children overcome their processing limitations from childhood. In the remaining sections, we discuss each hypothesis in turn, and then describe how a manipulation of object affordances (i.e. plausibility) can be used to disentangle them.

1.2. Do children have a broad deficit in using top-down cues to infer phrase structure?

Our first hypothesis is that children do not use referential context cues to infer phrase structure because they have a broad deficit in using *any* higher-level cue to inform on-going processes at lower levels. This hypothesis is rooted in an information-processing tradition of cognitive psychology. According to this tradition, perception is treated as a set (or stream) of processes in which sensory input is transformed, and then used to build distinct and increasingly abstract representations. For example, visual perception is the stream of processes by which photons (i.e. low-level information) give rise to representations of objects and scenes (i.e. high-level information). The bottom-up flow of information—from the retina to the lateral geniculate nucleus, and through a series of cortical regions—results in increasingly complex representations of the visual field, ranging from retinotopic center-surround cells to cortical representations of motion, shape, and object identity (see Maunsell & Newsome, 1987; Lamme & Roelfsema, 2000 for greater detail).

This bottom-up processing stream is influenced by expectations at higher-levels via top-down connections, resulting in context effects, priming, and even some visual illusions (e.g. Berry & Schwartz, 2011; Teufel & Nanay, 2017).

Given the ubiquity of top-down processing in cognitive systems, it might seem unlikely that children as old as 4–7 years of age would have a global deficit of this kind. We can imagine three possible forms that a top-down deficit would take, all of which vary in their a priori plausibility. First, children could be unable to use top-down information during language comprehension because they lack the relevant neural pathways to transfer information across these levels of representation. For example, children could have a bottom-up pathway linking words to syntax, and syntax to even higher levels of representations (e.g. combinatorial conceptual structures or intended messages)—but lack a parallel top-down pathway that allows for information from conceptual structures to affect low-level syntactic or lexical processing. On this hypothesis, child cognition would look a lot like the information-processing models of the 20th century, with a unidirectional flow of information from perception up to a central workspace (e.g. Atkinson & Shiffrin, 1968; for a recent review, see Malmberg, Raaijmakers, & Shiffrin, 2019).

In the case of language, this hypothesis is a non-starter because children not only comprehend language, but they also produce it. Language production requires connections from higher-level representations to lower-level ones, allowing us to select words and structures on the basis of intended messages. Thus, it seems reasonable to believe that the representations used in comprehension are the same as those used in production—at least at the word-level and above. Evidence for this notion comes from interactions between the two processes in priming and language acquisition (Stackhouse & Wells, 1997; Huttenlocher, Vasilyeva, & Shimpi, 2004; McCauley & Christiansen, 2011; Dell & Chang, 2013; Pickering & Garrod, 2013; see Meyer, Huettig, & Levelt, 2016 for an overview). Thus, in many contemporary theories, language production and language comprehension involve the same cognitive system and the same connections, but with a reversal of the dominant flow of information through the system. In comprehension, the dominant flow of information is bottom-up from sounds to messages, whereas, in production, the dominant flow is top-down from messages to sounds (see Dell & Chang, 2013 for further discussion). If we accept these contemporary theories, then the mere fact that children's earliest productions reflect their thoughts rules out the possibility that children are missing the top-down connections between communicative intentions and syntactic forms. Whether they use these connections during comprehension (and whether they accurately represent the intentions of others) are separate issues altogether.

Second, a top-down deficit could arise from architectural limits on how information is shared across levels of representation. Early theories of language comprehension argued for modular processes in which decisions at a given level were solely based on the input from the level immediately below it and then revised in consultation with higher-level cues (Forster, 1979; Frazier, 1978; Swinney, 1979). For example, syntactic parsing was initially argued to only involve information about the syntactic categories of the words (which had been identified at the level below) and not higher-level information (e.g. Friederici, Hahne, & Mecklinger, 1996). Prior findings rule out the possibility that children's parsing is strictly modular: by age five, children can rapidly integrate multiple information sources like prosody and verb bias (Snedeker & Yuan, 2008; Snedeker, 2013). But there is a weaker architectural hypothesis that is consistent with the findings to date—perhaps, children have difficulty integrating different streams of information in real time, and these capacity limitations lead them to rely primarily on those cues that reflect the dominant flow of information for a given process. This *limited cue integration* account would predict that children rely heavily on bottom-up information during comprehension (e.g. lexical biases, prosodic information) and on top-down information during production (e.g. referential intentions, pragmatic goals). As children's abilities to juggle multiple sources of information improves, perhaps due to improved working memory or executive function, they would begin to more effectively integrate information from the non-dominant stream.

Third, a top-down deficit could emerge as a side effect of slower processing speed. We will call this the *processing speed* hypothesis, and this hypothesis draws on the early work of Altmann and Steedman (1988), and Dell (1986). Altmann and Steedman (1988) outline the logic of how referential contexts could be used to infer that the intended analysis of an ambiguous sentence like "Brush the cat with the feather" was NP-attachment (i.e. brush the cat *that has a feather*). They conclude that any comprehension system (modular and non-modular) will go through a short initial processing stage in which perceptual input is used to construct the possible analyses (e.g. the NP and VP-attachment) before these analyses can be evaluated with respect to their referential implications at a higher level. Thus, the question of modularity, they argue, is really a question of how much bottom-up processing occurs before the implications at higher levels are evaluated. While much has changed in psycholinguistics, the logic of their argument still holds in all but the most constraining contexts. Unless you can predict the specific form (or the set of forms) that is likely to be produced, you will first have to recognize the words, and their potential relations, before determining which top-down constraints are relevant. Thus, the use of top-down constraints will often involve the construction of a multi-step feedback loop: 1) activate the relevant alternatives at the lower level, 2) propagate this information up to the higher level, 3) evaluate it relative to this higher-level knowledge, 4) send the results of that analysis back down to the lower level, and 5) repeat until the relevant interpretative commitments are made (again, see Fig. 1 for an illustration of these processes).

Dell (1986) explored the temporal dynamics of feedback loops in adult language production. In production, in contrast with comprehension, feedback loops will be involved anytime *lower-level* cues shape *higher-level* processes. One example of these dynamics is how segment selection can influence word selection, resulting in speech errors that are more likely to be words than non-words. For example, speech errors that turn "BROWN COW" into "CROWN BROW" (i.e. two real words) are more likely to occur than those that turn "BROWN HORSE" into "HOWN BRORSE" (i.e. two non-words). Dell (1986) finds that this bias towards real words (also called the *lexicality bias*) is greater when participants have more time to plan their speech, demonstrating that the effect of the feedback loop takes time to emerge. This raises an intriguing possibility about development: children, in general, process information more slowly than adults do (e.g. Hale, 1990; Kail, 1991; Kail & Salthouse, 1994; Kail & Ferrer, 2007; MacWhinney, 2012). This means that in any given period of time, if children and adults are doing the same task, the children will get less done. In other words, a young child might be a lot like an adult operating under time pressure—unable to construct a feedback loop quick enough to make use of it in real time. During language comprehension, this inability might prevent children (and adults under time pressure) from accessing top-down cues in time to affect their interpretation of the ambiguous phrase. The time limit (i.e. the need to commit to an interpretation) could come either from the pressure of new words coming in, or perhaps from the build-up of bottom-up cues coupled with children's inability to keep multiple possibilities in mind at once.

The present study does not try to distinguish between the *limited cue integration* hypothesis and the *processing speed* hypothesis presented above. Both of these hypotheses predict the same pattern—young children will still be far less likely than adults to use top-down cues when inferring phrase structure, regardless of how robust or informative the cue is about syntactic structure. Although these hypotheses make the same prediction, their utility is that they make *opposite* predictions than a hypothesis that says children do not use top-down cues that poorly predict which syntactic structure is being used (see section below for greater discussion). Note, we acknowledge that these two classes of hypotheses are not the only two possibilities, as there are many other factors that might affect when and how top-down cues are used by children during comprehension.

1.3. Do children fail to notice referential context cues because they poorly predict syntax?

As we mentioned above, an alternative explanation for the slow emergence of top-down cues in children's syntactic processing centers around the relative effectiveness or validity of the cue itself. Specifically, children may be slow to acquire certain top-down knowledge because the cue only weakly predicts which syntactic structure is being used (Trueswell & Gleitman, 2007). In models of this type (often called constraint-satisfaction models), cue acquisition and cue use depend on a cue's *validity* (Brunswik, 1956; Gibson, 1966; MacWhinney & Bates, 1989). Cue validity is a function of cue *availability* (how often is the cue present) and cue *reliability* (how often, when present, does the cue lead to a correct interpretation). These models predict that children should initially rely heavily on the most valid cues in their input, whereas adults, who have had more time to soak up the statistics of the input, might consider additional cues that are less valid but still useful like referential contexts (see MacWhinney, 1978; MacWhinney, Bates, & Kliegl, 1984; Sokolov, 1988; MacWhinney, Leinbach, Taraban, & McDonald, 1989). MacWhinney and colleagues tested this prediction in the domain of thematic role assignment (i.e. figuring out *who* did *what* to *whom*; see MacWhinney, 2012 for a review), but the logic applies equally well to structural ambiguity resolution.

The referential context manipulation taps into the following knowledge: Consider definite singular nouns like *the frog*. If there are many possible referents for the mentioned noun (i.e. if there are many frogs around you), there is a greater likelihood of hearing a *with*-phrase that disambiguates which referent is being discussed (e.g. the frog *with the spots*, not the frog *with the blue legs*). This correlation between *with*-phrases and number of referents is a complex contingency to track and learn from linguistic input: A child must first home in on the right category of things to be counting (e.g. definite singular nouns, and post-nominal *with*-phrases). Then, they must correctly infer the number of possible referents in the speaker's referential model. Finally, they must track the data to discover this contingency. The amount of time it takes for a child to acquire this cue likely depends on how robust the correlation is in their input.

For adult speakers, the contingency between the number of referents and the use of NP-modification (e.g. *with the spots*) appears to be quite weak. For example, Brown-Schmidt, Campana, and Tanenhaus (2002) found that adults often produced under-specified utterances like "Pick up the block" when there were multiple blocks in the immediate context. Curiously, their interlocutors were typically able to pick out the intended referent, presumably because the prior discourse and their shared goals disambiguated which object was being discussed. Thus, to pull out the contingency between referential ambiguity and modification from dialogs such as these, a learner would have to closely follow the prior discourse and correctly infer the referential intentions of adults. If the number of possible referents is often unavailable to the child (due to limited understanding of the discourse) or is often misleading (due to misunderstanding an adult's discourse model), then we should expect that children will need more data and more time to learn this relationship relative to more robust cues like verb biases or prosody (Trueswell & Gleitman, 2007).

1.4. Top-down plausibility as a reliable and accessible cue to sentence structure for children

Thus far, we have outlined two types of hypotheses: The first type argues that children's poor use of top-down cues during comprehension stems from their own processing limitations rather than the properties of the top-down cues themselves. The second type argues that the validity of the top-down cue determines whether children use it to infer phrase structure in real time, and moreover, that children do not struggle to integrate information from the non-dominant flow of information during processing. To distinguish between these two types of hypotheses, we need to disentangle cue validity from the direction of information flow (top-down vs. bottom-up). To do this, we explore whether children have difficulty using a top-down cue that is highly valid with respect to syntactic structure. We reasoned that if children continue to make less use of top-down information than adults, it would demonstrate that cue validity is not the *sole* reason for this processing bias, making the alternative hypotheses more appealing. If children, on the other hand, are essentially adult-like in their use of more valid, top-down cues, it would suggest that there is no general bias towards bottom-up processing, and the patterns that we see in the literature largely reflect the difficulty of learning each particular cue.

The cue that we focus on in this paper is the plausibility of one syntactic analysis given the affordances of the objects in the context. Specifically, we manipulate the plausibility of the instrument interpretation of a *with*-phrase by varying the object in the PP (e.g. "Tickle the frog with the *feather*" vs. "Tickle the frog with the *mirror*"). Prior adult work has shown that plausibility information can be rapidly used to guide structural interpretation of ambiguous phrases (Trueswell et al., 1994; McElree & Griffith, 1995; Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; cf. Clifton et al., 2003). For example, Garnsey et al. (1997) manipulated both verb bias and plausibility information in sentences like "The crooked politician *denied* the **accusation/election** <u>would change</u> things at all," and found that both manipulations impacted the earliest reading times on the critical noun (in bold) and subsequent spillover region (underlined).

To date, there are no published child studies that manipulate verb bias and plausibility to explore how these two cues affect their online comprehension. The closest study (Kidd, Stewart, & Serratrice, 2011) explored one cell of this design: all of the target sentences had a verb with a strong instrument-bias and a PP-object that made the instrument interpretation implausible (e.g. "Cut the cake with the candle"). The visual context consisted of two possible referents for the direct object noun (e.g. a cake with a candle and one without), a plausible instrument (e.g. a toy knife), and an implausible instrument (e.g. a large candle). Adults appeared to use plausibility or referential context cues to successfully arrive at the modifier (NP-attached) interpretation about 60% of the time (e.g. cutting the cake that had the candle by using the knife). In contrast, 5-year-olds appeared to rely on verb bias information, which led them to instrument (VP-attached) interpretations (e.g. cutting one of the cakes with the large candle).

While these findings are consistent with our first hypothesis (a developmental shift toward top-down processing), they have three limitations: First, because neither of the critical factors were manipulated, we cannot directly assess the effect of the top-down and bottom-up cues in each of the age groups. Second, because the two-referent contexts were used, we cannot determine whether the difference between the age groups is attributable to a developmental shift in the use of plausibility or merely to the well-documented shift in the use of referential context. Third, on every trial, there was a strong association between the verb and the unmentioned plausible instrument (see Koenig, Mauner, & Bienvenue, 2003). Both children and adults began looking at the plausible instrument (e.g. the knife) shortly after hearing the verb (e.g. cut) and continued doing so until the end of the sentence. Presumably, these early looks reflected either the semantic association between the words or the expectation that this instrument would be mentioned (see Nation, Marshall, & Altmann, 2003; Yuan, Fisher, Kandhadai, & Fernald, 2011). As a result, children showed little looking to the implausible (mentioned) instrument (e.g. the large candle) or the referent of the modified NP (e.g. the cake with the candle), and thus their eye-movements failed to provide insight into how they might be interpreting the prepositional phrase.

In a subsequent study, Bavin, Kidd, Prendergast, and Baker (2016) used a similar paradigm with older children ($M = 6.6$ years) and found a data pattern that one might be tempted to interpret as evidence of children's sensitivity to plausibility. In this study, children with and without Autism Spectrum Disorder (ASD) were shown scenes containing four images (e.g. a candle, a knife, a cake with candle, and a plain cake) while listening to various sentences. Children heard a set of unambiguous sentences with plausible instruments (e.g. "The girl will cut the cake with the knife") and a set of ambiguous sentences with implausible instruments (e.g. "The girl will cut the cake with the candle"). Bavin et al. (2016) found that children looked at the plausible instruments (e.g. the knife) more often in the unambiguous sentences—whereas, in the ambiguous sentences, children looked more at the modified NP (e.g. the cake with the candle). While this pattern could reflect school-age children's use of plausibility cues during syntactic processing, it could also be driven by the ability to recognize words and match them with objects in the scene (i.e. looking more at candles when candles are mentioned).[2]

The present study factorially manipulates verb bias and plausibility in a one-referent context (and does not provide a highly plausible instrument foil in the implausible conditions). We also eliminate the confound that is in the study by Bavin et al. (2016), ensuring that the referential scene makes each sentence equally ambiguous, which will distinguish effects of plausibility from effects of word recognition. As a result, our design addresses the limitations above, allowing us to study children's use of top-down cues in a context where eye-movements are likely to reflect children's moment-to-moment interpretation of the prepositional phrase.

2. Experiment 1

The goal of Experiment 1 was to determine the degree to which children and adults share intuitions about object affordances, and thus about the plausibility of using particular objects to perform an action. To test this, we conducted a forced-choice study in which participants were given three objects and asked questions like "What can you use to brush the cat?" The three objects were selected to be a plausible instrument, an implausible instrument, and a distractor for the mentioned action (e.g. using a comb, a toothbrush, or a rock for *brushing*). To preview our results, children largely agreed with adults about which instruments could be used for the actions, indicating that intuitions about object affordances are present by five years of age. These findings paved the way for Experiment 2 in which we investigated whether this knowledge of plausible instruments could be used as a top-down cue during moment-to-moment language comprehension.

2.1. Materials and methods

2.1.1. Participants

Twenty native English-speaking children 4;6–6;0 ($M = 5;2$, 8 males) and twenty native English-speaking adults from the Harvard University and Harvard Summer School communities participated in this experiment. Adult participants gave their consent and all participants under the age of 18 had parental consent to participate and assented to the procedure.

[2] In ambiguous conditions, we might have expected children to distribute their looks between the modified NP (the cake with the *candle*) and the implausible instrument (the large *candle*) after hearing the PP-object (the *candle*). However, children preferred to look at the modified NP. This preference is most likely driven by increased looks to the modified NP following the mention of the head noun (the cake). Thus, when children heard the PP-object, they did not need to look away from the modified NP to find a matching object, leading to the observed preference in looks.

2.1.2. Stimuli

2.1.2.1. Verb selection.
Our verbs were taken from Snedeker and Trueswell (2004). In their study, the verbs were selected by first conducting a corpus analysis and then refining the set of verbs with a ratings study. Specifically, adults completed sentence fragments containing a verb phrase up to the ambiguously attached preposition (e.g. "Tickle the bear with…"). Their responses were coded as *Modifier* continuations (Tickle the bear with *the devilish smile*), *Instrument* continuations (Tickle the bear with *a feather duster*), or *Other* (Tickle the bear with *great care*). Modifier and Instrument-biased verbs were those that elicited one continuation type (e.g. Instrument) three times more than the other (e.g. Modifier). The Equi-biased verbs fell somewhere in between, but on average showed no strong bias ($M = 52\%$ Instrument completions).

2.1.2.2. Object selection.
We conducted a ratings study with adults to identify instruments that varied in their plausibility. For each of our verbs, we made a list of six potential instruments for that action—three that seemed highly or moderately plausible and three that seemed implausible. For example, the verb *tickle* had a toothbrush, a stick, and a tissue as High Plausibility instruments, and a rock, a saltshaker, and a sock as Low Plausibility instruments. Sixty undergraduates from Harvard University then rated the plausibility (1–7) of sentences containing these verb-instrument pairings in an unambiguous syntactic frame (e.g. "Mary used the toothbrush to tickle the doll"). Plausible actions (7) were defined as being likely or easily imaginable, and implausible actions (1) were defined as being unlikely or hard to imagine. Verb-instrument pairs were manipulated between-participants such that each person only saw one instance of each verb. We then selected the best rated instrument from the three High Plausibility candidates and the worst rated instrument from the three Low Plausibility candidates to be used in critical sentences. This procedure resulted in plausibility differences between High Plausibility ($M = 4.46$, $SE = 0.12$) and Low Plausibility ($M = 2.65$, $SE = 0.10$) objects.

2.1.3. Procedure

Participants played a game that involved answering questions about the objects that were placed on an inclined podium directly in front of them. The podium had four shelves. The upper left shelf contained a stuffed elephant that stayed in place throughout the experiment. The remaining three shelves contained objects that changed from trial to trial depending on the verb: one was an object with High Plausibility as an instrument (based on the ratings study above), one had Low Plausibility, and the third was a Distractor object. The Distractors were selected randomly from the list of High and Low Plausibility objects. The objects' shelf positions (upper right, lower left, lower right) were counterbalanced across trials. The full set of stimuli can be found on the Open Science Framework (OSF; https://osf.io/wke5y/).

At the beginning of each trial, the experimenter labeled the elephant and then placed the three objects on the empty shelves, labeling them as they did so. After placing the objects, the experimenter would ask the following question using the critical verb: "What can you use to *VERB* the elephant?" Participants were told to answer by selecting one of the three objects. If the participant expressed uncertainty, the experimenter would repeat the question and ask them to make their best guess. At the end of the trial, the objects were removed, and the next trial began. All participants saw the same 24 trials in the same order. There were eight trials per verb bias condition, and they were presented in a blocked design. Participants first heard all eight Modifier-biased, then all eight Equi-biased, and finally all eight Instrument-biased trials.

2.1.4. Data coding and analyses

Offline coders indicated which object the participant selected (High, Low, or Distractor). No participants were excluded for failing to respond to the commands (6 out of 8 trials required). Two individual trials were removed due to experimenter error. For the analysis, selection of the High Plausibility object was coded as 1, and the other responses were coded as 0. We modeled this response with a generalized logistic mixed effects model using the *lme4* package in the R statistical computing environment (Bates, Maechler, Bolker, & Walker, 2014; R Core Team, 2020). Our model had a fixed effect of Age (Child = 0, Adult = 1), random intercepts for participants and items, and a random slope of Age for items. We also ran a model with additional fixed effects for Verb Bias (Instrument, Modifier, Equi) and the interaction of Verb Bias and Age; however, there was no improvement in model fit, $\chi^2(4, N = 40) = 2.28$, $p = .68$. As a result, we report the results from the simpler model with Age as the only predictor.

To evaluate significance, we adopt the convention of having an absolute value of z greater than 2 (Gelman & Hill, 2007) due to the debate about how to best calculate the appropriate degrees of freedom for the test statistics in mixed effects models (see Baayen, 2008). However, we also report the p-values as calculated by the *lmerTest* package, as both methods of evaluating significance arrive at the same conclusions.

2.1.5. Results and discussion

Table 1 presents the proportions of trials in which participants chose each of the objects and the standard errors for each age group. Both adults and children showed a strong preference for the High Plausibility instrument with probabilities of 93% and 83% respectively. There was an effect of Age ($\widehat{\beta} = 1.41$, $SE = 0.45$, $z = 3.13$, $p < .01$) such that adults selected the highly plausible object significantly more than children (Odds Ratio (OR) = 4.08, CI: [1.7, 11.5]). Critically, the model intercept was also significant ($\widehat{\beta} = 2.24$, $SE = 0.41$, $z = 5.42$, $p < .001$), demonstrating that children's selection of the High Plausibility instrument was well above chance (Odds = 9.44, CI: [4.4, 24.4]).

To directly compare instrument preferences between the two age groups, a Spearman's rank-order correlation was conducted on by-item averages, which found a strong positive correlation between adults' and children's use of the High Plausibility instrument

Table 1
Proportion of selecting each object type from Experiment 1. Participants were asked questions like "What can you use to *brush* the elephant?" and then selected one of three objects. This table shows the proportion of trials in which participants selected the High Plausibility object (e.g. a towel), the Low Plausibility object (e.g. a cup), or the Distractor object (e.g. a party hat) to complete the target action (e.g. *brush*). The averages for children and adults are presented separately with their respective standard errors in parentheses.

Age	High Plausibility	Low Plausibility	Distractor
Children	83% (1.3)	11% (0.8)	6% (0.8)
Adults	93% (0.8)	5% (0.7)	2% (0.5)

across items ($r_s(22) = .59$, $p = .002$; see Fig. 2 below).

Thus, Experiment 1 confirmed that both adults and children were sensitive to our manipulation of object affordances and recognized that the High Plausibility instrument could be used to carry out these actions. Moreover, there was a strong positive correlation between adults' and children's by-item instrument preferences. However, we did observe a significant difference in the overall performance across age groups, as adults selected the High Plausibility instruments (over the other two objects) more often than the children did (93% vs. 83% respectively). In the remainder of this section, we explore three hypotheses that might account for this difference.

First, we asked whether vocabulary knowledge might have played a role in children's lower performance. To address this question, we estimated how often each object label appears in the speech of 5-year-old children and their adult caregivers using the CHILDES corpus (MacWhinney, 2000) and the *childesr* package in R (Braginsky, Sanchez, & Yurovsky, 2020). Results indicated that the log frequency rates (per million words) for the object labels in adults' and children's speech are quite similar (see Fig. 3 below). We confirmed this finding with a linear mixed effects model that had fixed effects of Age (Child = 0, Parent = 1), Object Type (Distractor = 0, Target = 1), and their interaction, as well as a random intercept for items. The model revealed no significant effects of Age ($\widehat{\beta} = -0.54$, $SE = 0.34$, $t = -1.58$, $p = .12$), Object Type ($\widehat{\beta} = -0.24$, $SE = 0.28$, $t = -0.84$, $p = .40$), nor their interaction ($\widehat{\beta} = -0.11$, $SE = 0.40$, $t = 0.27$, $p = .79$).

Moreover, we also assessed whether there was a relationship between the frequency of an object label in children's speech and how often children selected that object during the task. To do this, we conducted a Spearman's rank-order correlation on the log frequency of the High Plausibility object labels with the proportion of children who selected that object. We found no significant relationship between the two variables ($r_s(23) = -.01$, $p = .97$; see Fig. 4 below), suggesting that vocabulary knowledge did not influence children's performance in Experiment 1.

Second, we considered whether the children knew the affordances of these objects as well as the adults. If children were less familiar with the objects' affordances, they may be less accurate in their judgments about which objects should be used to carry out each task. While we cannot rule out this possibility, there are features of our data that limit the scope of this concern. Critically, every child showed the expected pattern, selecting the High Plausibility objects far more often than the Low Plausibility or Distractor objects (see Fig. 5 below).

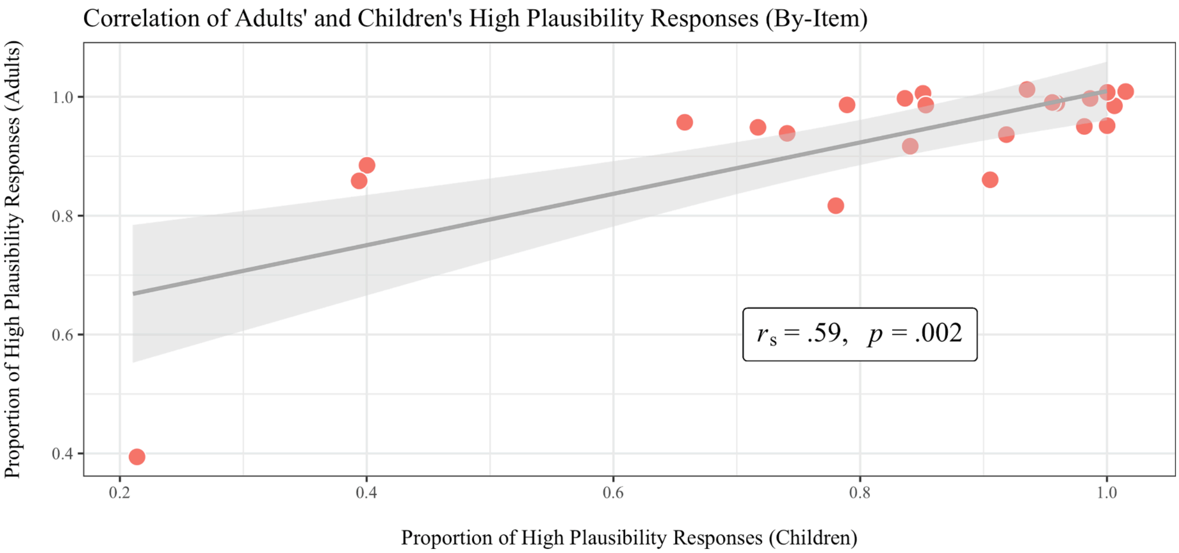

Fig. 2. *Correlation of by-item responses from Experiment 1.* Using a Spearman's rank-based order correlation of adults' and children's by-item proportions of High Plausibility responses, we found a positive correlation ($r_s(22) = .59$, $p = .002$) between adults' and children's use of plausible instruments across the items from Experiment 1.

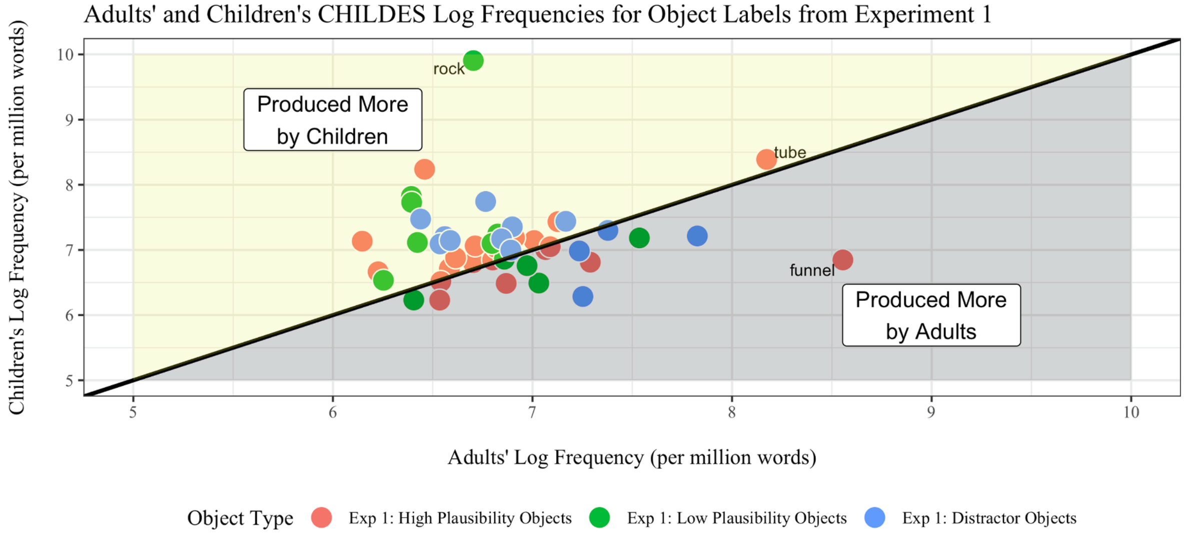

Fig. 3. *Log frequencies for object labels in adults' and children's (4;6-6;0) speech.* Log frequencies (per million words) for each object label were estimated from CHILDES. The diagonal line represents when frequencies are identical in adult and child speech. Most frequencies fall on the diagonal or above the diagonal, indicating the label is more frequent in children's speech. High Plausibility objects appear in red, Low Plausibility objects in green, and Distractors in blue. A linear mixed model revealed no significant differences across age group or object type. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Finally, in running the study, the first author noticed that there seemed to be two types of children that completed the task. Some children seemed to think of the study as a test, patiently waiting to hear the entire question before making their final actions. Other children had more fun with the task—often reaching for particular objects before the target sentence even began. These two approaches to the task anecdotally correspond to the child's performance. If the child was reaching for objects before hearing the sentences, they would often select the wrong item because they simply wanted to play with it. The child with the worst performance in Fig. 5 is an example of a child with this approach. In contrast, the children that focused on the task in order to answer the questions correctly, often waited and then remarked on whether or not there was a clear object that should be used. Children with this approach performed better—for example, the child with the best performance in Fig. 5 behaved in this way. Given these observations, it is possible that children's responses are not quite as tightly yoked to their top-down knowledge of plausibility as adults' responses in this task.

3. Experiment 2

In Experiment 2, we investigate whether children are able to recruit top-down information about the likelihood of the instrument interpretation given the objects in front of them to guide their structural interpretation of PP-attachment ambiguities. The prior literature has shown children failing to use top-down referential contexts during online comprehension. The present study investigates two hypotheses as to why children may not rely on the number of referents in the scene when inferring phrase structure. The first hypothesis is that children have a broad deficit in using *any* top-down cue during syntactic processing. The second is that children may struggle to use referential context cues because the cue itself has low validity and is a weak predictor of syntactic attachment in discourse. Experiment 2 addresses both hypotheses by using an alternative top-down cue, plausibility. There is good reason to think that plausibility has high cue validity and will be strongly predictive of syntactic structure in real-world language use (see Introduction). Experiment 1 showed that adults and children indeed share intuitions about likely and unlikely events and what makes an instrument plausible for a given action. Thus, in Experiment 2, we investigate children's use of top-down plausibility in a visual world eye-tracking experiment.

3.1. Materials and methods

3.1.1. Participants

Seventy-two native English-speaking children aged 4;6–6;0 ($M = 5;9$, 36 males) and seventy-two native English-speaking adults from the Harvard University community participated in the experiment. All adult participants gave their consent, and all participants under the age of 18 had parental consent to participate and assented to the procedure.

3.1.2. Stimuli

3.1.2.1. Selecting the target verbs and objects. Experiment 2 used 18 of the 24 verbs from Experiment 1 (6 in each verb bias class). For each verb, we selected a High Plausibility and a Low Plausibility instrument on the basis of another ratings study that we conducted

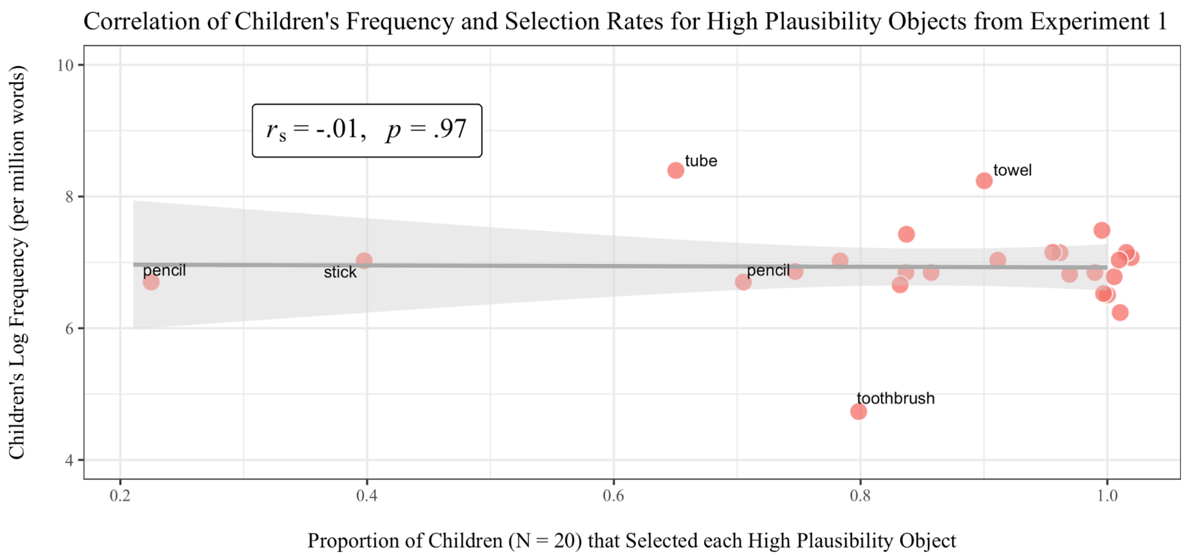

Fig. 4. *Correlating children's by-item object selection and label frequencies from Experiment 1.* This figure shows the Spearman's rank-based order correlation of the proportion of children that selected the High Plausibility object and the log frequency of that particular object. There was no significant relationship between the two variables for children in Experiment 1 ($r_s(23) = -.01$, $p = .97$).

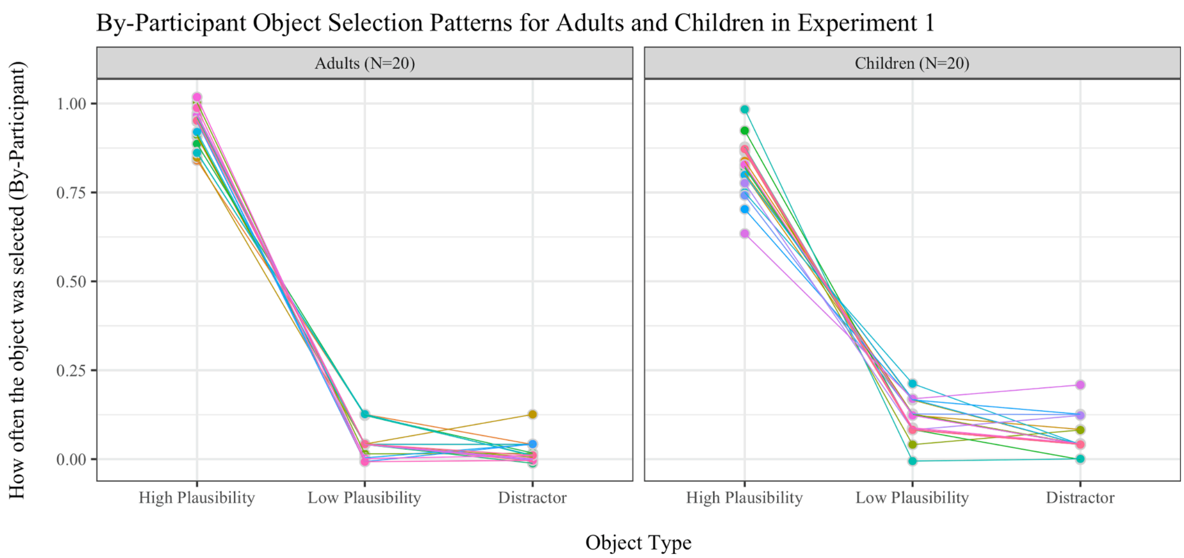

Fig. 5. *By-Participant object selection patterns for adults and children in Experiment 1.* This figure shows how often each participant selected the three object types during the task. Each colored line represents a single participant, and each dot represents the proportion of trials in which that participant selected that particular object type. All adults (on the left) and children (on the right) show the expected pattern of selecting the High Plausibility objects more often than the other two objects.

with a new set of 130 adults online. This new ratings study followed the same procedure as the ratings study from Experiment 1. The High Plausibility instruments all had ratings above 5.13 ($M = 5.96$, $SE = 0.11$) while the Low Plausibility instruments all had ratings below 3.21 ($M = 2.16$, $SE = 0.13$).

We confirmed the differences between our High and Low Plausibility instrument ratings with a non-parametric, rank-based linear regression, implemented using the *Rfit* package in the R statistical computing environment (Kloke & McKean, 2012; R Core Team, 2020). We analyzed by-item average ratings (1–7) and included fixed effects of Verb Bias (Instrument, Equi, Modifier-biased), Plausibility (High, Low), and their interaction. Simple difference coding was used for Plausibility (High = .5, Low = −.5). Treatment coding was used for Verb Bias, which resulted in two contrasts: one comparing Instrument-biased conditions to the Equi-biased and Modifier-biased conditions (Instrument = .66, Equi = −.33, Modifier = −.33) and the other comparing Equi-biased and Modifier-biased conditions (Equi = .33, Modifier = −.33). There was only a main effect of Plausibility ($\widehat{\beta} = 3.78$, $SE = 0.18$, $t = 20.58$, $p < .001$)

such that conditions with Low Plausibility instruments received lower ratings than those with High Plausibility instruments. There were no effects of Verb Bias nor any interaction.

In addition to norming the target verb-instrument pairs, we normed the verb-distractor pairings. We reasoned that having the same Distractor object in the High and Low Plausibility conditions would create a confound such that the High Plausibility object was always a much better instrument than any other object in the scene, whereas the Low Plausibility object would often be a worse instrument than the Distractor. We wanted to roughly equate the plausibility of the two potential instruments in the scene so that differences in the degree of competition from the Distractor object could not account for any observed differences in looks to the Target Instrument. For example, if the target sentence was "You can talk to the bunny with the rake," we included the mentioned rake (the Target Instrument) as well as another object that is an equally bad instrument for talking such as a piece of paper (the Distractor Instrument). In Experiment 2, all High Plausibility conditions received average ratings above 4.69 (Targets: $M = 5.96$, $SE = 0.11$; Distractors: $M = 5.46$, $SE = 0.12$), and all Low Plausibility conditions received average ratings below 3.79 (Targets: $M = 2.16$, $SE = 0.13$; Distractors: $M = 2.52$, $SE = 0.17$). A list of all conditions with their target instruments, their distractors, and their ratings can be found on OSF (https://osf.io/wke5y/).

To formally test the differences between Target and Distractor Instruments, we conducted another non-parametric, rank-based linear regression with fixed effects of Object Type (Target = .5, Distractor = −.5), Plausibility (High = .5, Low = −.5), and their interaction. There was a main effect of Plausibility ($\widehat{\beta} = 3.39$, $SE = 0.15$, $t = 23.09$, $p < .001$) and a significant interaction between Plausibility and Object Type ($\widehat{\beta} = 0.87$, $SE = 0.29$, $t = -2.96$, $p < .01$). Pairwise comparison revealed significant differences between Target and Distractor objects in the High Plausibility conditions ($\widehat{\beta} = 0.51$, $SE = 0.19$, $t = 2.63$, $p < .025$) but not the Low Plausibility conditions ($\widehat{\beta} = -0.35$, $SE = 0.25$, $t = -1.40$, $p = .17$). One implication of this interaction is that the effects of Plausibility may be inflated in the High Plausibility conditions, as participants may be more likely to look at the Target Instruments simply because they are more plausible than their Distractor Instrument. Thus, in our main analyses, we controlled for this interaction by adding an appropriate covariate to our mixed models—namely, the numerical difference between the plausibility ratings for the Target Instruments and their Distractors. To preview our findings, there were no differences in the overall pattern of effects when including this covariate (see the results in Section 3.2.2 below).

3.1.2.2. Creating the target trials. In Experiment 2, we manipulated both plausibility (High and Low) and verb bias (Instrument, Equi, and Modifier-biased) between participants. This was done to reduce the number of trials needed per participant and to eliminate possible effects of contrast across trials because prior work has found that children experience greater interference across conditions (Snedeker & Yuan, 2008). On each critical trial, there was an ambiguous PP-attachment ("You can tickle the frog with the feather") in which the *with*-phrase could potentially modify the verb (as an instrument) or the noun (as a modifier). Each trial had a set of four toys: 1) a *Target Instrument*: a full-sized object with high or low likelihood of being used to carry out the target action; 2) a *Target Animal*: a stuffed animal holding a smaller version of the Target Instrument; 3) a *Distractor Instrument*: another full-sized object with a similar likelihood of being used to carry out the target action as the Target Instrument; 4) a *Distractor Animal*: another animal with a small-scale version of the Distractor Instrument. An example of each condition is given in Table 2 below.

3.1.3. Procedure

Experiment 2 follows the procedure from Snedeker and Trueswell (2004). Participants sat in front of an inclined podium like that used in Experiment 1. The podium had a central hole so that a camera placed behind the podium could record participants' eye movements (see Fig. 6 below). Each trial started with the experimenter labeling and then placing the toy props on the podium. All toys were labeled, one-by-one (e.g. "Here, we have a frog, and a leopard, etc."). If a toy animal was holding an object, the experimenter introduced them separately (e.g. "a frog" and "a feather") rather than as a complex noun phrase (e.g. "a frog with a feather") in order to avoid priming participants to interpret the *with*-phrase as describing the Target Animal (e.g. the frog *that has* a feather").

After toy placement, the experimenter played a recording through external speakers, which instructed the participant to look at the center of the podium. The experimenter then played two recordings that told the participants how they should interact with the toys. The second recording was only played after the first action was completed. After the second action, all toys were removed, and the next trial began. When playing any of the recordings, the experimenter would step out of the participant's line of sight until they completed the trial. All actions were recorded on a second camera placed behind the participant and coded offline.

The outline of the entire procedure can be found in Table 3 below. Participants first saw three practice toy sets to get used to the task and to elicit both Instrument-type and Modifier-type actions (e.g. "Use the napkin to grab one of the cows", "Now, touch the cow that has a key"). After the practice, they saw a mix of target and filler toy sets. Each toy set had two sentence recordings associated with it. Target toy sets always played the target sentence first, followed by a filler sentence with no ambiguities (e.g. "Now, make the animals dance"). Filler toy sets used different toys and played two additional filler sentences. Children saw 5 filler toy sets and adults saw 15 sets, creating 16 and 36 filler sentences respectively. We used two presentation orders (forward, reverse) to reduce the effect of trial order on any given item. At the end of the experiment, all participants were debriefed and compensated.

3.1.4. Data coding and analyses

3.1.4.1. Action and eye gaze coding. All actions were categorized into five groups: 1) *No Action/Data Loss:* a refusal to act, an incomplete action (e.g. simply picking up an object), or an action that was not visible due to camera blockage; 2) *Target Instrument:* using the Target Instrument to complete the action on a stuffed animal; 3) *Mini Instrument:* using the smaller version of the Target

Table 2
Examples of target sentences and object sets for all conditions in Experiment 2. Each trial had four objects in the visual scene: a Target Instrument, a Target Animal, a Distractor Instrument, and a Distractor Animal. There were manipulations of plausibility (High, Low) and verb bias (Instrument-biased, Equi-biased, and Modifier-biased).

Condition	Ambiguous Sentence	Target Instrument	Target Animal	Distractor Instrument	Distractor Animal
High Instrument	You can poke the tiger with the stick	A stick	Tiger with a stick	A pen	Elephant with a pen
Low Instrument	*You can poke the tiger with the bathing suit*	*A bathing suit*	*Tiger with a bathing suit*	*A globe*	*Elephant with a globe*
High Equi	You can drag the fox with the string	A string	Fox with a string	A blanket	Cat with a blanket
Low Equi	*You can drag the fox with the ball*	*A ball*	*Fox with a ball*	*Puzzle pieces*	*Cat with puzzle pieces*
High Modifier	You can talk to the bunny with the headset	A headset	Bunny with a headset	A phone	Skunk with a phone
Low Modifier	*You can talk to the bunny with the rake*	*A rake*	*Bunny with a rake*	*A piece of paper*	*Skunk with paper*

Instrument to complete the action; 4) *Modifier:* performing the action without an instrument, e.g., using one's hand to tickle the frog that has a feather; 5) *Other:* performing an action that is not categorically interpretable as an Instrument or Modifier action. Additional details about each action type can be found in the Results section below.

All eye movements were recorded onto SONY DVCAM digital video tapes with audio-lock recording. The coding was performed on a digital VCR using frame-by-frame viewing. Each trial began at the onset of the target utterance and continued until the participant signaled that they had finished the target action. If the end of the action was not clearly defined, the trial ended two seconds after the end of the utterance. One coder watched the tape with the audio on, noting the onset and offset times for each target utterance. A second coder watched the same tape with no audio, noting the onset of each eye movement and the quadrant position of each fixation. Since both coders were blind to the position of the toys on the podium, the quadrant position data needed to be re-coded as looks to the *Target Instrument*, *Target Animal*, *Distractor Instrument*, or *Distractor Animal*. Frames could also be coded as looks to the *Central Camera* or as track loss (e.g. blinks). We excluded all blinks and other track loss data from our final eye gaze analyses, which removed 2.8% of all video frames from adults and 5.2% of all frames from children. We assessed intercoder agreement by having two coders categorize looks from 14 adults and 14 children using the five categories above. Intercoder agreement was 82.1% and 77.3% for adults and children respectively. If using a binary coding scheme (i.e. looks to the *Target Instrument* or *Other*), intercoder agreement was 94.2% and 91.6% for adults and children respectively. For additional information about intercoder agreement using this paradigm, we refer the reader to the Appendix in Snedeker and Trueswell (2004).

3.1.4.2. Statistical analyses for the action data. For our analysis of the action data, we eliminated trials with no actions or data loss. Target Instrument and Mini Instrument responses were coded as 1. Modifier and Other responses were coded as 0. We modeled our binary response using a generalized logistic mixed effects model with fixed effects of Plausibility (High, Low) and Verb Bias (Instrument, Equi, Modifier-biased), as well as their interaction. We used simple difference coding for Plausibility (High = .5, Low = −.5) and treatment coding for Verb Bias, which resulted in two contrasts: Instrument-biased vs. Equi-biased and Modifier-biased conditions (Instrument = .66, Equi = −.33, Modifier = −.33) and Equi-biased vs. Modifier-biased conditions (Equi = .33, Modifier = −.33). However, there were no significant effects of the second Verb Bias contrast, so we excluded those results from our discussion below. Snedeker and Trueswell (2004), who used the same verbs, also reported that the Equi and Modifier-biased conditions patterned together. Finally, we specified random intercepts only for participants and items (Baayen, Davidson, & Bates, 2008; Jaeger, 2008), as the maximal model did not converge (Barr, Levy, Scheepers, & Tily, 2013). We adopt the same conventions for evaluating significance as in Experiment 1.

3.1.4.3. Statistical analyses for the eye gaze data (cluster-based, permutation). To capture the full range of effects in our eye gaze data, we ran a series of cluster-based permutation analyses (Maris & Oostenveld, 2007; Wendt, Brand, Kollmeier, & Najbauer, 2014). Several studies using the visual world paradigm have used this statistical technique to analyze adults' and children's eye gaze data (e.g. Hahn, Snedeker, & Rabagliati, 2015; Borovsky, 2017; Weighall, Henderson, Barr, Cairney, & Gaskell, 2017; Yang, Chan, Chang, & Kidd, 2020). In general, these analyses follow a similar procedure: First, time-series data are divided into time-bins. At each time bin, a test is performed to obtain a test statistic for a given effect (e.g. Plausibility). If this statistic is greater than a predetermined threshold (e.g. z greater than 2), the time bin is considered significant. Then, if two or more significant time bins are adjacent, they become a "cluster," and their test statistics are summed to create a sum statistic. The original data are then randomly shuffled, and the previous steps are repeated. This shuffling procedure is repeated (e.g. 1,000 times or more) and each time, the largest sum statistic of any cluster is recorded. This iterative process creates a distribution of sum statistics that is expected by chance given the randomly shuffled data. The last step takes the original sum statistics for each cluster and compares them to this distribution. The cluster-level p-values are determined by where in the distribution the clusters' sum statistics fall. For example, if 3% of sum statistics in the distribution are equal to or more extreme than the observed statistic, the cluster would have $p = .03$.

For our cluster analyses, we first grouped the binary, frame-by-frame eye gaze data into 100 ms time bins (without averaging) such that each bin contained the data from 3 frames (e.g. the "300 ms" time bin had data points from 300, 333.34, and 366.67 ms). Note, these data were time-locked to the onset of the target object in the prepositional phrase (e.g. the onset of *stick* in "You can poke the tiger with the stick"). Thus, any negative time bins (e.g. −400 ms, −1000 ms) correspond to data that occurred before the onset of the target object whereas positive time bins (e.g. 500 ms, 2000 ms) correspond to data that occurred after. This initial binning procedure resulted in roughly 1,300 data points per time bin (i.e. 3 frames × 6 trials × 72 adults or children = 1,296 observations before exclusions due to

Fig. 6. *Stimuli and experimental set-up from Experiment 2.* The visual display for Experiment 2 with the Target Animal (the frog with the feather), Target Instrument (the normal-sized feather), Distractor Animal (the leopard with the candle), and Distractor Instrument (the normal-sized candle). The camera is positioned behind the central hole in the podium.

track loss). We chose to have multiple data points per time bin (rather than downsampling the data to a single observation or treating each observation as a time bin) in order to minimize issues surrounding model convergence and singular fits.[3] Our generalized logistic mixed effects models had the same fixed effects and contrast codings from our action analyses: Plausibility (High = .5, Low = −.5), Verb Bias (Instrument = .66, Equi = −.33, Modifier = −.33), and their interaction. For random effects, we had random intercepts for participants and items, and random slopes for Plausibility for items. We conducted our analyses on all 100 ms time bins between −1700 ms and 2000 ms. We selected this time window because it encompassed the time between the average onset of our target sentences (−1675 ms) and the average onset of participants' actions (2000 ms). At each time bin, the models provided z-values for Plausibility, Verb Bias, and their interaction. If the z-value was greater than the absolute value of 2 (roughly the 97.5th quantile for the z distribution at alpha = .05 for two-tailed tests), we labeled that time bin as significant. We then followed the general clustering procedure (outlined above) by permuting item labels and repeating this process 1,000 times per analysis. All statistical analyses can be found in our annotated R scripts on OSF (https://osf.io/wke5y/).

3.2. Results and discussion

3.2.1. Action data

The action data clearly demonstrated that both adults and children understood our instructions. In both age groups, nearly all of the actions were consistent with the verb and were performed on the Target Animal. For example, after hearing "You can talk to the bunny with the headset," participants would turn towards the bunny and start speaking. Actions that did not follow this pattern were coded as *Other* (2.6% of all actions for adults, 2.1% for children). The remaining actions were all consistent with a grammatical interpretation of the ambiguous sentences; they simply varied in whether (and how) participants used an instrument to complete the action. *Modifier* actions were those performed on the Target Animal that did not involve an instrument (43.2% for adults, 46.9% for children). For example, after hearing "You can drag the fox with the string," many participants grabbed the fox (who was holding a tiny string) by the arm and dragged him across the stage. *Target Instrument* actions, on the other hand, involved using the Target Instrument to perform

[3] At each time point, the mixed model was assessed for issues related to model convergence and singular fits. If there were any warnings, we did not use the model estimates—rather, we took the model estimates from the last time point that converged properly. If the model was at the very first time point, then we set its estimates to zero. This approach is overly conservative; however, it prohibits potentially misspecified models from initiating a cluster or prematurely ending an ongoing cluster. This procedure was adopted for the main mixed models, as well as for all of the models computed during resampling.

Table 3
Demonstration of all trials for participants in the Low Plausibility, Instrument-biased condition in Experiment 2. Adults and children observed the same target sentences (highlighted in gray) depending on which plausibility and verb bias group they were in. All filler trials were identical regardless of experimental group; however, adults saw more filler trials than children. In each trial, participants heard two sentences, which are presented in the last two columns.

Adult Trials	Child Trials	First Sentence in Item Set	Second Sentence in Item Set
Demo 1	Demo 1	You can use the napkin to grab one of the cows.	Now, touch the cow that has a key.
Demo 2	Demo 2	You can use the sticker to pet the dog.	Now, shake hands with the cat who is wearing a dress.
Demo 3	Demo 3	You can sniff the panda.	Now, say goodnight to the horse.
Target 1	**Target 1**	**e.g., You can brush the horse with the cup.**	**Now, make the animals crawl onto the truck.**
Filler 1	Filler 1	You can kiss the bunnies.	Now, have them sit down on the eraser.
Filler 2		You can squeeze the animals.	Now, make them play leapfrog.
Filler 3		You can toss the tissue in the air.	Now, give the raccoon a flower.
Target 2	**Target 2**	**e.g., You can clean the pig with the chair.**	**Now, make him lick the button.**
Filler 4	Filler 2	You can have the zebra draw with the chalk.	Now, let the rooster have a turn.
Target 3	**Target 3**	**e.g., You can cover the raccoon with the pen.**	**Now, make the animals dance.**
Filler 5	Filler 3	You can drop the duck on the horse.	Now, make the horse throw the ball.
Filler 6		You can place the animals on top of the plate.	Now, dump them off.
Filler 7		You can make the cats slide down the bottle.	Now, give them both a drink of soda.
Target 4	**Target 4**	**e.g., You can feed the dog with the painting.**	**Now, make him bark.**
Filler 8	Filler 4	You can make the animals swim in the cup.	Now, shake them dry.
Target 5	**Target 5**	**e.g., You can poke the tiger with the bathing suit.**	**Now, place him on the globe.**
Filler 9	Filler 5	You can hide one of the bears in the washcloth.	Now, push the sponge off the shelf.
Filler 10		You can lift one of the foxes off the shelf.	Now, have him kick the other fox.
Filler 11		You can slap the lion.	Now, make him stand on his head.
Filler 12		You can pretend the rock is a spaceship.	Now, let it land on the shelf.
Target 6	**Target 6**	**e.g., You can tickle the bear with the mirror.**	**Now, have both animals jump over the couch.**
Filler 13		You can make the animals wrestle.	Now, lay the lion down on the book.
Filler 14		You can give the spoon to the panda.	Now, help him dig for buried treasure.
Filler 15		You can wave the post-it.	Now, give it to one of the tigers.

the expected action (50.8% for adults, 45.7% for children). For example, a child might wrap the big string around the fox (who was holding a tiny string) and use it to drag him across the stage. Lastly, *Mini Instrument* actions involved using the tiny object attached to the Target Animal to perform the action on that animal (3.3% for adults, 5.1% for children). For example, participants might grab the tiny string attached to the fox, and then pull it in order to drag the fox across the stage. Note, prior research has found that *Mini Instrument* and *Target Instrument* actions pattern together (see Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008). Thus, we collapsed these measures into a single measure of Target Instrument actions for our statistical analyses.

In our mixed model for adults, there were main effects of Plausibility ($\widehat{\beta} = 3.08$, $SE = 0.48$, $z = 6.38$, $p < .001$) and Verb Bias ($\widehat{\beta} = 2.81$, $SE = 0.56$, $z = 5.03$, $p < .001$) with High Plausibility and Instrument-biased conditions eliciting more instrument actions. There was also a significant interaction of Plausibility and Verb Bias ($\widehat{\beta} = -3.48$, $SE = 0.95$, $z = -3.65$, $p < .001$) such that the Plausibility effect for Instrument-biased verbs was smaller because these verbs resulted in more Target Instrument actions regardless of the instrument plausibility.

Children also showed main effects of both Plausibility ($\widehat{\beta} = 2.19$, $SE = 0.62$, $z = 3.51$, $p < .001$) and Verb Bias ($\widehat{\beta} = 4.51$, $SE = 0.91$, $z = 4.96$, $p < .001$) with High Plausibility and Instrument-biased conditions eliciting more Target Instrument actions. However, unlike adults, children showed no significant interaction.

In adults, we found that the effect sizes for Verb Bias and Plausibility were relatively similar—however, in children, the effect size for Verb Bias appeared to be much larger than that for Plausibility. To quantify how much Plausibility and Verb Bias effects differed in children and in adults, we compared the odds ratios (OR) for both effects in each age group separately. In adults, Plausibility (OR = 21.75, CI: [8.44, 56.04]) and Verb Bias (OR = 16.63, CI: [5.56, 49.75]) had a comparable effect on the likelihood of an instrument action. However, in children, we saw that the influence of Plausibility (OR = 8.95, CI: [2.63, 30.43]) was much smaller than the

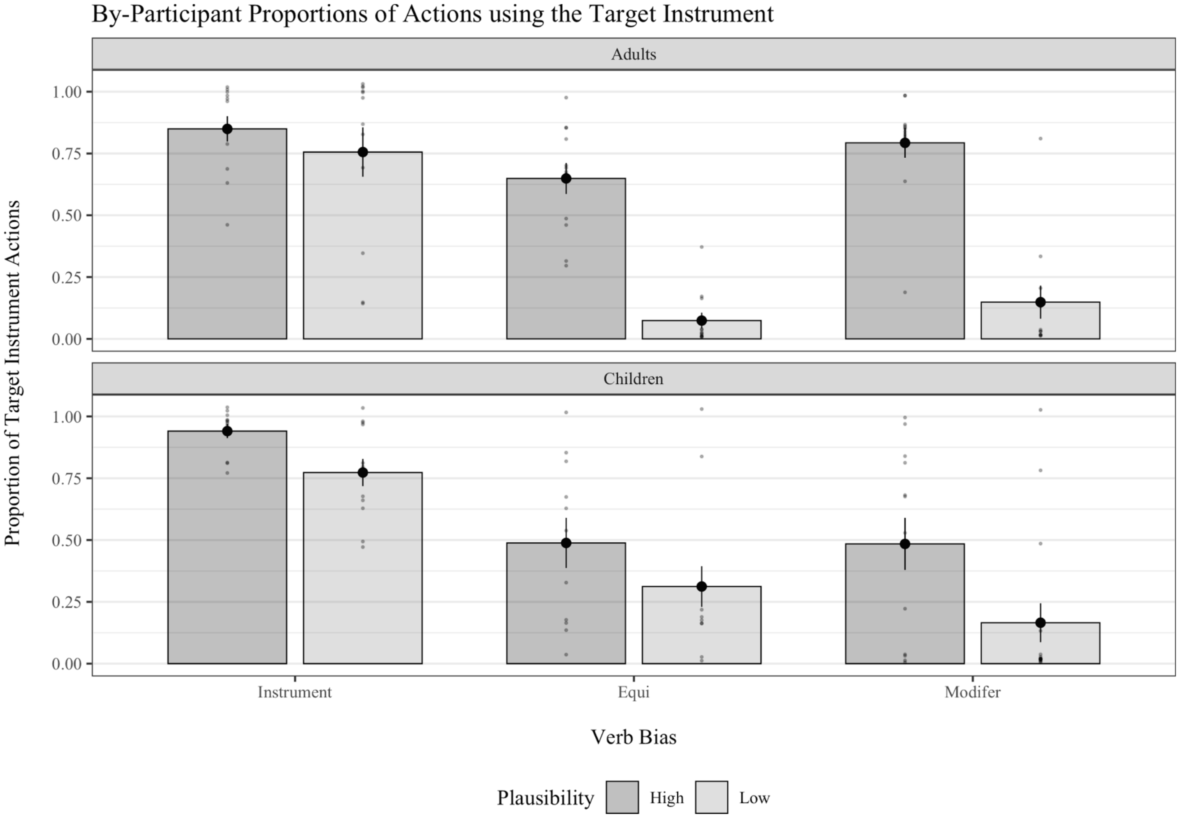

Fig. 7. *Action response data from Experiment 2.* Adults (top) show an overwhelming preference for Target Instrument actions regardless of verb bias. They also entertain Low Plausibility objects when the verb is instrument-biased. Children (bottom) show similar preferences for Target Instrument responses; however, the majority of their instrument actions are restricted to the Instrument-biased condition. Each data point represents how often a single participant performed an action using the Target Instrument or Mini Instrument (i.e. the by-participant average proportions of Target Instrument actions).

influence of Verb Bias (OR = 91.32, CI: [15.37, 542.65]). We confirmed the significance of this difference between Plausibility and Verb Bias in children by calculating a z-value for the difference between the two model estimates in the child analysis and the corresponding p-value for that difference ($z = 2.11$, $p < .05$).

Thus, the action data demonstrate that both adults and children are sensitive to object affordances in their ultimate interpretation of PP-attachment ambiguities. Adults presented with High Plausibility objects largely interpreted the instructions as requiring an instrument; however, adults often interpreted Low Plausibility objects as instruments when the verb was Instrument-biased. Children's actions were strongly influenced by Verb Bias but were also affected (to a smaller degree) by Plausibility (see Fig. 7).

3.2.2. Eye gaze cluster analyses

For the effect of Plausibility, we observed a significant cluster for adults such that the High Plausibility conditions elicited more looks to the Target Instrument, collapsing across Verb Bias (see Fig. 8, top panel). This cluster emerged between −200 ms to 1000 ms relative to the onset of the PP-object (Summed z-Statistic = 38.14, $p = .001$).[4] The timing of this cluster suggests that adults deployed plausibility information predictively, showing increased looks to the highly plausible Target Instrument 200 ms before the object was explicitly mentioned in the sentence. For children, there were no significant clusters for Plausibility, as no two consecutive time bins met the threshold to be considered a cluster in the analyses (see Fig. 8, bottom panel). However, there seems to be a slight increase in children's looks to the Target Instrument in the High Plausibility conditions towards the end of the trial (around 1100 ms), perhaps suggesting a delayed sensitivity to top-down plausibility cues and/or the emergence of Plausibility effects in children's actions. We will return to this data pattern in the discussion below.

[4] As mentioned in Section 3.1.2.1, we conducted a set of analyses with an additional covariate to capture the difference in plausibility ratings between the Target Instruments and their Distractors from Experiment 2. When including this covariate, the magnitude of adults' Plausibility effect was smaller (Summed z-Statistic = 28.51), and the duration of the cluster was shorter (−100 ms to 900 ms)—however, the main pattern of significance remained the same.

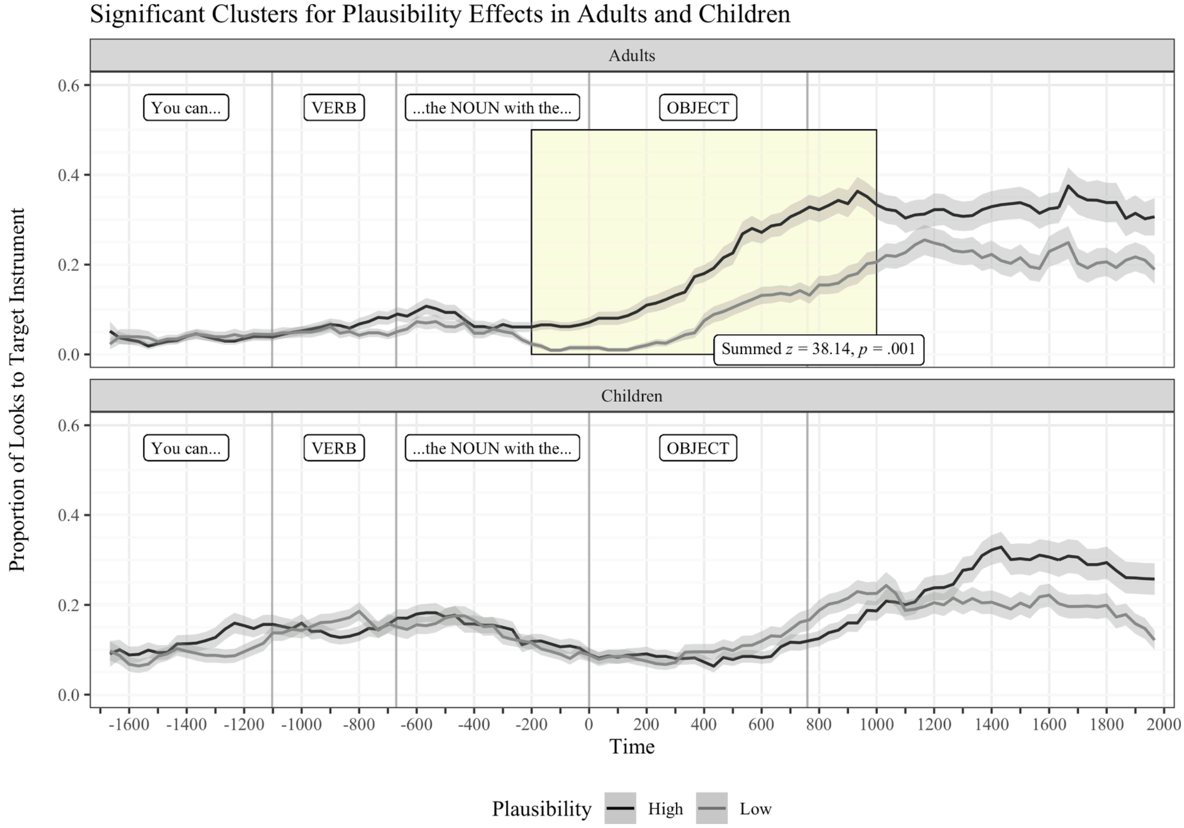

Fig. 8. *Results of the cluster analysis for Plausibility effects.* Adults (top) show a large time window in which the High Plausibility condition resulted in increased looks to the Target Instrument (-200–1000 ms post-PP object onset). Children (bottom) showed no significant clusters for Plausibility.

For the effect of Verb Bias, there were no significant clusters for adults—again, this was because no two consecutive time bins met the threshold to be considered a cluster in the analyses. The lack of significant clusters in adults suggests that they prioritize a different source of information over verb bias cues (namely, plausibility) when comprehending these syntactic ambiguities (see Fig. 9, top panel). For children, one significant cluster emerged between 300 and 1000 ms after the PP-object onset (see Fig. 9, bottom panel), which is quite fast given that humans take ~200 ms to plan and launch an eye movement (Allopenna et al., 1998; Hallett, 1986; Salverda, Kleinschmidt, & Tanenhaus, 2014). During this window, children looked to the Target Instrument more for Instrument-biased verbs than for Modifier and Equi-biased verbs combined (Summed z-Statistic = 23.16, p = .012).[5] This significant cluster in children seems to suggest that bottom-up information from the verb (and its lexical bias towards VP or NP-attachment) drives children's pattern of looking for a potential instrument upon hearing the PP-object.

Finally, for the interaction between Verb Bias and Plausibility, we observed a significant cluster for adults between 800 and 1700 ms after the PP-object onset (Summed z-Statistic = 26.11, p = .002).[6] The timing of this cluster roughly corresponds to the average offset of the target sentences (750 ms), suggesting that adults may start integrating both top-down and bottom-up cues prior to making their ultimate interpretations. To unpack this interaction, we conducted pairwise cluster-based permutation tests that looked for effects of Verb Bias in the High and Low Plausibility conditions separately. There were no significant clusters for Verb Bias in the High Plausibility conditions; however, in the Low Plausibility conditions, there was a significant cluster between 900 and 1700 ms (Summed z-Statistic = 22.62, p = .006) during which adults showed increased looks to the Target Instrument in Low Plausibility, Instrument-biased conditions (see Fig. 10, top panel). For children, there were no significant clusters for the interaction, as no two consecutive time bins met the threshold to be considered a cluster in the analyses (see Fig. 10, bottom panel).

In sum, the findings from our eye gaze data were broadly consistent with those from the action data. Adults were sensitive to plausibility information from very early on in the sentence but also integrated verb bias information towards the end of the utterance. In particular, adults made greater use of verb bias cues when the instrument plausibility was low (e.g. You can poke the tiger with *the*

[5] The analysis with the *plausibility difference* covariate also found a significant cluster for Verb Bias in children between 300 and 1000 ms—although, the magnitude of the effect was smaller (Summed z-Statistic = 21.33).

[6] The analysis with the *plausibility difference* covariate also found a significant Interaction cluster in adults between 800 and 1700 ms—although, the magnitude of the effect was smaller (Summed z-Statistic = 25.63).

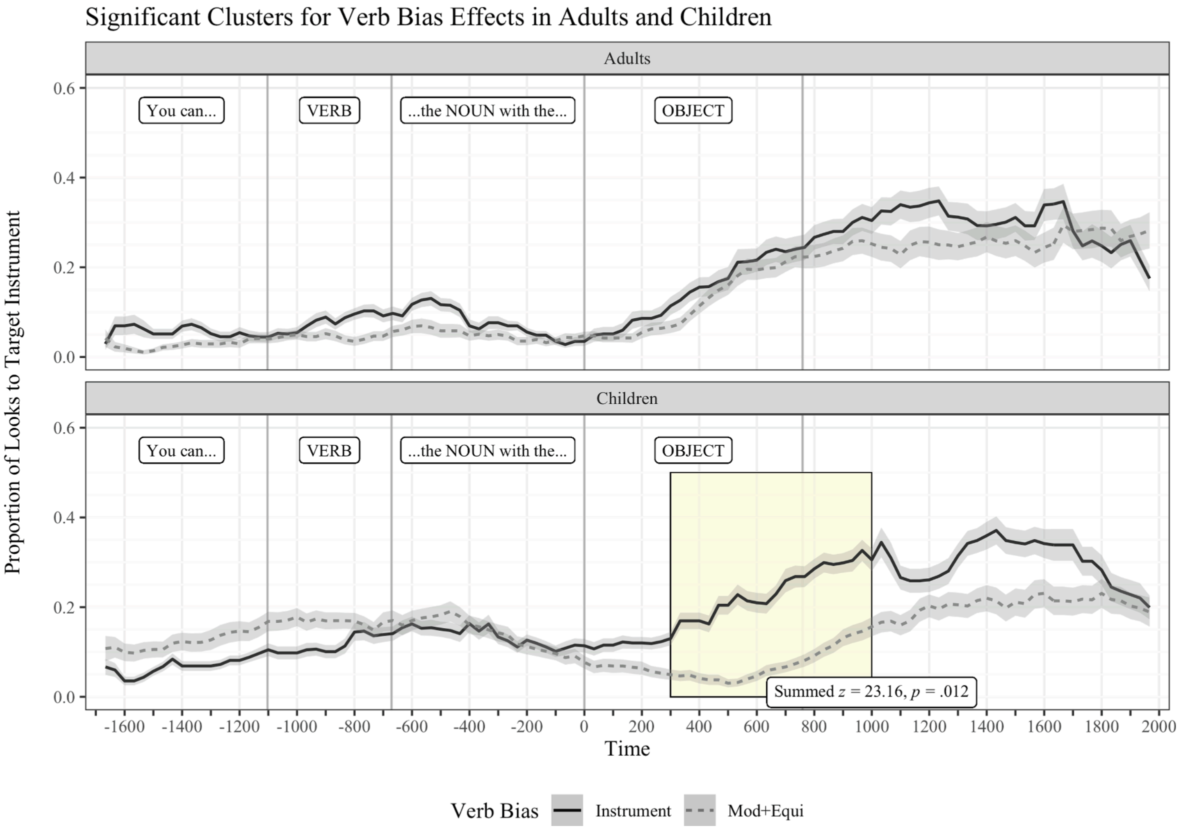

Fig. 9. *Results of the cluster analysis for Verb Bias effects.* Adults (top) show no significant clusters for Verb Bias. Children (bottom) show one cluster in which the Instrument-biased condition resulted in more looks to the Target Instrument (300–1000 ms post-PP object onset).

bathing suit). This pattern closely resembles the pattern in the action data, which showed that adults were more likely to use Low Plausibility objects to carry out an Instrument action (e.g. poking the tiger *by using the bathing suit*) when the verb was instrument-biased.

In contrast, children's gaze patterns were strongly shaped by verb bias but showed no sensitivity to plausibility (even at the end of the trial). These results are somewhat surprising given that the children were affected by plausibility, albeit to a smaller extent, when making their final actions. We see two potential explanations for this inconsistency: First, the lack of Plausibility effects in children's eye-movements could be attributed to small effect sizes and/or the intrinsic noisiness of eye gaze data relative to action data. Second, it is possible that the Plausibility effects are limited to a subset of the items. Our sentences varied considerably in the length of the final word. If children are slow to calculate plausibility, and quick to act once the utterance is over, then we might only expect to see plausibility effects in the longer sentences. We explore this possibility below.

3.2.3. *Exploratory eye gaze cluster analyses*

The results from our primary eye gaze analyses raised two questions, both of which warranted further investigation: The first question is why did adults' sensitivity to verb bias information only emerge *after* their sensitivity to plausibility? Verb bias is often conceptualized as stored information culled through years of experience with language, which is accessed during word recognition in an automatic, bottom-up fashion regardless of its situational utility. Plausibility, on the other hand, is a top-down inference based on a situation model that must be constructed as the sentence unfolds. Thus, if verb bias is a robust cue to structure, we should expect adults to have access to this information shortly after hearing the verb and well before accessing top-down cues like plausibility. However, in our analyses, we see no evidence of an effect of Verb Bias early in the trial (see verb regions in Fig. 9 above).

One potential explanation for this data pattern is that, over the course of the study, adults learned to use the plausibility cues predictively. Our study used between-participant manipulations of both plausibility and verb bias (i.e. if a participant heard an Instrument-biased, High Plausibility sentence on one trial, they also heard Instrument-biased, High Plausibility sentences on all subsequent trials). Thus, as the experiment progressed, participants might have come to anticipate that they would hear sentences with meanings parallel to those that they have heard before (e.g. act on an animal with an instrument). If plausibility was the dominant force shaping their ultimate interpretation, this could result in adults searching for plausible instruments in the high plausibility conditions before encountering the ambiguous target word. Our primary analyses suggest that this is happening: the eye gaze data show that adults are sensitive to plausibility well before they even hear the target word (see adults' Plausibility effects emerging 200

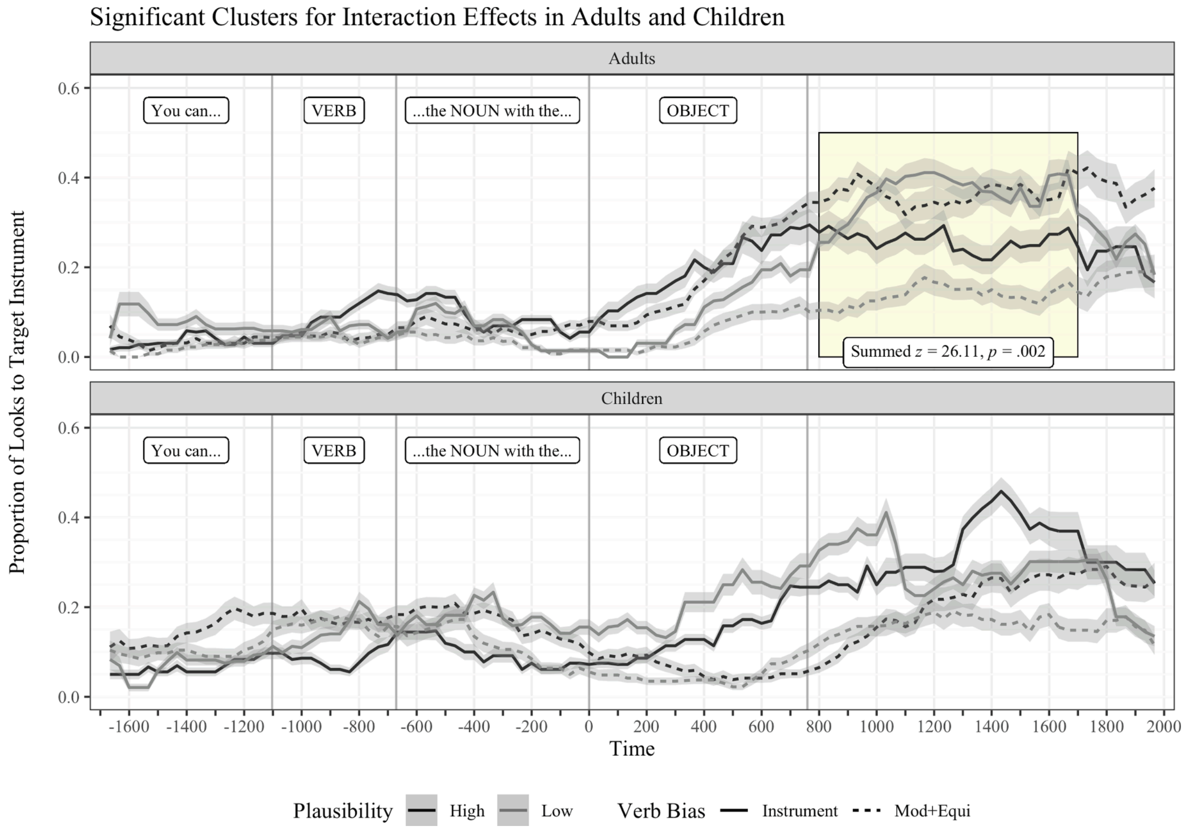

Fig. 10. *Results of the cluster analysis for Interaction effects.* Adults (top) show one significant cluster between 800 and 1700 ms. This interaction was driven by increased looks to the Target Instrument in the Low Plausibility, Instrument-biased conditions between 900 and 1700 ms. Children (bottom) did not show any significant clusters for the interaction.

ms before the onset of the PP-object in Fig. 8). On this account, we would predict that adults' gaze patterns would change over the course of the study. More specifically, in the early trials, eye-movements might be primarily sensitive to verb bias cues, particularly early in the sentence (given the cue's automaticity). In contrast, over time, Plausibility effects might emerge earlier and become more robust, as participants develop expectations about the sentences that they will hear.

To investigate these possibilities, we conducted a series of post-hoc cluster-based permutation tests with mixed models, which compared adults' looking patterns in the first three trials against their patterns in the last three trials. Specifically, we looked at how effects of Trial Order interacted with Verb Bias and then with Plausibility separately (see Fig. 11 below). The first model included fixed effects of Verb Bias (Instrument = .66, Equi = −.33, Modifier = −.33), Trial Order (First Half = .5, Second Half = −.5), and their interaction, as well as random intercepts for participants and items. The second model included fixed effects of Plausibility (High = .5, Low = −.5), Trial Order, and their interaction, as well as random intercepts for participants and items and a random slope for Plausibility for participants. For Verb Bias, the model revealed a significant interaction between Verb Bias and Trial Order between 700 and 1700 ms (Summed z-Statistic = 33.25, p = .001), indicating that adults relied more on verb bias cues in the first half of the experiment. For Plausibility, the model revealed a significant interaction between Plausibility and Trial Order between 100 and 1000 ms (Summed z-Statistic = 31.48, p = .003), indicating that adults relied more on plausibility cues in the second half of the experiment in early time windows.

These effects are consistent with three of the four predictions that we made above: specifically, we found that 1) Plausibility effects increased with experience; 2) Plausibility effects emerged early and only in the second half of the experiment; and 3) Verb Bias effects were larger in the first three trials relative to the last three. However, we found no evidence for our fourth prediction: we expected to find an early emerging Verb Bias effect in the adults in the first half of the experiment—and instead, we found a Verb Bias effect that only became reliable about 700 ms after the onset of the object in the prepositional phrase.

There are several limitations to this data set that leave us reluctant to interpret the absence of an early Verb Bias effect as strong evidence *against* adults' rapid use of verb-based information in this task: First, this absence is a null effect in an exploratory analysis that has less power than the primary analyses (because only half of the trials are being used). There are good reasons to assume that the expected effect (i.e. early emerging Verb Bias effects) would be smaller than effects that emerge later in the trial. In general, predictive effects are often smaller than effects that appear at (or after) the word itself—and the effects that emerge shortly after hearing a word are typically smaller than the effects that emerge much later. Second, in order to execute a look to the Target Instrument, a participant

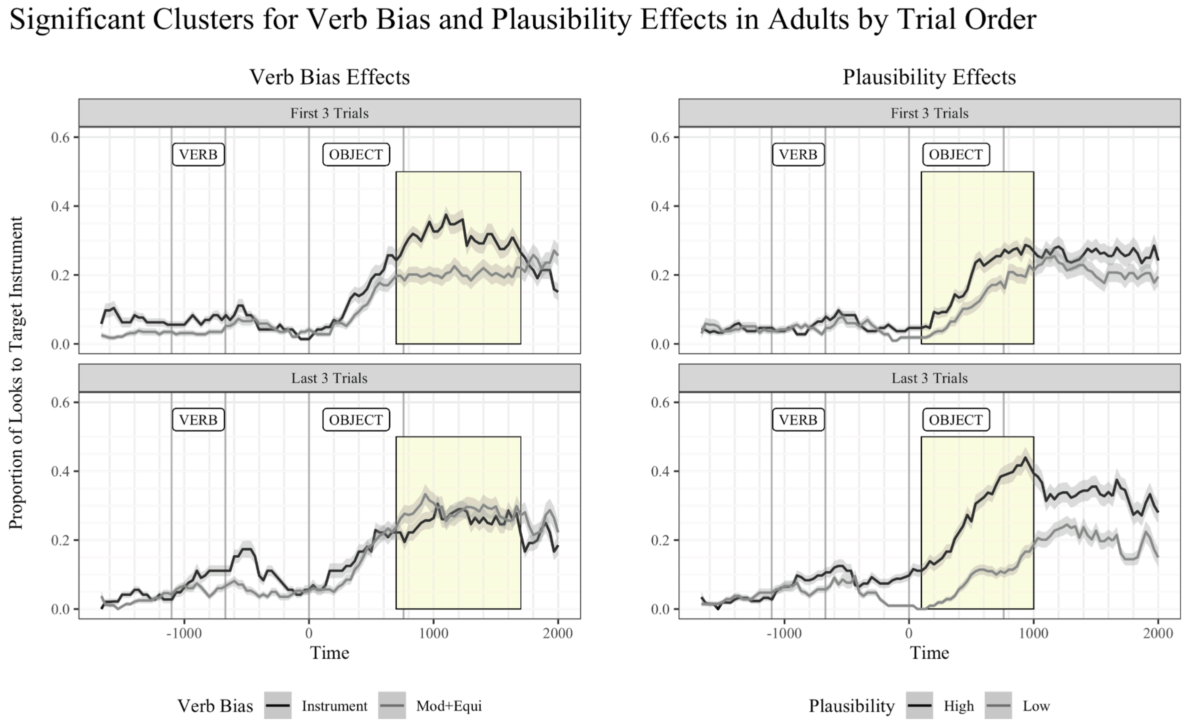

Fig. 11. *Results of the cluster analyses for adults' effects across Trial Order.* Adults' Verb Bias effects (left) show one significant cluster between 700 and 1700 ms (Summed z-statistic = 33.25, p = .001). This interaction was driven by increased looks to the Target Instrument in the Instrument-biased conditions in the first three trials. Adults' Plausibility effects (right) also showed a significant cluster for the interaction between 100 and 1000 ms, suggesting that adults learn to use plausibility cues later in the experiment (Summed z-statistic = 31.48, p = .003).

in our task needs more information than simply the prediction that an instrument will be mentioned (i.e. the information obtained from instrument-biased verbs). Specifically, looks to the Target Instrument require that the participant has inferred that *this particular object* is likely to fill that predicted instrument role. In our study, this inference might be hard for adults to make before hearing the target PP-object because there are two different objects in the display that were chosen to be equally good (or poor) instruments for the action in question. Thus, in the Low Plausibility condition, participants might expect an instrument (due to the instrument-biased verb) but see no viable candidate—and in the High Plausibility condition, they might expect an instrument but fail to shift their gaze because they are torn between two equally plausible targets.[7]

Taken together, these findings support the idea that adults' relative insensitivity to verb bias cues during comprehension is partly due to the fact that, across trials, adults learn to rely more on the plausibility cues that guide their ultimate interpretations. Moreover, these findings motivate the need to replicate these effects in a design that is not between-participants. This *cue learning* hypothesis would predict that, in studies using within-participant designs, the effects of Plausibility will emerge later in the trial (i.e. be used less predictively) but will continue to dominate adults' final interpretations of the utterances.

The second question raised by our primary analyses is why are children sensitive to top-down plausibility cues when making their final actions but not during comprehension? As we mentioned above, plausibility is an inference based on a situation model that must be constructed as the sentence unfolds. This process may be more effortful, less automatic, and more variable across contexts. Thus, it may take children longer to construct this information.

To explore this hypothesis further, we conducted a post-hoc analysis of our own data. Specifically, we compared target sentences that had longer and shorter ambiguous regions (e.g. You can poke the tiger *with the bathing suit* vs. *with the stick*) to determine whether the longer sentences resulted in greater use of plausibility in children. It is important to note that, regardless of the length of the ambiguous region, all sentences were disambiguated at the onset of the object in the prepositional phrase. So, although children must process more phonetic material in the longer sentences, the amount of information (relevant to their syntactic analyses) to process remains the same across conditions. Thus, the only difference between longer and shorter sentences is that when the sentence has a

[7] Note, these competing pressures were not present in Snedeker and Trueswell (2004) because the Distractor Instruments were not matched in plausibility to the Target Instruments. At first glance, it may seem surprising that children in our study show relatively early use of verb bias information (given the limitations above). But critically, our primary analyses show that children's Verb Bias effect emerges *after* the onset of the PP-object, and thus, *after* the relevant instrument has been disambiguated. In children, unlike adults, verb bias dominates their ultimate interpretations, meaning that cumulative learning across trials for children should make the Verb Bias effects in later trials earlier and larger than those in the beginning of the experiment.

longer PP-object, children have more time to consider the plausibility of the instrument interpretation before acting.

To do this analysis, we conducted a median split to categorize our items based on the total duration from the point of ambiguity (i.e. with…) until the sentence offset. We then took the group with the shorter sentences ($M = 2.28$ s, $SE = 0.17$) and compared it to the group with the longer sentences ($M = 2.61$ s, $SE = 0.25$). Each duration group was balanced in terms of verb bias and plausibility conditions.[8] To compare the groups, we conducted another cluster-based permutation test using a mixed model that had fixed effects of Plausibility (High = .5, Low = −.5), Verb Bias (Instrument = .66, Equi = −.33, Modifier = −.33), and Sentence Duration (Longer = .5, Shorter = −.5), as well as all the two-way and three-way interactions. This model also had random intercepts for participants and items.

Results confirmed the original children's Verb Bias effect, as a significant cluster emerged between 400 and 1000 ms (Summed z-Statistic = 19.94, $p = .018$). Additionally, in a later time window, there was a significant two-way interaction between Plausibility and Sentence Duration, such that the longer sentences showed an effect of Plausibility emerging very late in the trial (1200–1900 ms, Summed z-Statistic = 26.64, $p = .005$) whereas the shorter sentences did not (see Fig. 12 below). This analysis tentatively suggests that children may simply require more time than adults to deploy top-down information. In many contexts, this time is not available, and thus children often show weaker effects of context than adults, or no effect at all.

4. General discussion

The present study explored two hypotheses about why young children fail to use top-down referential context cues when inferring the phrase structure of ambiguous sentences. Our first hypothesis was that children have a broad deficit in using *any* top-down cue to inform lower-level syntactic processes, perhaps due to architectural or processing limitations. Our second hypothesis was that

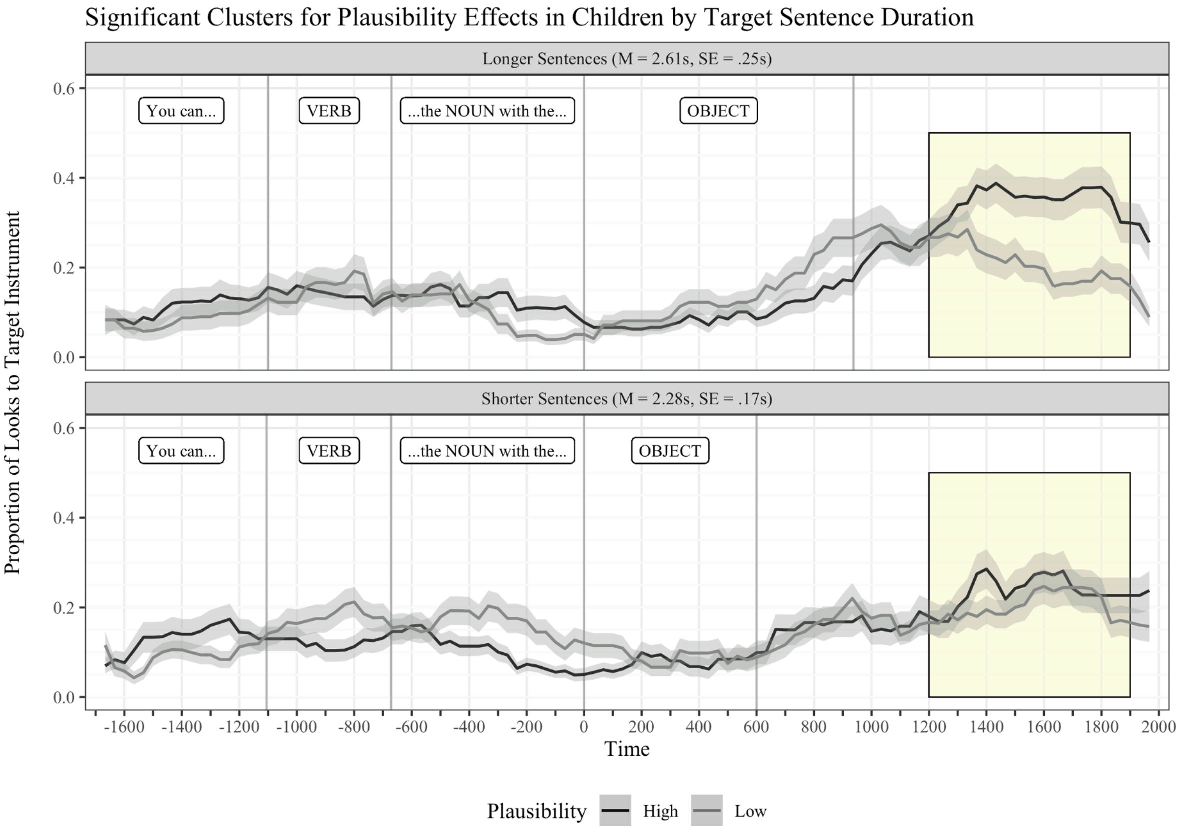

Fig. 12. *Results of the cluster analysis for children's Plausibility effects in longer and shorter items.* There was a significant cluster for the interaction between Plausibility and Sentence Duration between 1200 and 1900 ms (Summed z-statistic = 26.64, $p = .005$). This effect was driven by a difference between plausibility conditions in the longer sentences (top) but not in the shorter sentences (bottom).

[8] Each group had 3 High and 3 Low Plausibility items with instrument-biased verbs. Similarly, there were 6 High and 6 Low Plausibility items with modifier/equi-biased verbs (when combined)—the shorter sentences had 5 High and Low Plausibility items with equi-biased verbs, whereas the longer sentences had 5 High and Low Plausibility items with modifier-biased verbs. The number of observations in each time bin did not significantly differ across conditions.

children's difficulties in using top-down cues like the referential context is solely a reflection of cue validity—rather than a reflection of a broader developmental difference in top-down processing abilities. We tested these hypotheses by assessing children's abilities to use another top-down cue to phrase structure that has high cue validity. Specifically, we manipulated the plausibility of the instrument interpretation for ambiguous sentences like "You can tickle the frog with the feather/mirror" by changing the affordances of the object in the prepositional phrase (e.g. feathers are better than mirrors for tickling).

Experiment 1 confirmed that our manipulation of object affordances was valid. Adults and 5-year-old children shared intuitions about which objects make better instruments for particular actions (by-item correlation, $r_s(22) = .59$, $p = .002$; see Fig. 2), and these intuitions largely patterned with the adult ratings task that we used to generate our stimuli. Experiment 2 directly tested children's ability to use plausibility and verb bias when inferring phrase structure and found that children relied almost exclusively on the bottom-up verb bias information. Children looked more to the Target Instrument in conditions with instrument-biased verbs relative to modifier and equi-biased verbs (see Fig. 9), and they showed no preference to look at the Target Instrument when it was more plausible (see Fig. 8). Thus, these results support our first hypothesis that children have a broad deficit in using top-down cues regardless of cue validity. Children, however, were not completely insensitive to our plausibility manipulation; it did affect their ultimate interpretation of the sentences. Children were more likely to perform instrument actions when the object was a highly plausible instrument for that verb (see Fig. 7), though this effect was much smaller than the effect of verb bias.

Taken together, these data support the hypothesis that children's failure to use referential context cues when inferring phrase structure is not due *solely* to the cue's poor validity. Rather, children's failures are consistent with a broad deficit in using top-down cues to guide language comprehension. This deficit is not absolute; children's actions are influenced to a small degree by the plausibility of the event, showing that they have some ability to use this cue to settle on an interpretation. But it is clear that children make less use of this information than adults—they are slower to use it, and they give it far less weight in their final interpretation.

We recognize that ruling out one hypothesis can never provide definitive support for an alternative hypothesis. We have demonstrated that children's failures with top-down cues are not due solely to cue validity, but our findings are compatible with a range of alternatives. First and most obviously, school-aged children could have a global deficit in rapidly using top-down information during language comprehension (or perceptual processes, more generally) but may otherwise be able to use low validity cues to the same degree as adults. Next, perhaps children are both less adept at using top-down information, and less adept at using low validity cues. To investigate this hypothesis, we would need to identify a bottom-up cue with low validity that adults are skilled at using, and then test it in children. Finally, there could be other factors that we have not yet considered, which may better explain why children heavily rely on some cues (e.g. verb bias, prosody) but not others (e.g. referential context, plausibility). Exploring this final possibility would require identification of a third theoretically relevant dimension (beyond reliability and top-downness) that correctly predicts the observed data pattern. We are not aware of any plausible candidates. For this reason, the remainder of our discussion will focus on the hypothesis that children are less adept at top-down processing, without making a commitment about their ability to otherwise use low-validity cues.

Thus, in this General Discussion, we first integrate our adult findings with the prior literature on how adults use lexical biases and contextual cues when inferring phrase structure. We then integrate our child findings with the prior developmental literature. Finally, we return to our alternative hypotheses about the origins of children's top-down deficit (as outlined in the Introduction) and end by discussing future research that might distinguish between them.

4.1. Adults' use of bottom-up lexical biases and top-down contextual cues during comprehension

The present study is broadly compatible with a rich body of work demonstrating that adults integrate bottom-up lexical cues and top-down contextual cues to phrase structure during comprehension (e.g. Altmann & Steedman, 1988; Just & Carpenter, 1992; MacDonald, 1994; Trueswell et al., 1994; Spivey-Knowlton & Sedivy, 1995; Kidd & Bavin, 2007). In this section, we discuss two studies that specifically explore the contributions of verb biases and plausibility during adult ambiguity resolution (Garnsey et al., 1997; Kidd et al., 2011). We conclude that, while these studies emphasize the role of verb bias in adult's early language comprehension, we can reconcile their findings with ours by considering the differences in the tasks, the syntactic constructions, and the specific plausibility cues used in each study.

The first study is very similar to ours. Kidd et al. (2011) used the visual world paradigm to see how adults (and children) resolve ambiguous sentences with conflicting verb bias and plausibility cues. This study had one condition: all sentences had instrument-biased verbs and implausible instruments (e.g. "Cut the cake with the candle"). Their findings are consistent with ours in many ways: First, both studies find that when verb bias and plausibility are in conflict, adults split the difference in their final interpretation. Kidd and colleagues found that adults often used implausible instruments to perform nonsensical actions like using a large candle to cut a cake (40% of adults' actions were implausible). In the present study, adults performed actions with the implausible instruments 75% of the time in the parallel Low Plausibility Instrument-biased condition (see Fig. 7). Second, Kidd and colleagues find evidence for adults' early sensitivity to plausibility during online comprehension. In their study, after hearing the verb, adults showed increased looks to a Distractor Instrument with high affordances for the mentioned action (e.g. a knife for cutting a cake). We also find adults using object affordances to guide interpretation. However, because Kidd and colleagues did not manipulate plausibility nor verb bias, we cannot assess the relative magnitude of verb bias and plausibility cues or compare their findings directly with our own.

In contrast, Garnsey et al. (1997) used a design that is more similar to ours, as they systematically manipulated both plausibility and verb bias in written sentences. They found that adults make rapid use of both cues when reading temporarily ambiguous sentences like those in (2). Verb biases were manipulated by selecting verbs that varied in how often they appeared with different kinds of post-verbal arguments: DO-biased verbs typically appeared with Direct Objects (see 2a), SC-biased verbs typically appeared with Sentential

Complements (see 2b), and Equi-biased verbs appeared equally often with complements of both kinds (see 2c). Plausibility was manipulated by changing the thematic fit of the NP after the verb such that some NPs (e.g. *the problems, the issue, the accusation*) were plausible direct objects for the specific verb while others (e.g. *the schools, the witness, the election*) were implausible. They also had unambiguous control sentences with overt complementizers (e.g. The angry father emphasized *that* the problems/the schools were continuing to get worse) that disambiguated the sentence before the target NP.

(2) a. The angry father *emphasized* **the problems/the schools** were continuing to get worse.
b. The divorce lawyer *argued* **the issue/the witness** was irrelevant to the case.
c. The crooked politician *denied* **the accusation/the election** would change things at all.

Garnsey et al. (1997) found effects of *both* verb bias and plausibility on fixation times at the critical NP region (in bold) and the point of disambiguation (underlined), i.e., where it became clear that the noun was the subject of a sentential complement.

The pattern of effects, however, suggests that verb bias plays the dominant role in the adults' initial interpretation of these ambiguities, while plausibility plays a more modest role and is used later in comprehension. Specifically, in the SC-biased conditions, the ambiguous sentences were read as quickly as the unambiguous controls regardless of plausibility, suggesting that readers quickly arrived at the intended interpretation due to verb bias. In DO-biased conditions, however, readers were slower to read the disambiguation region in ambiguous conditions relative to unambiguous controls regardless of the plausibility of the NP, suggesting that the NPs immediately following the DO-biased verbs are initially interpreted as direct objects. Nevertheless, there were signs in the DO-biased conditions that plausibility information was available early and played a role in revising this initial commitment. First, adults were slower to read implausible NPs than plausible NPs in ambiguous sentences, suggesting they had recognized that the implausible noun was unexpected on their preferred analysis (DO). Second, there were more regressions at the point of disambiguation when the NP was a plausible direct object in ambiguous sentences, suggesting that the preferred argument structure for the verb was more difficult to rule out when it was supported by plausibility as well. In sum, their findings suggest that adults are sensitive to plausibility cues but make use of them only after verb biases guide their initial syntactic commitments.

Thus, while both our study and the Garnsey study find effects of these two variables, the pattern of effects is quite different. In our study, plausibility is the early and dominant cue, and, in their study, it is verb bias. This discrepancy between studies could reflect a difference in how top-down cues are used in different discourse contexts. For example, Spivey, Tanenhaus, Eberhard, and Sedivy (2002) have noted that comprehenders' use of lexical constraints and referential cues differs across paradigms. In most reading studies, lexical constraints are dominant while referential cues are weak, whereas in visual world studies, referential cues seem to dominate. They attribute this difference to the way in which context is built in the two paradigms: in reading studies, the discourse referents are established in the text itself and must be held in memory; in visual world studies, the referents are continuously visible in the scene, making the referential information more salient (also see Tanenhaus et al., 1995; Tanenhaus & Trueswell, 2005 for discussion). Plausibility information may also be more salient and easier to construct when the context is physically co-present.

The particular cue that we manipulated in our study, object affordances, was largely motivated by J.J. Gibson and E.J. Gibson's ecological theory of perception (Gibson, 1977, 1979, 1982, 2014; see also Chemero, 2003; Dotov, Nie, & De Wit, 2012; Adolph & Kretch, 2015; Wagman & Blau, 2019 for discussion). The Gibsons argued that perception exists to allow for action and that, consequently, the retrieval of affordances is a fundamental part of seeing an object. On this hypothesis, we should expect that other visual world studies will find early and robust effects of plausibility in adults. Kidd et al. (2011) confirm this expectation. First, they find early looks to plausible instruments, which suggests that object affordances are being rapidly calculated and considered. Second, on 60% of the trials, adults did not use the implausible instrument for the action, suggesting that object affordances may be used to overcome the strong instrument bias from the verb. Finally, the spontaneous use of the unmentioned instrument in their study suggests that, in planning their actions, adults considered object affordances even when the object in question had not been mentioned.

In sum, our findings on adult's use of bottom-up and top-down cues during syntactic ambiguity resolution are consistent with the larger literature at the broadest level: adults use both types of information when inferring phrase structure online. Unlike the previous literature, we find that adults will rely almost exclusively on top-down plausibility cues when the cue is robust and reliable in the wider language and is not systematically invalid in the experimental context. This conclusion is consistent with most theories of cue validity, which argue that adults will prioritize a single cue with high reliability (if available) rather than rapidly integrating it with multiple weaker cues that might collectively improve reliability (e.g. MacWhinney et al., 1984; Kempe & MacWhinney, 1999; MacWhinney, 2005; 2008).

4.2. Children's use of bottom-up and top-down cues during comprehension

The present study produced three main findings about children's use of bottom-up and top-down cues during comprehension—all of which are broadly consistent with the prior developmental literature but go beyond it in critical ways.

First, we find that 5-year-old children are highly sensitive to bottom-up lexical information when inferring phrase structure: verb biases were the dominant factor in shaping children's moment-to-moment understanding and their ultimate interpretations of syntactic ambiguities. This is consistent with previous studies that have found a similar reliance on verb biases in 5-year-olds' online and offline comprehension measures (e.g. Snedeker & Trueswell, 2004; Kidd & Bavin, 2005; Kidd et al., 2011). These studies, however, did not rate or control for the global plausibility of their target sentences, and thus did not cleanly disentangle bottom-up and top-down cues. For example, Snedeker and Trueswell (2004) collected ratings for how good the Target Instrument was for the given action, considering that the instrument was being used to complete the task. They explicitly instructed raters to factor out plausibility differences in instrument use across verbs. This almost certainly resulted in a confound—instrument interpretations were presumably more plausible for sentences with instrument-biased verbs relative to sentences with modifier-biased verbs simply because instrument-biased verbs describe events in which instrument use is expected and/or necessary (e.g. "Cut the cake with the knife"). This is critical to understanding adults' performance in their study and in our own: Snedeker and Trueswell (2004) found strong effects of verb bias in adults while the present study did not. However, the apparent difference in these studies may merely reflect a consistent reliance on plausibility cues in adults, which were confounded with verb bias in one study and orthogonal to verb bias in the other. The present study is, to the best of our knowledge, the only study in the developmental literature that carefully disentangles verb bias from global plausibility cues, allowing us to observe how (and when) these two factors interact during 5-year-old children's syntactic ambiguity resolution.

Second, in Experiment 1, we demonstrated that 5-year-olds have knowledge about object affordances, specifically which instruments are better suited for particular actions. This finding is consistent with a broader developmental literature that shows children are able to select the best tool for a given task by judging the affordances of the objects in front of them (e.g. Brown, 1990; Klatzky, Lederman, & Mankinen, 2005; Beck, Apperly, Chappell, Guthrie, & Cutting, 2011). For example, Beck et al. (2011) showed 4- to 7-year-old children two pipe cleaners—one that was bent into a hook-shape and one that was straight—and then asked them to retrieve a bucket that was stuck inside of a narrow tube. They found that children reliably chose the hooked pipe cleaner over the straight one to complete the task.

Third, we find that 5-year-olds, despite knowing about object affordances, make minimal use of this information when interpreting structurally ambiguous sentences. We found no overall effects of plausibility on children's eye-movements and a small effect (relative to the effect of verb bias) on children's actions (see Figs. 7–8). This is consistent with the two prior studies looking at plausibility and structural ambiguity resolution (Kidd & Bavin, 2005; Kidd et al., 2011), as well as with a larger literature on children's limited use of top-down information in language processing. Kidd and Bavin (2005) presented children with sentences that varied in the affordances of the object in the prepositional phrase and the bias of the verb (3).

(3)	a. **Tickle** the hippo with the *feather*.	(**Instrument-biased verb**, *plausible instrument*)
	b. **Chop** the tree with the *leaves*.	(**Instrument-biased verb**, *implausible instrument*)
	c. **Look** at the duck with the *glasses*.	(**Modifier-biased verb**, *plausible instrument*)

Their visual displays had two possible referents for the NP (e.g. a tree with leaves, and one without leaves), the prepositional phrase object (e.g. a pile of leaves), and a distractor object (e.g. a knife). For instrument-biased verbs, children made fewer Target Instrument actions when the mentioned object was an implausible instrument (3b), suggesting that they can use plausibility cues to overcome the strong lexical bias. Children also made fewer Target Instrument actions in the condition with modifier-biased verbs (3c) relative to the conditions with instrument-biased verbs (3a and 3b), again showing children's sensitivity to verb bias manipulations. While this study is consistent with our findings, it has two limitations relative to the present study: First, the plausible and implausible instrument sentences are separate items that are not matched for their critical properties. The items use different verbs and distractor instruments, making it impossible to isolate the effect of plausibility per se. Second, Kidd and Bavin (2005) did not collect eye gaze data, and thus their findings cannot tell us how plausibility is used by 5-year-olds in real time. As we noted previously, in a later study, Kidd et al. (2011) did look at eye-movements for implausible sentences with instrument-biased verbs; however, this study was limited to a single condition, which means that the data cannot show us how effects of verb bias or plausibility emerge in real time. Their findings, however, are again consistent with ours: 5-year-old children used the implausible instrument most of the time, and more often than adults.[9]

At first glance, the broader literature on children's use of top-down cues during ambiguity resolution seems messier and less consistent. Most studies have found that children under the age of eight fail to use the referential context and related cues when inferring phrase structure (e.g. Trueswell et al., 1999; Snedeker & Trueswell, 2004; Hurewitz et al., 2000; Kidd & Bavin, 2005; 2007; Anderson et al., 2011; Weighall, 2008). However, there are a handful of studies that find effects of top-down cues on children's eye movements (Snedeker & Trueswell, 2004), actions (Rabagliati, Pylkkänen, & Marcus, 2013), or both (Qi, Love, Fisher, & Brown-Schmidt, 2020). For example, Qi et al. (2020) found effects of referential context on 5- to 6-year-olds' interpretation of temporary ambiguities like "Put the frog on the pond into the tent." When the visual scene included two frogs (one on a pond, another on a leaf), children made fewer looks to the incorrect destination (the pond) and fewer incorrect actions relative to conditions with only one frog. Both effects, however, were fragile and small, each appearing in one of their experiments but not the other. Qi and colleagues conclude that, despite their limited success in this task, 5-year-olds are poor at using referential cues, perhaps because they are slow in calculating the relevance of top-down constraints. In sum, every study we are aware of finds that children under age 8 are far more likely than adults to rely more on bottom-up information to resolve ambiguities.

[9] As we noted in the Introduction, a third study by Bavin et al. (2016) found effects of instrument plausibility in an eye-tracking paradigm with slightly older children. The effects in this study, however, could reflect bottom-up word recognition (i.e. looking more at candles when hearing "candle" and at knives when hearing "knife") instead of effects of plausibility on syntactic ambiguity resolution (i.e. looking to a potential instrument when it is mentioned and plausible for the action, but not when it is mentioned but implausible).

4.3. Why might children have a broad deficit in top-down processing?

In the Introduction, we introduced three alternative hypotheses about why children may have a broad deficit in using top-down cues to inform lower-level processes. The first hypothesis is that children simply lack top-down connections between the various representations in question. We argued that this hypothesis was a non-starter in the domain of language because these top-down connections are necessary for language production. Because children are clearly capable of expressing their thoughts and needs, they must be able to select lower-level linguistic representations on the basis of higher-level plans. Of course, one could envision a computational system in which production and comprehension were carried out by completely separate networks, but human beings seem to use the same networks and representations for these tasks (see Meyer et al., 2016 for an overview).

Our second hypothesis, the *limited cue integration* account, is a wrinkle on classic modularity (Fodor, 1983). This account argues that children are limited in the degree to which they can integrate information from multiple information sources, and instead they rely heavily on cues that follow the dominant flow of information for a given process—top-down processing from ideas to words in production, and bottom-up processing from words to ideas in comprehension. This bottleneck could result from architectural constraints and/or limitations in working memory (see Snedeker & Huang, 2015; Omaki & Lidz, 2015; Allen & Behrens, 2019 for reviews). As children grow older, they overcome this bottleneck, developing the capacity to hold onto multiple representations and evaluate them with respect to both bottom-up and top-down information.

Our third and final hypothesis is that children's limitations in top-down processing are a side effect of their slower processing speed. This *processing speed* hypothesis draws on two lines of research. The first is work by Kail and colleagues, which demonstrates that, across a wide range of tasks, young children process information more slowly than older children and adults do (Kail & Ferrer, 2007; Kail, 1991; Kail, Hall, & Caskey, 1999). For example, 5-year-olds take about three times as long as adults to classify objects, match pictures to sentences, tap their fingers, or determine the number of objects in an array, while 8-year-olds take roughly twice as long (Kail, 1991). The improvement in processing speed is initially steep but decelerates across development with performance peaking somewhere around 18 years of age (Kail & Ferrer, 2007; Kail & Salthouse, 1994). For example, if a task takes adults 1000 ms to complete, 5- to 8-year-olds will take 2000 ms, 8- to 11-year-olds will take 1500 ms, 11- to 14-year-olds will take 1300 ms, and 14- to 18-year-olds will take 1200 ms (Kail, 2000, 1991). Thus, the period of steepest growth roughly coincides with the period in which children come to make greater use of top-down cues in processing. The gradual improvement that we see throughout this period, with considerable individual differences, is compatible with the small and variable effects of context that we see in children around the age of five (Qi et al., 2020).

The second line of work that inspires this hypothesis is research by Gary Dell on the role of time in establishing feedback loops in multilayered dynamic systems like the language processing system. His work focuses largely on language production in which the dominant flow of information is top-down from ideas to words. Dell (1986) investigated if (and when) bottom-up information from phonological selection processes influenced higher-level processes related to lexical selection. He points out that, in models of production, feedback effects emerge slowly over time because they require information from higher levels (e.g. lexical selection) to first filter down to lower levels (e.g. phonological selection) and then percolate back up. He tested the predictions of these models by using a speech error induction paradigm and by varying the time that participants had to respond. Participants with longer response intervals produced a greater proportion of errors that showed the influence of a feedback loop.

The same logic can be applied to models of comprehension but in the opposite direction: feedback effects can only emerge after information from lower levels spreads to higher ones and then filters back down to influence ongoing processes at levels below. To translate this logic to the present study, plausibility information can only affect a listener's syntactic processes after perceptual representations of the incoming sounds allow the listener to activate words, which act as inputs for syntactic processes like structure building. In contemporary models of comprehension, bottom-up representations include information that is stored with a word (i.e. lexical biases, verb semantics). These representations can inform which candidate syntactic structures are built and then considered as potential interpretations of the utterance. As soon as these candidate structures are constructed, they get sent up to higher levels for further evaluation about which are likely and unlikely given a listener's knowledge of the world. For example, "Tim poked the cat with the black and white fur" does not readily translate into a plausible VP-attached structure (i.e. Tim used fur to poke the cat). This sentence does, however, translate into a plausible NP-attached structure (i.e. there is a black and white cat that got poked by Tim). Once the plausibility of these structures is evaluated, this information is sent back down to the syntactic level as signals to downweight certain analyses (e.g. the implausible ones) and upweight others (e.g. the ones that are plausible and/or relevant). Thus, on this account, listeners' access to top-down cues like plausibility will often be delayed and dependent on some degree of bottom-up processing during comprehension. Whether these top-down cues get used will depend on whether the processes at the lower level(s) of interest (e.g. syntactic analyses) terminate before the top-down information filters back down.

The present study was not designed to distinguish between the *processing speed* and the *limited cue integration* hypotheses outlined above—however, these two hypotheses make distinct and testable predictions about when children will be most successful in using top-down information during comprehension.

The processing speed hypothesis is grounded in work demonstrating that, across a wide range of tasks, children show consistently slower processing than adults (e.g. Kail, 1991; Kail & Salthouse, 1994; Kail, 2000; Kail & Ferrer, 2007). In other words, for each unit of time that passes, a child completes fewer mental operations than an adult does. This difference has consequences for top-down processing, as comprehenders must complete many mental operations to rapidly construct higher-level representations that feed back down to lower levels and constrain ongoing interpretation. For example, prior work has demonstrated that adults are less apt to construct feedback loops of this kind if they are under time pressure (see Dell, 1986). Under the processing speed hypothesis, a child hearing normally paced language is much like an adult operating under time pressure. Thus, all other things being equal, the processing speed account predicts that if we can slow down the speech or give children more time to process what they are hearing, then they should be able to make greater use of top-down information.

In contrast, the limited cue integration hypothesis does not predict that the use of top-down information will vary with the time available to comprehenders. On this hypothesis, children's limitations are linked to their inability to integrate cues, which forces them to rely on the dominant stream of information for any given process (e.g. bottom-up information for comprehension, top-down for production). As they grow, children gradually overcome this bottleneck, perhaps due to improvements in working memory or executive function. But critically, their success is not related to the amount of time that they have to process the input. Thus, under this account, we should expect children's processing difficulties to persist, even at slower speech rates.

To date, there is no strong evidence bearing on these predictions. One study by Qi et al. (2020), however, provides modest support for each of these hypotheses. Their primary task and their stimuli were similar to those of Trueswell et al. (1999), e.g., "Put the frog on the pond into the tent" with either one or two frogs present in the scene. In contrast with earlier studies, they found effects of referential context on children's interpretation of the sentences, though these effects were small and variable across their experiments. The difference between this study and the prior studies could be due to the presentation speed: the researchers used a slower speech rate in hopes that it would improve performance. However, because many other factors varied as well (e.g. a between-participants design, a different participant pool, and different sets of objects), no strong inferences can be drawn. The researchers also conducted correlational analyses of individual performance and found that the use of referential context was greatest in children with higher executive function. This finding would be consistent with the limited cue integration hypothesis. But given the complex web of things that are correlated with childhood executive function (e.g. SES, performance IQ, socio-emotional development) and the myriad of ways in which executive function could influence task performance (e.g. better attention, better learning across trials, better revision), this particular finding is consistent with many alternative hypotheses as well.

In our post-hoc analyses, we found further evidence in support of the processing speed hypothesis. Specifically, we found that when children have longer to comprehend the sentence, they are more likely to use plausibility information to guide their interpretation (see Fig. 12 above). The processing speed hypothesis also makes predictions about adult comprehension. On this hypothesis, effective use of top-down cues often takes more time than the use of bottom-up cues because it typically requires the construction of a feedback loop: the syntactic analyses under consideration must be constructed, then evaluated at a higher level, and then updated accordingly. If adults are put under time pressure, we expect that they will be less able to construct these feedback loops before responding (Dell, 1986), and thus, they will be less likely to make use of top-down cues. This prediction could be tested by increasing the speed of presentation or by using extremely short ambiguities and creating pressure to respond quickly.

5. Conclusion

The present study investigated why children struggle to use the same information as adults do when inferring the structure of the sentence they are hearing. Prior work has demonstrated that children do not rely on high-level, top-down information about their immediate context when interpreting PP-attachment ambiguities like "You can tickle the frog with the feather." We investigated two hypotheses about why children may ignore top-down cues like the referential context: First, children may be less likely to use *any* top-down cue during syntactic processing. Alternatively, children may only struggle with top-down cues that are not strongly informative with respect to sentence structure. To address this question, we manipulated a more reliable top-down cue (plausibility) and pitted it against a reliable bottom-up cue (verb bias). We find that adults' and children's interpretations are shaped by both sources of information. Adults, however, rely primarily on plausibility during comprehension (although, they shift to using this top-down cue only after learning more about the trial structure). Children, on the other hand, rely primarily on bottom-up verb bias information during moment-to-moment comprehension. In a post-hoc analysis, we find that sensitivity to top-down information emerges in children's comprehension when they have more time to compute the possible interpretations of the utterance. This finding suggests that children have a bias to follow easily accessible, bottom-up cues; however, if given enough time, they will begin to integrate cues that rely on more time intensive inferences. Thus, we propose that children's bias to use bottom-up information is most likely a reflection of the architecture of the comprehension system and differences in processing speed.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Maribelle Dickins, Karen Andres, Paulina Piwowarczyk, and Moshe Poliak for their assistance; Susan Carey and members of the Harvard community for their recommendations and feedback on early drafts; Arunima Sarin for providing much needed support, perspective, and clarity to the piece; Renée Baillargeon and two anonymous reviewers for their attention to detail and helpful critiques; and finally, all of the families and individuals who participated in this work. This research was presented at the 34th Annual Boston University Conference on Language Development (November 2009; Boston, MA) and the 31st Annual CUNY Sentence Processing Conference (March 2018; Davis, CA).

Funding

This work was supported by a grant from the National Science Foundation (NSF BCS-0623845 awarded to Jesse Snedeker) and by an award from the Harvard University's Barbara Ditmars Fund (awarded to Anthony Yacovone).

References

Adolph, K. E., & Kretch, K. S. (2015). Gibson's theory of perceptual learning. *International Encyclopedia of the Social and Behavioral Sciences, 10*, 127–134.
Akhtar, N. (2002). Relevance and early word learning. *Journal of Child Language, 29*, 677–686. https://doi.org/10.1017/S0305000902005214
Allen, S. E. M., & Behrens, H. (2019). In *Insights into Understanding Human Language from Children's Acquisition of Morphology and Syntax: A Historical and Current Perspective on Central Questions in the Field* (pp. 127–145). MIT Press.
Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439.
Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*(3), 191–238.
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. *Experience, Variation, and Generalization: Learning a First Language, 7*, 203–221.
Atkinson, R. C., & Shiffrin, R. M. (1968). *Human memory: A proposed system and its control processes* (Vol. 2,, 89–195.
Baayen, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. Cambridge University Press.
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* (4), 390–412.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.
Bavin, E. L., Kidd, E., Prendergast, L. A., & Baker, E. K. (2016). Young children with ASD use lexical and referential information during on-line sentence processing. *Frontiers in Psychology, 7*, 171.
Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., & Cutting, N. (2011). Making tools isn't child's play. *Cognition, 119*(2), 301–306.
Berry, M. J., & Schwartz, G. (2011). The retina as embodying predictions about the visual world. Predictions in the brain: Using our past to generate a future, 295.
Bever, T. G. (1970). The cognitive basis for linguistic structures. Cognition and the development of language.
Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*(6), 774–806.
Borovsky, A. (2017). The amount and structure of prior event experience affects anticipatory sentence interpretation. *Language, Cognition and Neuroscience, 32*(2), 190–204.
Braginsky, M., Sanchez, A. & Yurovsky,D. (2020). childesr: Accessing the 'CHILDES' Database. R package version 0.2.1. https://CRAN.R-project.org/package=childesr.
Brown, A. L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive science, 14*(1), 107–133.
Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2002). Reference resolution in the wild: On-line circumscription of referential domains in a natural, interactive problem-solving task. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 24, No. 24).
Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Univ of California Press.
Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47*(1), 30–49.
Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15*(2), 181–195.
Choi, Y., & Mazuka, R. (2003). Young children's use of prosody in sentence parsing. *Journal of Psycholinguistic Research, 32*(2), 197–217.
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*(1), 41–61.
Clifton, C., Jr, Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*(3), 317–334.
Connine, C., Ferreira, F., Jones, C., Clifton, C., & Frazier, L. (1984). Verb frame preferences: Descriptive norms. *Journal of Psycholinguistic Research, 13*(4), 307–319.
Dell, G. S., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321.
Dotov, D. G., Nie, L., & De Wit, M. M. (2012). Understanding affordances: History and contemporary development of Gibson's central concept. Avant: The Journal of the Philosophical-Interdisciplinary Vanguard.
Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*(6), 409–436.
Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11*(3), 127–163.
Fennell, C., & Waxman, S. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development, 81*(5), 1376–1383. https://doi.org/10.1111/j.1467-8624.2010.01479.x
Fodor, J. A. (1983). *The modularity of mind*. MIT press.
Forster, K. I. (1979). Levels of processing and the structure of the language processes. *Sentence processing: Psycholinguistic studies presented to Merril Garret*, 27–84.
Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*(4), 291–325.
Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Doctoral dissertation. University of Connecticut.
Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22*(5), 1219.
Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers, 36*(3), 432–443.
Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*(1), 58–93.
Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel hierarchical models* (Vol. 1). New York, NY, USA: Cambridge University Press.
Gibson, J. J. (1966). The senses considered as perceptual systems.
Gibson, J. J. (1977). The theory of affordances. *Hilldale, USA, 1*(2).
Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
Gibson, J. J. (1982). *Notes on affordances*. Hillsdale, USA: Reasons for realism.
Gibson, J. J. (2014). *The ecological approach to visual perception* (classic edition). Psychology Press.
Hahn, N., Snedeker, J., & Rabagliati, H. (2015). Rapid linguistic ambiguity resolution in young children with autism spectrum disorder: Eye tracking evidence for the limits of weak central coherence. *Autism Research, 8*(6), 717–726.
Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Development, 61*(3), 653–663.

Hallett, P. (1986). Eye movements (and human visual perception). *Handbook of Perception and Human Performance, 1*, 10–11.
Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*(2), 151–171.
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*(6), 597–626.
Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language, 50*(2), 182–195.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.
Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 61*(5), 708–723.
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.
Kail, R. V., & Ferrer, E. (2007). Processing speed in childhood and adolescence: Longitudinal models for examining developmental change. *Child Development, 78*(6), 1760–1770.
Kail, R. (1991). Processing time declines exponentially during childhood and adolescence. *Developmental Psychology, 27*(2), 259–266.
Kail, R. (2000). Speed of information processing: Developmental change and links to intelligence. *Journal of School Psychology, 38*(1), 51–61.
Kail, R., Hall, L. K., & Caskey, B. J. (1999). Processing speed, exposure to print, and naming speed. *Applied Psycholinguistics, 20*(2), 303–314.
Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86*(2-3), 199–225.
Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133–156.
Kempe, V., & MacWhinney, B. (1999). Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes, 14*(2), 129–171.
Kidd, E., & Bavin, E. L. (2005). Lexical and referential cues to sentence interpretation: An investigation of children's interpretations of ambiguous sentences. *Journal of Child Language, 32*(4), 855–876.
Kidd, E., & Bavin, E. L. (2007). Lexical and referential influences on on-line spoken language comprehension: A comparison of adults and primary-school-age children. *First Language, 27*(1), 29–52.
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*(1), 222–234.
Klatzky, R. L., Lederman, S. J., & Mankinen, J. M. (2005). Visual and haptic exploratory procedures in children's judgments about tool function. *Infant Behavior and Development, 28*(3), 240–249.
Kloke, J. D., & McKean, J. W. (2012). Rfit: Rank-based estimation for linear models. *The R Journal, 4*(2), 57–64.
Koenig, J.-P., Mauner, G., & Bienvenue, B. (2003). Arguments for adjuncts. *Cognition, 89*(2), 67–103.
Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571–579.
MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9*(2), 157–201.
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703.
MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development, 43*(1/2), 1. https://doi.org/10.2307/1166047
MacWhinney, B. (1987). The competition model. *Mechanisms of Language Acquisition*, 249–308.
MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (Third Edition). Mahwah, NJ: Lawrence Erlbaum Associates.
MacWhinney, B. (2001). Psycholinguistics: Overview. *International Encyclopedia of the Social & Behavioural Sciences, 18*, 12343–12349.
MacWhinney, B. (2005). A unified model of language acquisition. Handbook of bilingualism: Psycholinguistic approaches, 4967.
MacWhinney, B. (2008). *A unified model*. Routledge/Taylor & Francis Group.
MacWhinney, B. (2012). The logic of the unified model. *The Routledge Handbook of Second Language Acquisition*, 211–227.
MacWhinney, B., & Bates, E. (1989). The crosslinguistic study of sentence processing.
MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 127–150.
MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28*(3), 255–277.
Malmberg, K. J., Raaijmakers, J. G. W., & Shiffrin, R. M. (2019). 50 years of research sparked by Atkinson and Shiffrin (1968). *Memory & Cognition, 47*(4), 561–574.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG- data. *Journal of Neuroscience Methods, 164*(1), 177–190.
Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two- and four-year-olds learn to adapt referring expressions to context: Effects of distracters and feedback on referential communication. *Topics in Cognitive Science, 4*(2), 184–210. https://doi.org/10.1111/j.1756-8765.2012.01181.x
Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience, 10*(1), 363–401.
McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 33, No. 33).
McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 21*(1), 134.
Meyer, A. S., Huettig, F., & Levelt, W. J. M. (2016). Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language, 89*, 1–7.
Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86*(4), 314–329.
Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition, 22*(2), 158–192.
Pickering, M. J., & Garrod, S. (2013). How tightly are production and comprehension interwoven? *Frontiers in Psychology, 4*, 238.
Qi, Z., Love, J., Fisher, C., & Brown-Schmidt, S. (2020). Referential context and executive functioning influence children's resolution of syntactic ambiguity. *Journal of Experimental Psych.: Learning, Memory, and Cognition., 46*(10), 1922–1947.
R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rabagliati, H., Pylkkänen, L., & Marcus, G. F. (2013). Top-down influence in young children's linguistic ambiguity resolution. *Developmental Psychology, 49*(6), 1076–1089.
Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 358–374.
Roland, D., & Jurafsky, D. (2002). Verb sense and verb subcategorization probabilities. *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues, 4*, 325–345.
Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71*(1), 145–163.
Sedivy, J. C., K. Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*(2), 109–147.
Snedeker, J. (2013). Children's Sentence Processing. In S. Processing (Ed.), *van Gompel* (pp. 189–220). New York, New York: Psychology Press.
Snedeker, J., & Huang, Y. (2015). Sentence Processing. In E. Bavin, & L. Naigles (Eds.), *The Handbook of Child Language* (2nd Edition, pp. 409–437). Cambridge University 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*(3), 238–299.
Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58*(2), 574–608.

Sokolov, J. L. (1988). Cue validity in Hebrew sentence comprehension. *Journal of Child Language, 15*(1), 129–155.
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*(4), 447–481.
Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*(3), 227–267.
Stackhouse, J., & Wells, B. (Eds.). (1997). *Children's speech and literacy difficulties: A psycholinguistic framework* (Vol. 1). John Wiley & Sons Incorporated.
Swinney, D. A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 645–659.
Tanenhaus, M. K., & Trueswell, J. C. (2005). Eye movements as a tool for bridging the language-as-product and language-as-action traditions. Approaches to studying world- situated language use: Bridging the language-as-product and language-action traditions, 3-37.
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*(5217), 1632–1634.
Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language, 27*(6), 597–632.
Teufel, C., & Nanay, B. (2017). How to (and how not to) think about top-down influences on visual perception. *Consciousness and Cognition, 47*, 17–25.
Traxler, M. J. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology Section A, A, 55*(1), 75–96.
Trueswell, J., & Gleitman, L. (2004). Children's eye movements during listening: Developmental evidence for a constraint-based theory of sentence processing. The interface of language, vision, and action: Eye movements and the visual world, 319-346.
Trueswell, J. C., & Gleitman, L. R. (2007). Learning to parse and its implications for language acquisition. In The Oxford handbook of psycholinguistics.
Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution.
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* (2), 89–134.
Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*(3), 285–318.
Wagman, J. B., & Blau, J. J. C. (2019). A guided tour of Gibson's theory of affordances. Perception as information detection: Reflections on Gibson's Ecological approach to visual perception (ed. Wagman, J. B., and Blau, J. J. C.) 130–148.
Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology, 99*(2), 75–95.
Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2017). Eye- tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language, 167*, 13–27.
Wendt, D., Brand, T., Kollmeier, B., & Najbauer, J. (2014). An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities. *PLoS One, 9*(6), e100186.
Yacovone, A., Rigby, I., & Omaki, A. (2020). Children's comprehension and repair of garden- path wh-questions. *Language Acquisition*, 1–34.
Yang, W., Chan, A., Chang, F., & Kidd, E. (2020). Four-year-old Mandarin-speaking children's online comprehension of relative clauses. *Cognition, 196*, 104103. https://doi.org/10.1016/j.cognition.2019.104103
Yuan, S., Fisher, C., Kandhadai, P., & Fernald, A. (2011). You can stipe the pig and nerk the fork: Learning to use verbs to predict nouns. In *Proceedings of the 35th annual Boston University conference on language development* (pp. 665–677). Boston, MA: Cascadilla Press.