Available online at www.sciencedirect.com
ScienceDirect
Cognition 107 (2008) 850–903
COGNITION
www.elsevier.com/locate/COGNIT

# Putting lexical constraints in context into the visual-world paradigm

Jared M. Novick *, Sharon L. Thompson-Schill, John C. Trueswell

*Department of Psychology and Institute for Research in Cognitive Science, University of Pennsylvania, Suite 302C, 3401 Walnut Street, Philadelphia, PA 19104, USA*

Received 9 January 2007; revised 26 November 2007; accepted 21 December 2007

**Abstract**

Prior eye-tracking studies of spoken sentence comprehension have found that the presence of two potential referents, e.g., two frogs, can guide listeners toward a Modifier interpretation of *Put the frog on the napkin*... despite strong lexical biases associated with *Put* that support a Goal interpretation of the temporary ambiguity (Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634; Trueswell, J. C., Sekerina, I., Hill, N. M. & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*, 89–134). This pattern is not expected under constraint-based parsing theories: cue conflict between the lexical evidence (which supports the Goal analysis) and the visuo-contextual evidence (which supports the Modifier analysis) should result in uncertainty about the intended analysis and partial consideration of the Goal analysis. We reexamined these *put* studies (Experiment 1) by introducing a response time-constraint and a spatial contrast between competing referents (a frog *on* a napkin vs. a frog *in* a bowl). If listeners immediately interpret *on the*... as the start of a restrictive modifier, then their eye movements should rapidly converge on the intended referent (the frog *on* something). However, listeners showed this pattern only when the phrase was unambiguously a Modifier (*Put the frog that's on the*...). Syntactically ambiguous trials resulted in transient consideration

* Corresponding author. Address: Center for Advanced Study of Language, University of Maryland, Box 25, College Park, MD 20742-0025, USA. Tel.: +1 301 226 8841.
*E-mail address:* jnovick@casl.umd.edu (J.M. Novick).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2007.12.011

of the Competitor animal (the frog *in* something). A reading study was also run on the same individuals (Experiment 2) and performance was compared between the two experiments. Those individuals who relied heavily on lexical biases to resolve a complement ambiguity in reading (*The man heard/realized the story had been...*) showed increased sensitivity to *both* lexical and contextual constraints in the *put*-task; i.e., increased consideration of the Goal analysis in 1-Referent Scenes, but also adeptness at using spatial constraints of prepositions (in vs. on) to restrict referential alternatives in 2-Referent Scenes. These findings cross-validate visual world and reading methods and support multiple-constraint theories of sentence processing in which individuals differ in their sensitivity to lexical contingencies.
© 2008 Elsevier B.V. All rights reserved.

*Keywords:* Parsing; Syntactic ambiguity resolution; Lexical constraints; Visual-world paradigm; Contextual constraints; Reading time measures; Individual differences

## 1. Introduction

Language input enters the human parsing system in an incremental fashion regardless of modality. During reading, words reach our eyes moment by moment as we make successive fixations across text on a page. Similarly, during listening, words make contact with our ears as speakers' utterances unfold word by word. An important consequence of incremental production and comprehension is that readers and listeners are frequently faced with temporary ambiguities about how best to structure the input in real-time. Consider, for example, the following illustration of this:

(1) *Jared put the apple on the towel into the box.*

Here, a temporary syntactic ambiguity arises when encountering *on the towel.* This Prepositional Phrase (PP) could be linked to either the verb *put* as a Goal, indicating where Jared put the apple, or it could be linked to the immediately preceding Noun Phrase (NP) as a Modifier, providing more information about the apple.

In principle, a wide range of evidence could be used by readers and listeners at the point of ambiguity to inform their processing commitments. For instance, lexical evidence in this example sentence strongly supports an initial Goal analysis of *on the towel*. This is because the verb *put* requires a Goal argument and commonly introduces one with a Prepositional Phrase (PP) headed by *on*, *in*, *onto*, or *into*. In addition, readers and listeners could also take into account contextual factors. In this case, no other apples have been mentioned, so there isn't any particular reason to further modify *the apple* with a preposition. Thus, contextual factors also support a Goal analysis. If the context included multiple apples, one of which was on a towel, then a reader or listener who is aware of such information could in theory use it to inform parsing commitments and thereby pursue a Modifier interpretation of this temporarily ambiguous phrase.

An important research agenda within psycholinguistics has been to explore how and when these and other evidential sources are integrated by the reader or listener

to resolve ambiguity. By parametrically manipulating the evidence supporting different alternatives and examining how these manipulations impact processing commitments, one can differentiate various theories of the human sentence parsing process. One broad class of parsing theories, referred to as the interactive constraint-based theories, will be the focus of the current experimental efforts. These theories predict that the effectiveness of various constraints on the parsing process will be a function of their availability at the point of ambiguity and a function of their reliability in predicting a particular parse (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994). If multiple reliable constraints are available, they are weighed simultaneously and integrated into a reader's or a listener's parsing commitment. In example (1) above, a strong commitment to the Goal analysis is expected given the highly reliable lexical predictors for this structure. Contextual factors, such as the presence of multiple apples, are also expected to be available and used, but would have to battle against the lexical support for the Goal analysis. Indeed, one might expect that lexical factors are in general such strong predictors of structure that they would play a particularly important role in most parsing commitments (hence, many have adopted the term Constraint-Based Lexicalist (CBL) theory for this particular perspective).

To date, there is a fair amount of experimental evidence consistent with the CBL theory. In particular, numerous studies, some of which historically precede the CBL theory, have found that contextual and plausibility factors influence syntactic ambiguity resolution, often at the earliest stages of processing (e.g., Altmann & Steedman, 1988; Crain & Steedman, 1982; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Taraban & McClelland, 1988; Trueswell, Tanenhaus, & Garnsey, 1994). Manipulations of the detailed structural preferences of verbs exhibit similar effects (e.g., Holmes, Stowe, & Cupples, 1989; Novick, Kim, & Trueswell, 2003; Trueswell & Kim, 1998; Trueswell, Tanenhaus, & Kello, 1993). And, when lexical and contextual/plausibility evidence have been experimentally pitted against each other, it has been found that both factors contribute simultaneously, never completely eliminating the effect of lexical constraints (e.g., Britt, 1994; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Snedeker & Trueswell, 2004; Sedivy & Spivey-Knowlton, 1994).[1]

There is, however, an especially well-known parsing result that, although generally consistent with interactive parsing theories, is particularly difficult to explain under a CBL account (Tanenhaus et al., 1995; see also Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; and Trueswell, Sekerina, Hill, & Logrip, 1999). These studies have examined the ambiguity illustrated in (1) above, and used a method for studying parsing preferences in the auditory domain. In particular, participants followed spoken instructions to move real objects around a visual workspace while their eye

[1] Although there are some recent studies that appear to provide evidence against the CBL theory (Pickering & Traxler, 1998; Van Gompel, Pickering, Pearson, & Liversedge, 2005), these studies have been criticized for not taking into account relevant constraints (Elman, Hare, & McRae, 2004) or for not accurately deriving the predictions of a constraint-based theory (Green & Mitchell, 2006). These findings and counter-findings will not, however, be the topic of the present paper.

movements were recorded. Target utterances involved temporarily ambiguous and unambiguous imperative sentences like (2a) and (2b).

(2) a. *Put the apple on the towel into the box.*
b. *Put the apple that's on the towel into the box.*

The visual referent world that accompanied each particular target sentence either supported the Goal analysis or supported the Modifier analysis. The Goal analysis was supported by having just one apple present sitting on a towel, and a 'luring' Incorrect Goal (an empty towel). The Modifier analysis was supported by having two apples in the scene, one of which was on a towel, plus the luring Incorrect Goal. Within 1-Apple Scenes, eye fixation patterns showed that listeners rapidly committed to the Goal interpretation of *on the towel* and were 'surprised' by the presence of a second Goal phrase such as *into the box*. This was illustrated by a high proportion of early looks to the Incorrect Goal in the scene, the empty towel, at the onset of hearing *towel*. Upon encountering *into the box*, listeners then redirected their eyes and engaged in a process of finding a new analysis of *on the towel* that permitted *into the box* to be the Goal. Crucially, however, when aspects of the contextual scene supported the Modifier interpretation, Tanenhaus and colleagues found that listeners were unsurprised by *into the box*, and that lexical biases associated with *put* were completely overridden in light of a 2-Apple Scene that supported a Modifier interpretation (Tanenhaus et al., 1995). Virtually no early looks to an Incorrect Goal were observed, and eye movement patterns were essentially identical to those that arose in response to syntactically unambiguous control sentences (for similar studies and replications of these results, see also Spivey et al., 2002 and Trueswell et al., 1999). In addition, Chambers, Tanenhaus, and Magnuson (2004) found that pragmatic factors also influence ambiguity resolution for sentences of this type. When hearing *Pour the egg in the bowl onto the flour,* the affordances of task-relevant objects modulated looks to an Incorrect Goal. In particular, the presence of two liquid eggs generated eye movement patterns similar to the 2-Referent Scenes described above (i.e., no increased looks to the Incorrect Goal relative to unambiguous controls). However, changing one of the liquid eggs to a hard-boiled egg generated eye movements similar to 1-Referent Scenes (i.e., increased Incorrect Goal looks).

These results have been widely regarded as a compelling demonstration of how multiple sources of evidence from both the linguistic input and the non-linguistic visual context can rapidly conspire to guide listeners toward the correct analysis of the sentence, thus supporting constraint-based interactive perspectives (Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). In particular, the prominent visual salience of potential referents in the comprehension environment rapidly influenced the time-course of comprehension to the extent that these contextual sources eliminated any trace of parsing commitments toward an analysis that was consistent with highly constraining verb biases, for instance, an analysis of *on the towel* as the Goal of *put.*

However, as Spivey et al. (2002) point out, constraint-based theories have some trouble accounting for this strong contextual effect. Under such theories, it would

not be expected that highly reliable syntactic constraints, such as those associated with the verb *put,* could be completely overridden by contextual factors. These theories assume that multiple parses (e.g., Goal, Modifier) are temporarily considered in parallel and rapidly resolved on the basis of highly supportive convergent evidence; they do not predict elimination of any alternative before all the constraints have been weighed. Some degree of consideration of the Goal analysis is expected even in the 2-Apple case, which would perhaps be localized to the processing of *on the towel*, with significant effects of context. Indeed, contrary to these visual-world findings, past reading studies that have compared contextual and lexical factors have not observed a complete override of lexical preferences by countervailing contextual constraints (e.g., Britt, 1994; Spivey & Tanenhaus, 1998; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Spivey-Knowlton & Sedivy, 1995).

One explanation that has been given to this incongruous finding has been to appeal to the salient and co-present nature of the referential factors in these visual-world studies (Spivey et al., 2002). In particular, the influence of contextual constraints on parsing during reading could be fundamentally different from its influence on parsing during the comprehension of spoken language. Because contextual factors are introduced linguistically in the reading modality, local lexical constraints such as verb biases may impose a stronger effect on processing. By contrast, when listeners are involved in goal-directed comprehension (i.e., carrying out spoken instructions), the co-present visual context must necessarily be consulted to accurately perform the task; thus, contextual constraints may assert a much greater influence on processing in this domain. Nevertheless, as Spivey et al. (2002) note, it is still quite surprising that highly salient contextual factors could completely override the semantic and syntactic preferences generated by hearing the verb *put*.

Other questions arise concerning a salient-context explanation of the *put*-results when one considers a recent visual-world study by Snedeker and Trueswell (2004), which used different linguistic materials and found that 2-Referent visual scenes do not impose a decisive constraint on parsing choice; rather, such scenes act only as one constraint (among many) that could not single-handedly override opposing lexical preferences. In particular, it was found that adults' interpretations of *with*-phrases of globally ambiguous sentences (e.g., *Feel the frog with the feather*) oscillated between Instrument and Modifier interpretations depending on two modulating sources of information: (1) Verb Type (e.g., whether they heard an Instrument- or a Modifier-biased verb); and (2) Referential Scene Type (a 1-Frog or a 2-Frog Scene). Specifically, 1-Frog Scenes resulted in increased Instrument interpretations and decreased Modifier interpretations as compared to 2-Referent Scenes. Likewise, verb bias influenced these measures as well: increased Instrument-biased verbs resulted in increased Instrument interpretations and decreased Modifier interpretations. 2-Referent Scenes were not sufficient on their own to exclusively evoke a Modifier analysis of *with the feather;* visual context instead appeared to provide only partial support for modification, and was always weighed alongside lexical factors. This coordination of lexical and contextual constraints was observed using both on-line (eye movement) and off-line (hand action) measures.

Given these findings, the results of the *put*-studies are even more surprising: why should one contextual constraint unilaterally override the strong syntactic preferences of *put*? One possibility worth serious examination is that in the *put*-studies, signs of consideration of the Goal analysis in 2-Referent contexts could have gone undetected for methodological reasons. In particular, the primary measure of processing commitments in these studies was the proportion of looks to the Incorrect Goal (the empty towel) – looks that typically occur after hearing *towel* and well into hearing the disambiguating phrase *into the box*. As a result, these looks could have been influenced by this post-ambiguity information, which strongly supports the Modifier interpretation of *on the towel*. It is entirely possible, therefore, that this disambiguating evidence conspired to eliminate signs of consideration of the Goal interpretation in 2-Referent contexts, and maybe even reduced these signs but did not eliminate them in the 1-Referent contexts.[2] (See also Farmer, Anderson, & Spivey, 2007, whose modeling efforts show contextual and post-ambiguity constraints conspiring to mitigate strong verb biases.)

The current study therefore revisits visual-world evidence that supports the claim that visual context can be so salient that it can sometimes completely override countervailing linguistic evidence (i.e., strong verb constraints) that would otherwise guide the parser toward a different analysis (Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). In particular, we report a study in which participants were placed under a time constraint to complete an action, with the intent that this might reveal difficulty overriding lexical biases, which may have gone undetected previously. That is, the absence of a response deadline in previous studies may have veiled any temporal advantage that verb constraints may have had over other information sources (for a general discussion of this issue, see, e.g., McElree & Griffith, 1995, 1998).

In addition, the setup of a listener's 2-Referent contextual environment in the current experiment was designed to construct a spatial contrast between the potential referents; the Target object always appeared on a flat-surfaced platform (e.g., a towel), whereas the Competitor object always appeared inside a container (e.g., a bowl). This contrast permitted new time-course analyses of referent resolution that were not possible in prior studies. In particular, we compared ambiguous and unambiguous materials for how rapidly listeners could use the preposition *on* to fixate the Target (which was on something) rather than the Competitor (which was in something). Past *put*-studies have compared looks to the Target vs. looks to the Competitor, but not under these spatial contrast conditions. Spatial contrasts of this sort

[2] It is also possible that looks to the Incorrect Goal were artificially reduced in 2-Referent Scenes because of confusion over the intended referent for the direct object (*the frog*). If listeners were partially considering the Goal analysis in 2-Referent Scenes, then there is uncertainty about which frog should be the referent, pulling looks toward the two frogs and away from the Incorrect Goal. In 1-Referent Scenes, by contrast, consideration of the Goal interpretation does not introduce uncertainty about the referent of the direct object (because there is only one frog); as such, looks could be drawn to the Incorrect Goal. However, Spivey et al. (2002) rule out this concern via additional experimental conditions (see 3 & 1 condition of Spivey et al., 2002).

have been found to influence real-time processing when the utterances are syntactically unambiguous. In particular, Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002) found that listeners can launch anticipatory eye movements to a container (e.g., a can) as opposed to other objects (e.g., a plate) upon hearing the preposition *inside* in sentences like *Put the cube inside the*...

To summarize our own results in advance, we replicate many of the patterns reported previously (e.g., proportion of looks to an Incorrect Goal during the processing of the temporarily ambiguous PP; Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). Nevertheless, our newly added spatial-contrast measure provides clear evidence that, even in supportive 2-Referent Scenes, some difficulty exists in arriving at the Modifier interpretation of *on the towel* as compared to unambiguous controls (*that's on the towel*) – a result that is consistent with the CBL theory.

### *1.1. Individuals' reliance on lexical constraints across modality*

It is important to ask whether participants' difficulty with arriving at the correct interpretation of ambiguous *put* materials is in fact, as we suggest, arising from countervailing verb preferences. The Snedeker and Trueswell (2004) findings suggest that this is probably the case. One might want to test this, for instance, by doing similar verb manipulations using the Goal/Modifier ambiguity found in the *put* materials. However, it is difficult to adequately manipulate verb biases in an experiment involving sentences with double PPs (...on *the towel into the box.*). The reason is that the use of less biasing verbs (e.g., *move*, *slide*, etc.) would introduce another interpretation of the ambiguous phrase as a locative adjunct of the verb (e.g., *move the apple on the towel/slide the apple on the towel*, etc.), whose referential implication would be the Target area – namely, the towel that the apple is already on – and not the other towel in the scene. Thus, eye fixation patterns would probably not be informative regarding which parse people assigned.

As a result, we took a different, multiple method strategy to test whether lexical constraints may be driving this effect, hypothesizing that individuals might differ in their reliance on lexical constraints to drive the initial structuring of their parsing decisions. That is, if consideration of the Goal interpretation can be observed in the visual-world paradigm even under referential conditions that support the Modifier interpretation, then this consideration might be related to individual differences in the use of lexical cues to structure: listeners who are reliably tempted to take the ambiguous phrase *on the towel* as the Goal argument of *put,* despite contextual or disambiguating evidence to the contrary, may also rely heavily on lexical factors generally under different comprehension settings to drive their parsing choices.

To test this hypothesis, we compared our participants' performance on syntactic ambiguity resolution in the visual-world paradigm with their ambiguity resolution performance in a reading task using a very different kind of syntactic ambiguity. Comparing across different ambiguity types is important to ensure that any observable variation among individuals is not merely a reflection of individual differences in experience with PP-attachment ambiguities in particular. To this end, individuals who participated in our visual-world listening task also completed a reading task

to assess the extent to which some individuals exploit the reliability of lexical preferences more than others to drive processing decisions. The reading task used the Direct Object/Sentence Complement (DO/SC) ambiguity illustrated in (3):

(3) *The gossipy neighbor heard the story had been told to everyone but her.*

Here, the post-verbal Noun Phrase (NP) *the story* could temporarily be considered the direct object of the verb *heard* or the subject of an embedded sentence, which is how the sentence ultimately resolves in this example. The verb *heard* frequently appears with direct objects throughout the language; but, as illustrated here, it also permits a sentence complement. In cases like (3), when the sentence unfolds with an SC continuation, readers often slow down when they reach the post-NP disambiguating region (e.g., *had been*...) and often attempt to reread the sentence (Ferreira & Henderson, 1990, 1991; Garnsey et al., 1997; Trueswell et al., 1993). This processing difficulty suggests that readers initially take the NP *the story* as the DO of *heard* but then experience conflict when the second (subject-less) verb phrase is encountered (*had been told.*) In other words, like in the *put* example above, accumulating lexico-syntactic evidence supports a particular analysis of a temporarily ambiguous phrase, but later syntactic evidence conflicts with this analysis.

The strong ambiguity effect in DO/SC materials need not arise, however: changing the verb in (3) from *heard* to *realized* – a verb that allows a DO but strongly prefers a SC – eliminates signs of processing difficulty at the disambiguating region *had been* (Garnsey et al., 1997; Holmes et al., 1989; Trueswell et al., 1993; see Pickering & Traxler, 1998 for an alternative finding; but see also Elman et al., 2004). Thus, in these studies, an interaction is typically observed between Verb-Bias (DO-bias verbs vs. SC-bias verbs) and Ambiguity (*heard the story*... vs. *heard that the story*...), such that an effect of Ambiguity is observed only in the DO-bias verbs. However, like most behavioral studies, there is variation across individuals – most people show this interaction, but some do not.

As we describe below, we used our reading study as a diagnostic tool to identify those individuals who rely especially on verb information to guide parsing choices (Lexicalist individuals). We expect that those individuals who use verb information in the reading study will also be the ones to show larger consideration of the Incorrect Goal in the visual-world study, regardless of context. If this is the case, we can be fairly confident that Goal interpretations in the *put* study reflect in part the use of lexical (e.g., verb) biases, rather than other available constraints, like possible prosodic cues. The most parsimonious interpretation for such a relationship would be that an individual relies on lexical cues to similar extents across both reading and spoken language modalities.

In addition, it is plausible to expect that individuals who are adept at weighing lexical factors may also be quite good at weighing contextual factors when they are present. Those who are adept at tracking and using fine-grained lexical contingencies should also be quite good at tracking fine-grained referential contingencies, and in theory should facilitate the tracking of such contingencies (see, e.g., Kim, Srinivas, & Trueswell, 2002). If so, then lexically-sensitive individuals ought to show

enhanced sensitivity to both lexical *and* contextual factors in the *put*-study, arriving for instance at the Modifier interpretation more quickly in 2-Referent Ambiguous trials.

Finally, it is important to note that any correspondences that we identify across these two experiments would provide important cross-validation of the reading and visual-world paradigms that are traditionally used to infer interpretation commitments: positive correspondences and correlations across measures would address whether processing difficulty associated with increased reading time and eye fixation patterns in response to spoken instructions are sensitive to the same linguistic and cognitive processes. During reading, interpretation commitments are inferred by detecting processing slowdowns, i.e., signs that things have gone wrong during the comprehension process. That is, increased reading times in the case of ambiguity are typically construed as an indication that the reader initially misinterpreted the ambiguity and started to repair his or her misinterpretation in light of either new linguistic material that was just encountered or contextual material that he or she just realized was relevant. For instance, difficulty reading sentences like (1), compared to unambiguous controls, is typically taken as evidence that readers erroneously considered the unintended meaning of the ambiguous phrase and had to rescind (or re-rank, or reject) that consideration. By contrast, measures of a listener's interrogation of his or her visual environment are not typically used to measure processing difficulty per se. Rather, looks to possible referents while the sentence is unfolding are translated as indicators of what the listener is considering to be the interpretation of the utterance at particular moments within the speech stream (e.g., Cooper, 1974; Tanenhaus et al., 1995). How measures of real-time processing commitments in the visual-world paradigm relate to indices of processing difficulty in reading tasks therefore remains an important open empirical issue.

### *1.2. Experimental preliminaries*

All participants completed the visual-world listening task containing the *put* materials and a reading task containing the DO/SC ambiguity. Again, the idea was to test whether correlated variation among individual syntactic choices could be observed regardless of three important differences: (1) Modality (reading vs. listening in the visual-world paradigm); (2) Type of Ambiguity (DO/SC vs. PP-Attachment); and (3) Task (pressing a button vs. carrying out spoken instructions).

Section 2 below revisits the interactive claims of contextual-guidance in the visual-world task by examining additional indices of syntactic commitments to the Goal interpretation in 2-Referent contexts and by placing participants under a time constraint to carry out an instruction. Using standard measures (e.g., looks to an Incorrect Goal), we replicate most of the previous findings using these materials; however, imposing a response deadline and performing additional time-course measures reveal that listeners do in fact temporarily consider the ambiguous PP (e.g., *on the towel*) as a Goal, even when the visual scene supports a Modifier analysis.

Section 3 reports the results of the reading study. Section 4 then reports co-variation in performance between reading and listening processes to explore the extent to

which individuals are more likely to use lexical constraints to drive parsing decisions, and the extent to which this use is consistent within comprehenders across reading and listening modalities.

#### *1.2.1. Participants*

Forty students from the University of Pennsylvania community volunteered for the study (19 male; all native speakers of American English). Each participant visited the laboratory once for approximately two and half hours; this study was part of a larger-scale individual differences study (hence the long duration of each visit to the lab), which we are exploring in other research and is therefore not relevant to the current discussion. All participants received course credit or were paid $8 per hour. Order of participation in the experiments was randomized and counterbalanced.

## 2. Experiment 1: Reassessing lexical and visual context effects in the *put* task

The design of this study was similar to that of Trueswell et al. (1999) (see also Snedeker & Trueswell, 2004; Spivey et al., 2002; Tanenhaus et al., 1995). However, some important modifications were made to the paradigm in order to further test the effect of visual context on syntactic ambiguity resolution and how it interacts with the linguistic input, particularly in 2-Referent cases. As discussed in detail below, participants were placed under a time constraint that forced speeded responses with the intent of detecting any temporal priority of lexical or contextual constraints on parsing. In addition, the referential scenes were adapted from earlier studies in a way that allowed for a measure of parsing commitments to the Goal interpretation during the ambiguous phrase itself. These changes are marked with Sub-sections 2.1.1.1 and 2.1.2.1.

### *2.1. Methods*

#### *2.1.1. Procedure*

Participants were told that they would listen to and follow prerecorded instructions to manipulate stuffed animals and other toy objects. Each participant sat in front of an inclined platform. There was an aperture at the center of the platform, behind which a digital video camera was positioned that focused on the participant's face. In each quadrant of the platform there was a shelf on which one of several props could be placed. At the beginning of a trial, one experimenter laid out the props and introduced each one using indefinite noun phrases (e.g., *This is a frog, a plate*...).

Prerecorded sound files were then played from a laptop connected to external speakers and to the video camera. On each trial, the participant was first told to look at a fixation point at the center of the display. Then s/he was given two or three single sentence commands involving the props. The participant heard the first command, performed that action, and then heard the second command. A second camera, placed behind the participant, recorded the participant's actions and the locations of the props.

*2.1.1.1. Changes from previous studies.* Unlike previous experiments using this method (e.g., Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999), a time constraint was imposed on listeners to complete an action after hearing each instruction. The intention was to encourage failure to revise initial parsing commitments in some participants. Specifically, a tone sounded 750 ms after the offset of an instruction, which signaled to the listener that his or her action must be completed by this time. Thus, listeners had to rapidly commit to an interpretation and start executing an action before this tone sounded.

On a few occasions the participant asked for clarification or requested that the instruction be repeated. The experimenter responded by playing the sound file again but the eye movements were coded for only the initial presentation of the sentence.

### 2.1.2. Stimuli

Sixteen target stimuli were constructed and modeled after Trueswell et al. (1999). The first sentence of each critical trial contained the verb *Put* and appeared in one of two sentence types as shown in example (2), repeated here as (4):

(4) a. *Put the frog on the napkin into the box.* (Ambiguous)
b. *Put the frog that's on the napkin into the box.* (Unambiguous)

In (4a) the PP *on the napkin* is temporarily ambiguous between indicating either a location/goal (where the frog should be put) or a restrictive modifier (indicating that the frog to be put somewhere is currently on a napkin). The inclusion of *that's* (4b) removes this temporary ambiguity and syntactically imposes the Modifier analysis of *on the napkin.* All instructions were prerecorded by a female speaker. As best as possible, the prosody in these utterances was intended to be unbiased with respect to the relevant parsing alternatives: large prosodic breaks and pauses were avoided. Prosody was, however, controlled within the experimental design, since the same audio files were used across 1- and 2-Referent conditions.

The referential scene was also manipulated by changing critical features of the display configuration between target trials. 1-Referent Scenes (Fig. 1A) contained a Target animal (a toy frog sitting on a napkin), an Incorrect Goal (a second, unoccupied napkin), a Correct Goal (a box), and a Competitor animal (a toy horse sitting in a basket). These scenes should encourage the Goal interpretation of the phrase *on the napkin* in Ambiguous sentences because modification of *the frog* would be redundant in this situation. 2-Referent Scenes, by contrast, contained two toy referents of the same kind of animal, e.g., two toy frogs (Fig. 1B). These scenes were configured just like 1-Referent Scenes except that a second toy frog replaced the Competitor toy horse. These displays should support an NP-Modifier interpretation of *on the napkin* because the definite NP *the frog* does not uniquely specify which frog is the intended referent. Thus, the PP *on the napkin* is necessary to pick out which frog is being referred to.

Both factors were manipulated within participants. Four presentation lists were designed containing 16 target trials pseudorandomly intermingled with 26 filler trials. Of the target items, eight were Ambiguous (four within 1-Referent Scenes and

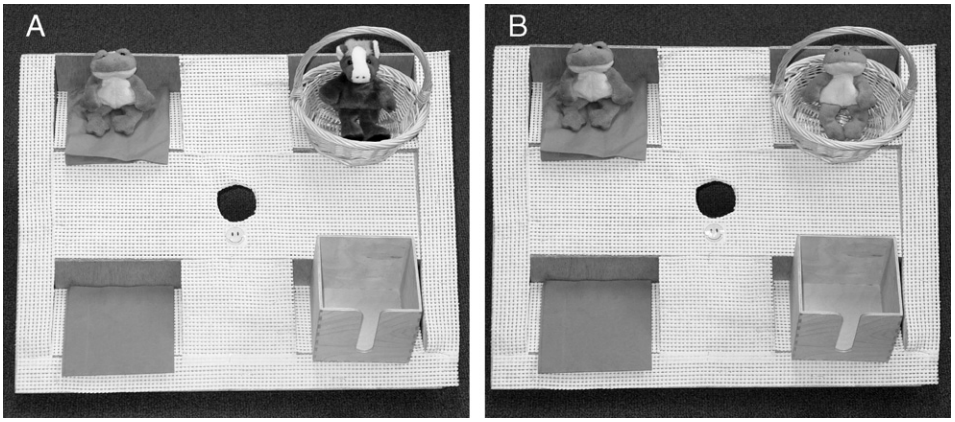


Fig. 1. (A) 1-Referent Scene type, which supports the Goal interpretation for *on the napkin*. These scenes contain, for instance, a Target animal (frog on a napkin), a Competitor animal (horse in a basket), an Incorrect Goal (an empty napkin), and a Correct Goal (a box). (B) 2-Referent Scene type, which supports the Modifier interpretation for *on the napkin.* These scenes contain the same props as 1-Referent Scenes except that a second frog replaces the horse as the Competitor animal.

four within 2-Referent Scenes), and eight were Unambiguous (four within 1-Referent Scenes and four within 2-Referent Scenes). Each critical trial was rotated through these four conditions, generating the four different presentation lists, and each participant was assigned to one of the four lists. Lists were run in both forward and reverse orders alternating between participants. Displays for filler trials looked similar to those of critical item displays so that participants could not predict what sort of instruction they would hear. Each trial (both target and filler) contained two or three sentences; the first sentence of each trial was always the critical sentence (e.g., *Put the frog…*). The second and third sentences of every trial also served as built-in fillers as they were syntactically unambiguous and distracted from the experimental manipulations. These instructions, for instance, asked participants to *Now spin the duck around* or *Now make the other animal stand on its head.* Filler items also included unambiguous forms of the verb *Put* and indicated its destination in various ways (e.g., *Put the X in the Y, next to the Z,* etc.). Participants completed five practice trials before the start of the task in order to become familiar with the procedure. These practice items were similar to filler trials.

*2.1.2.1. Changes from previous studies.* Unlike previous studies using these visual and linguistic stimuli (e.g., Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999), the Target and Competitor animals in this study were designed to differ on each trial concerning whether one was on something and the other was in something (e.g., a frog on a napkin vs. a horse/frog in a basket). Half of all target trials had the Target animal on something and the Competitor animal in something, whereas the opposite was true for the other half of trials (e.g., *Put the horse in the bowl onto the plate).* Prior visual-world studies of this sort (Spivey et al., 2002; Tanenhaus et al.,

1995; Trueswell et al., 1999) did not contrast between the prepositions *in* and *on* in terms of the platforms that carried the objects; all the objects in earlier studies were always on something. This new setup permitted us to examine whether listeners immediately follow a Modifier interpretation in 2-Referent Scenes. In particular, if the presence of two frogs is enough to guide listeners toward a Modifier interpretation of *on the napkin* without consideration of the Goal interpretation, then the preposition itself (*on*) should immediately allow listeners to distinguish between the two referents (the frog that's on a napkin and not the one that's in a basket) (Fig. 1B) (see Chambers et al., 2002). Thus, one would expect listeners' fixations to converge on the Target frog (the one on the napkin) upon hearing *on.* As such, the preference to look at the Target animal over the Competitor animal should increase at the same rate for both Ambiguous and Unambiguous linguistic stimuli, because the 2-Referent contextual constraint would in effect be enough to essentially mark the phrase *on the napkin* unambiguously as a Modifier.

On the other hand, if listeners do consider the Goal analysis for Ambiguous sentences despite the referential support from the visual environment in 2-Referent Scenes, which increase the probability of getting an NP Modifier compared to 1-Referent Scenes, then eye movements should converge early on the Target animal *for Unambiguous cases only.* This is because during Ambiguous instructions, listeners may look toward the Incorrect Goal (e.g., the empty napkin) in order to satisfy the verb-driven Goal analysis. Furthermore, listeners may exhibit a tendency to look at the Competitor frog because the temporarily ambiguous PP *on the napkin* would not be provisionally interpreted as a Modifier; this frog is not already on a napkin and thus needs to be put there (i.e., to satisfy, pragmatically, the (incorrect) Goal interpretation) (Meroni & Crain, 2003; Trueswell et al., 1999). That is to say, the difference in looking times to the Target vs. the Competitor might not be as great in the Ambiguous condition compared to the Unambiguous condition.

### *2.1.3. Coding*

Following Snedeker and Trueswell (2004), gaze direction was coded from the video of the listener's face on a frame-accurate digital VCR (SONY DSR-30) with audio-lock so that direction of eye gaze could be determined with respect to the speech stream on a scale of every 33 ms. The trained coder documented the beginning of each target trial by logging the onset of the verb *Put* and the listener's eye position at that moment. From that point on, frame-by-frame changes in gaze direction were recorded with the VCR's timestamp indicating when a new eye-movement occurred; and the coder then noted the new direction of fixation. Fixation direction was coded as one of the following: on one of the quadrants (e.g., upper left, lower right, etc.), in the center of the platform display, or elsewhere (i.e., away from the experimental scene). "Track loss" was coded if the participant's eyes were closed or occluded, for example, by a reaching arm in front of the face camera's lens. A trial was dropped from analysis if track loss accounted for more than 33% of the frames (this made up less than 5% of the analyzed data). A trial's offset was coded when a participant released an object after performing an action. The actual object on which a listener was fixating was later confirmed by the scene videotapes, which captured

where in the scene each object was placed, so that "upper left" could be matched with "empty napkin", for example. This procedure allowed the trained coders to be blind to the experimental conditions when logging each direction of gaze.

Participants' hand-actions on critical trials were coded by inspecting the scene videotapes. An action was coded as correct if the Target animal was moved directly to the Correct Goal without any involvement of the Incorrect Goal. Actions were also coded as correct if the Target animal was moved directly to the Correct Goal along with the modifying object that shared the Target's quadrant (e.g., the napkin on which the frog was sitting; this accounted for only 3% of the trials overall). Any actions involving the Incorrect Goal were coded as errors and can be categorized most frequently as 'Hopping' errors, in which the Target animal (or Competitor animal in 2-Referent Scenes) was moved first to the Incorrect Goal (e.g., the empty napkin) and *then* to the Correct Goal (e.g., the box) (see Trueswell et al., 1999).

### 2.2. Results and discussion of Experiment 1

The results of this experiment will be presented in two parts. First, we will discuss analyses of participants' actions, i.e., how well they carried out the movement of the objects in response to the instructions. We will then present the analysis of participants' eye movements.

As we discuss below, the data as a whole reveal signs of processing difficulty on Ambiguous trials, even in supportive 2-Referent contexts. Participants carried out more incorrect actions on Ambiguous as compared to Unambiguous trials, an effect which held even for 2-Referent Scenes. Even when participants carried out the correct action, their eye movement patterns indicated that they had difficulty correctly interpreting Ambiguous sentences, again even in 2-Referent Scenes. Consistent with our expectations however, the specific measure of looks to the Incorrect Goal was not sufficient for detecting this difficulty in 2-Referent Scenes. Difficulty with 2-Referent Ambiguous trials was instead manifested in other ways. In particular, relative to syntactically unambiguous controls, 2-Referent Ambiguous trials resulted in listeners showing the following: (1) increased uncertainty over which animal was the intended referent (Target vs. Competitor looks); (2) extended dwell times on the Target; and (3) delayed looks to the Correct Goal (the empty box) upon hearing the second PP *into the box*. Taken together, the data suggest that 2-Referent visual contexts reduced – but did not completely eliminate – the difficulty associated with arriving at the Modifier interpretation of the ambiguous phrase *on the napkin* in the presence of the verb *put*. The near elimination of looks to the Incorrect Goal within 2-Referent Ambiguous trials appears to be related to listener uncertainty over which animal (e.g., which frog) is being referred to (drawing looks to the Competitor and/or Target animal). Even when a listener is looking to the correct (Target) animal, they linger a little longer there as compared to the Unambiguous trials.

#### 2.2.1. Action responses reveal difficulty in 2-Referent Ambiguous contexts

Fig. 2 presents the proportion of incorrect actions carried out by participants. All but two of these errors involved moving an animal to the Incorrect Goal (e.g., the

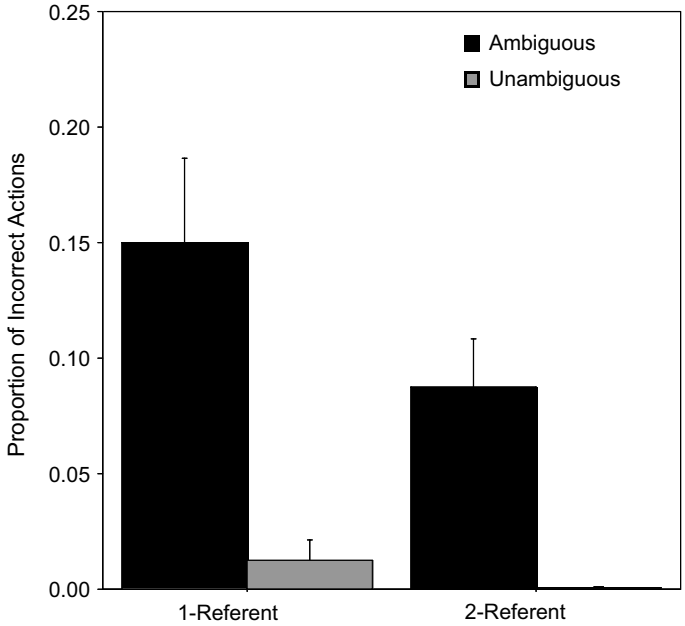


Fig. 2. Errors in hand actions for the *put* study across conditions. These rates are similar to those found for adults in Trueswell et al. (1999). Error bars reflect the 95% confidence interval.

empty napkin). The remaining two errors involved moving the Competitor animal (e.g., the frog in the basket) to the Correct Goal (e.g., the empty box), both instances occurring in 2-Referent Scenes. As can be seen in the figure, errors occurred almost exclusively for temporarily ambiguous sentences, and occurred most often in 1-Referent Ambiguous Scenes, although a sizeable proportion of errors also occurred in 2-Referent Scenes.

Subject and Item means were computed and entered into separate ANOVAs with three factors: Ambiguity (Ambiguous, Unambiguous); Context (1-Referent, 2-Referent); and List/Item-Group factor (4 lists in the subject analysis and 4 groups in the item analysis).[3] These analyses revealed a main effect of Ambiguity [$F1(1,36) = 24.81$, $p < .01$; $F2(1,12) = 22.52$, $p < .01$], and a main effect of Referential Scene [$F1(1,36) = 4.10$, $p < .05$; $F2(1,12) = 8.57$, $p < .05$]. Although more errors were observed in 1-Referent Ambiguous than 2-Referent Ambiguous conditions (15% vs. 8%), no interaction between Ambiguity and Context was found (both

[3] Throughout the paper, whenever ANOVAs were conducted on proportions, identical ANOVAs were also conducted on an arcsin transformation of the data, arcsin $((2 * p) - 1)$. This was done to adjust for the fact that the proportion ($p$) is bounded at 0 and 1. For the sake of accuracy, we report $F$ values and $p$ values from the transformed data and the means from the untransformed data (for clarity). Unless otherwise noted, statistically significant effects that were found for the transformed data were also significant for the untransformed (proportion) data.

$F$s < 2). Moreover, a reliable effect of Ambiguity was observed in both 1-Referent Contexts [$F1(1,36) = 17.37$, $p < .001$; $F2(1,12) = 19.69$, $p < .005$] and 2-Referent Contexts [$F1(1,36) = 22.70$, $p < .001$; $F2(1,12) = 12.46$, $p < .005$] when analyzed independently.

This pattern of errors indicates that listeners had difficulty resolving the temporary ambiguity associated with the PP *on the napkin*, even when the context supported a Modifier interpretation (2-Referent Scenes). The 2-Referent Scenes reduced the difficulty but did not completely eliminate it. Given prior studies showing that verb information influences ambiguity resolution in 2-Referent Scenes (Snedeker & Trueswell, 2004), the most parsimonious conclusion is that the effect of Ambiguity in the present 2-Referent Scenes reflects the lexical constraints associated with the verb *put* and the preposition *on*, which strongly support the Incorrect Goal interpretation of the first PP (e.g., *on the napkin*).

As mentioned above, participants in this study were placed under a time constraint to carry out their actions. This change could have increased errors in responding accurately to the instructions. Somewhat surprisingly though, the error rates and the pattern across conditions are quite similar to those reported for adults in the Trueswell et al. (1999) study (1-Referent Ambiguous: 17%; 1-Referent Unambiguous: 0%; 2-Referent Ambiguous: 7%; 2-Referent Unambiguous: 0%). Participants in that study were encouraged to go quickly but were not given a response deadline. The similarity in error rates suggests that imposing a response deadline did not increase the likelihood that participants would fail to revise initial misinterpretations.

Other studies that have used similar materials do not report error rates (Spivey et al., 2002; Chambers et al., 2004). However, Tanenhaus (personal communication) reports that a small number of errors (two or three) were observed in the Spivey et al. (2002) study, all of which occurred in the 1-Referent Ambiguous condition. Chambers (personal communication) also reports a small number of errors for the Chambers et al. (2004) study. In particular, three actions, all on Ambiguous trials, involved moving the Target to the Incorrect Goal. Such low numbers of errors clearly justify not reporting them in the published works. However, it is important to note that far fewer subjects and items were used in these past studies as compared to the present study. When this is taken into account, the results are quite similar across all experiments. For instance, Experiment 1 in Chambers et al. (2004) used sixteen subjects, but only eight items per list, resulting in four ambiguous items per subject (and hence two per contextual condition). This means that overall, there were 64 Ambiguous trials contributing to the data (all in 2-Referent Scenes). This is three out of 64, or roughly a 5% error rate. The present 2-Referent error rate was 8% for Ambiguous items. In Experiment 2 of Chambers et al., there were three cases where listeners selected the Competitor object for the action (one even in the Unambiguous condition), plus three other errors. Again, taking into account the number of subjects (24) and items (12), these proportions are quite similar to those reported here. Thus the error rates that we report in the present experiment are in line with previous studies; it is just that past studies had fewer items and fewer subjects, making these low probabilities difficult to detect. Together then, the present and past studies indicate that

(a) there is some difficulty on Ambiguous trials even in contextually supportive contexts; and (b) the overall pattern of difficulty is unlikely to be related to time pressure.

### 2.2.2. Eye movements

Eye movement analyses were conducted on correct-action trials only. We report three different types of analyses, the first two being similar to those used in previous studies: (1) looks to the Incorrect Goal (e.g., the empty napkin); (2) preference to look at the Target animal over the Competitor animal (e.g., the frog on something vs. the frog in something); and (3) looks to the Correct Goal (the empty box). The data indicate that although processing difficulty was reduced on 2-Referent Ambiguous trials as compared to 1-Referent Ambiguous trials, it was not completely eliminated relative to Unambiguous controls. The measure of looks to the Incorrect Goal was not able to detect this difficulty in 2-Referent Ambiguous trials, but the other measures show that listeners did have trouble arriving at the correct interpretation even in this condition.

*2.2.2.1. Looks to the Incorrect Goal.* As shown in Fig. 3, the average proportion of time spent looking at the Incorrect Goal (from the onset of *napkin* until an action was carried out) was larger in the 1-Referent Ambiguous condition as compared to its Unambiguous control. A similar increase was not seen in 2-Referent contexts. ANOVAs on Subject and Item means, using the same factors as above, supported this interpretation. In particular, although there was an effect of Ambiguity [$F1(1,36) = 16.89$, $p < .001$; $F2(1,12) = 15.94$, $p < .005$] and Referential Context [$F1(1,36) = 11.83$, $p < .005$; $F2(1,12) = 8.21$, $p < .05$], these two factors interacted [$F1(1,36) = 9.86$, $p < .005$; $F2(1,12) = 8.75$, $p < .05$].[4] Within 1-Referent Contexts there was a reliable effect of Ambiguity [$F1(1,36)= 26.64$; $p < .001$; $F2(1,12) = 30.05$, $p < .001$] but a similar effect was not present in 2-Referent Contexts [$F$s $< 1$].

As shown in Fig. 4, a similar pattern is observed over time relative to the onset of the first preposition (e.g., *on the napkin*...). Looks to the Incorrect Goal begin to increase after the onset of the noun (*napkin*), and are largely restricted to the 1-Referent Ambiguous condition. To test the significance of this effect, we calculated the proportion of time spent looking at the Incorrect Goal during a 1-s window after the onset of the noun (*napkin*). This window was offset by 200 ms (i.e., 200–1200 ms after *napkin*) to take into account the time it takes to launch an eye movement after it has been programmed (Matin, Shao, & Boff, 1993). Subject and Item means were entered into separate ANOVAs having the same factors as above. Even when restricting to this time window, although there is a main effect of Ambiguity [$F2(1,36) = 12.89$, $p < 005$; $F2(1,12) = 12.55$, $p < .005$] and Referential Scene [$F1(1,36) = 18.42$, $p < .001$; $F2(1,12) = 13.12$, $p < .005$], there is also an interaction between the two factors [$F1(1,36) = 17.06$, $p < .001$; $F2(1,12) = 9.01$, $p < .05$]. Again, the effect of Ambi-

[4] An unexpected (marginally significant) main effect of List was observed by the subject analysis [$F1(3,36) = 2.87$; $p = .05$], which also interacted with Context and Ambiguity [$F1(3,36) = 6.05$; $p < .005$].

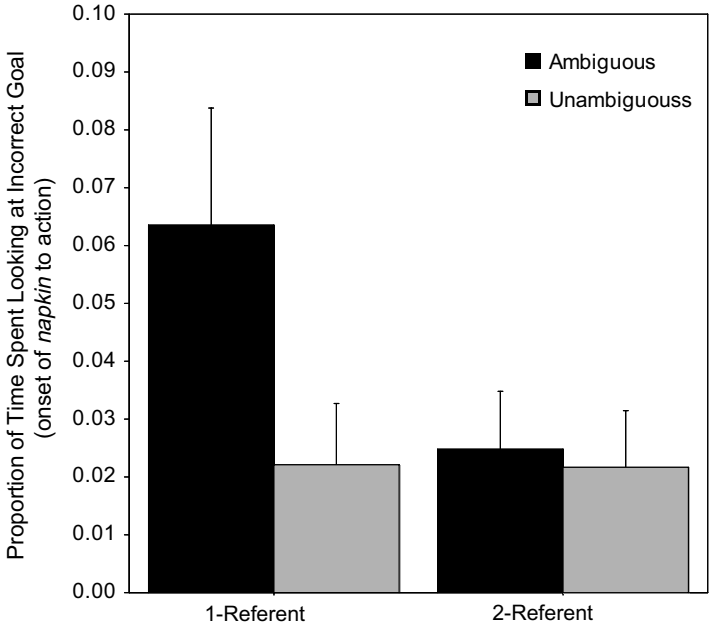


Fig. 3. Proportion of time spent looking at the Incorrect Goal (e.g., an empty napkin) from the onset of *napkin* until a hand action was completed. Main effects of Referential Context and Ambiguity were observed as well as an interaction between these two factors (i.e. 1-Referent Ambiguous contexts elicited the greatest proportion of looks to the Incorrect Goal; see text). Error bars reflect the 95% confidence interval.

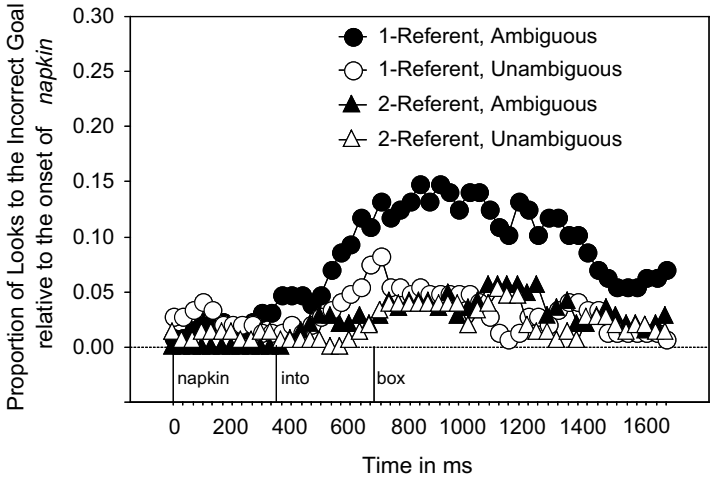


Fig. 4. Proportion of looks to the Incorrect Goal over time relative to the onset of *napkin*... across all four conditions.

guity was significant in the 1-Referent Context [$F1(1,36) = 18.47$, $p < .001$; $F2(1,12) = 12.21$, $p < .005$] but not in the 2-Referent Contexts [$F1(1,36) = 1.14$; $p > .2$; $F2(1,12) = 2.85$; $p > .1$].

If we take this pattern at face value, it means that when participants carried out the *correct* action, they temporarily considered the Goal interpretation of *on the napkin* in 1-Referent contexts but not in 2-Referent contexts. However, our action–error analyses showed that listeners in 2-Referent contexts do indeed consider the goal interpretation on some trials, and even choose it as their final interpretation. This would mean that on a small proportion of trials, listeners fail to take into account relevant aspects of the scene, with the result being errors in actions. When listeners *do* take into account the context, they behave flawlessly in both actions and their initial eye movement patterns. However, such a categorical explanation of the results is ruled out when one examines other aspects of the eye movement patterns: these other data, discussed below, indicate that the goal interpretation of *on the napkin* was partially considered even in 2-Referent contexts.

*2.2.2.2. Difficulty arriving at the Modifier interpretation in 2-Referent Scenes: Target vs. Competitor looks.* As discussed above, referential scenes were designed such that the Target and Competitor animals always differed regarding whether one was in something and the other was on something. This allowed us to obtain a measure of when people correctly interpreted *on the napkin* as a Modifier, while the ambiguous phrase was being heard. In particular, an analysis was conducted on each participant's preference to look at the Target animal instead of the Competitor animal over time as the ambiguous PP (e.g., *on the napkin*) unfolded. Preference for the Target animal is indicative of interpreting this phrase as an NP Modifier because such an interpretation would allow the listener to determine which frog is being referred to. Target preference was quantified as the proportion of time spent looking at the Target minus the proportion of time spent looking at the Competitor, and was plotted as the speech unfolded in time on the millisecond scale relative to the onset of the preposition *on* (Fig. 5a).

As can be seen in Fig. 5, only Unambiguous sentences produced a sharp increase over time in the preference to look at the Target animal upon hearing *on* (*i.e., the frog that's on*...). For the syntactically Ambiguous cases, listeners appear to compute this interpretation more slowly as shown by the more gradual increase over time in Target Preference compared to sentences in the Unambiguous condition. This pattern strongly suggests that 2-Referent Scenes alone are not sufficient to guide the listener exclusively toward the Modifier interpretation of the ambiguous sequence *Put the frog on the napkin*...

ANOVAs were performed in five different 200-ms time windows, starting from the onset of *on*. Significant differences (marginally significant by Item analysis) between the Unambiguous and Ambiguous conditions begin to emerge in the 200- to 400-ms window [$F1(1,36) = 6.15$, $p < .05$; $F2(1,12) = 3.67$, $p < .07$]. This effect becomes significant by both Subject and Item analyses in the 400- to 600-ms window [$F1(1,36) = 15.24$, $p < .001$; $F2(1,12) = 6.16$, $p < .05$], and persists marginally by Subject analysis (but significantly by Item analysis) until 800 ms after the onset of

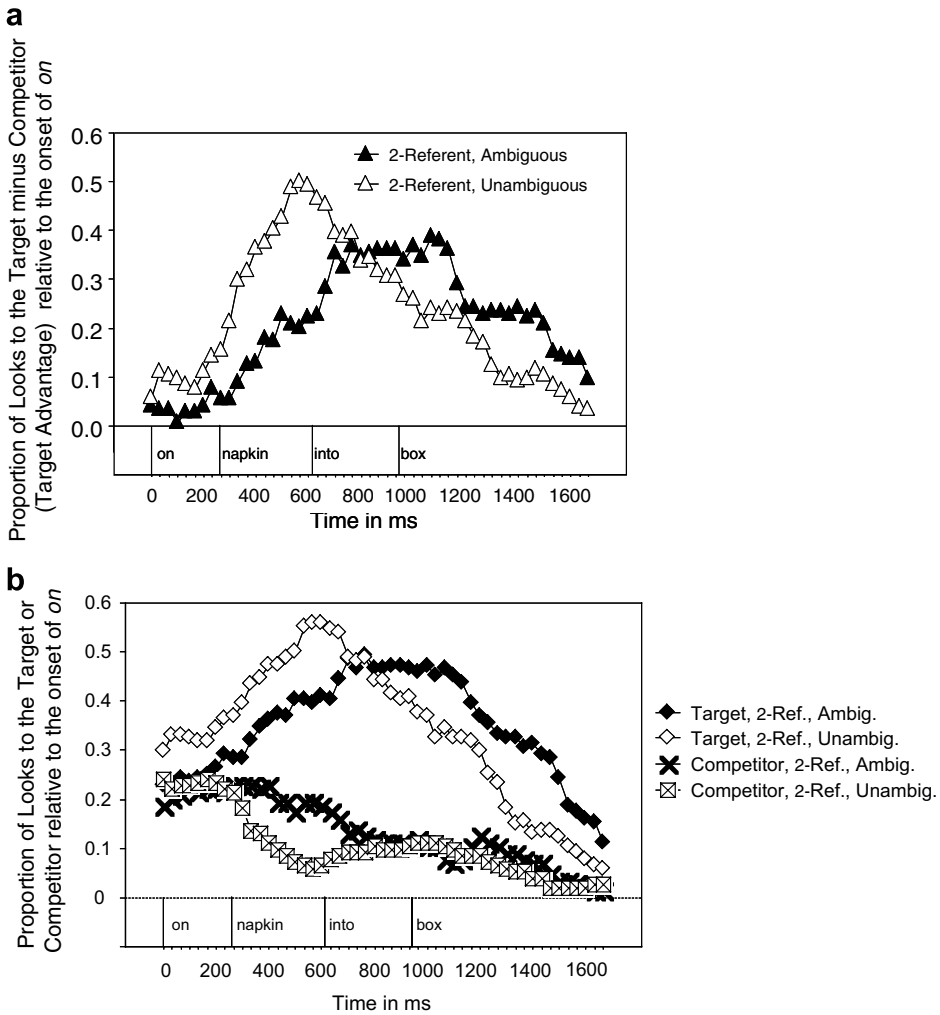


Fig. 5. (a) *Target advantage*: Probability of fixating the Target animal (e.g., the frog on the napkin) instead of the Competitor animal (e.g., the frog in the basket) over time for 2-Referent Ambiguous and Unambiguous conditions. Zero milliseconds on the *x*-axis marks the onset of the preposition *on* in the temporarily ambiguous phrase *on the napkin.* (b) Probability of fixating the Target animal and Competitor animal separately over time in 2-Referent Contexts.

the preposition [$F1(1,36) = 3.32$, $p < .07$; $F2(1,12) = 8.41$, $p < .05$]. This effect then disappears between 800 and 1,000 ms after *on* [$F1(1,36) = .58$, $p > .4$; $F2(1,12) = .03$, $p > .8$].

This temporal advantage for the syntactically Unambiguous version compared to the Ambiguous version cannot be explained away as a difference in the time between hearing *frog* and *on* in the Unambiguous condition (because of the intervening *that's* and possible durational differences of the preceding noun). In particular, the time

between the onset of the first noun (e.g., *frog*) and the first preposition (e.g., *on*) was 294 ms in the Ambiguous condition and 421 ms in the Unambiguous condition, a difference of only 127 ms. Informal inspection of Fig. 5a shows that the rise in the Unambiguous condition precedes the Ambiguous condition by more than 127 ms.[5] Moreover, the peak Target Advantage is greater in the Unambiguous as compared to the Ambiguous condition; thus, listeners are not only delayed in interpreting *on* as a Modifier in the Ambiguous case but they show greater uncertainty overall about which object *the frog* is referring to.

For completeness, time-course plots are also provided for looks to the Target and looks to the Competitor separately (Fig. 5b). Given that the Target Advantage measure is a difference score, it is possible that the observed effects of Ambiguity were due to changes only in Target looks, or changes only in Competitor looks. However, as can be seen in the figure, Ambiguity effects are evident in both the Target looks and the Competitor looks in 2-Referent Scenes. Listeners consider the Target more quickly and exclude the Competitor more quickly in Unambiguous Modifiers than in Ambiguous modifiers. Indeed, ANOVAs that were performed separately on Target looks and Competitor looks support this observation, revealing significant effects of Ambiguity. For looks to the Target, the difference emerges in the 200–400 ms window after the onset of the preposition: [$F1(1,36) = 8.19; p < .01; F2(1,12) = 4.11; p < .07$]. For looks to the Competitor, this difference became reliable in the 400–600 ms window after the preposition [$F1(1,36) = 9.09, p < .005; F2(1,12) = 5.24, p < .05$].[6]

These Target Advantage measures strongly suggest that listeners were initially uncertain about the correct interpretation of *on the napkin* in the 2-Referent Ambiguous condition. Context supported one interpretation and the lexical biases supported another. Partial support of the Goal interpretation of *on the napkin* produced uncertainty about the referent of the direct object NP *the frog*.

In fact, relative to Unambiguous controls, listeners appeared to prolong inspection of the Target prior to carrying out an action (see right side of Fig. 5b). This could indicate some delay in correctly executing the action in Ambiguous as compared to Unambiguous conditions. We consider this in the next analysis.

*2.2.2.3. Looks to the Correct Goal.* Fig. 6 plots looks to the Correct Goal relative to the onset of the second PP (*into the box*). As can be seen in Fig. 6a, collapsing across

[5] It is also not the case that the mere passage of time is predicting looks to the Target. On an item by item basis within any given 200 ms time window, the proportion of time spent looking at the Target did not correlate with the corresponding time in ms between the onset of the first noun (e.g., *frog*) and first preposition (e.g., *on*) within Ambiguous or Unambiguous conditions (all $ps > 0.1$).

[6] Inspection of Fig. 5b raises another concern, which is that there appears to be a baseline difference at the preposition *on*. Participants are looking at the Target more in the Unambiguous condition compared to the Ambiguous condition, even at frame zero. One way of addressing this concern (which was suggested by a reviewer) is to perform a series of contingent analyses in which one examines changes in the proportion of looks separately for trials on which listeners were already looking at the Target, Competitor, or elsewhere at the onset of the preposition *on*. We report these analyses in Appendix A. In brief, the results support the conclusions drawn here. However, we raise questions about the appropriateness of using such analyses.

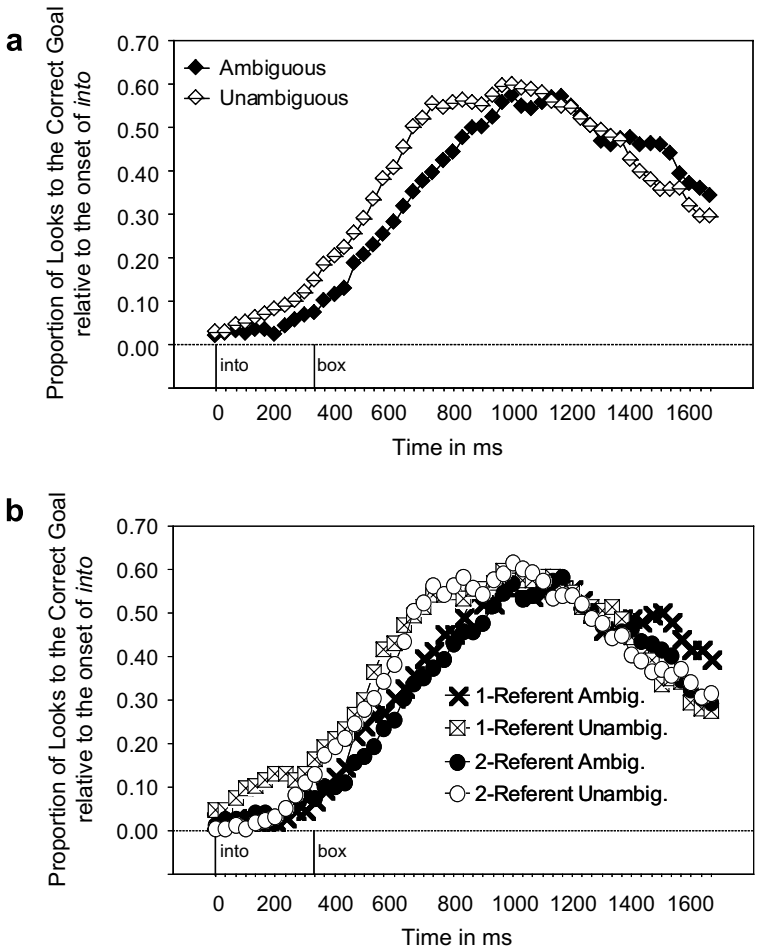


Fig. 6. (a) Proportion of looks to the Correct Goal over time in 2-Referent Scenes, relative to the onset of the second PP (e.g., *into*; marked by 0 ms on the *x*-axis), split by Ambiguity. (b) Proportion of looks to the Correct Goal over time, relative to the onset of *into*, split by Referential Scene and Ambiguity.

1- and 2-Referent contexts shows that there is a sizeable delay in looking to the Correct Goal for Ambiguous stimuli as compared to Unambiguous stimuli. Fig. 6b shows that this delay is quite similar in both 1- and 2-Referent scenes. Proportion of time spent looking at the Correct Goal during a 1-s window (200–1200 ms) after the onset of the second PP *into the box* was calculated and entered into Subject and Item ANOVAs with the same factors as above. Consistent with the pattern seen in the figure, there was a significant effect of Ambiguity [$F1(1,36) = 17.13$, $p < .001$; $F2(1,12) = 11.12$, $p < .005$], with Unambiguous items showing an advantage over Ambiguous items, no effect of Referential context [$F1 = 1.09$; $F2 = 0.76$], and no interaction [$F$s $< 1$].

This pattern suggests that even when correctly interpreting these sentences, participants are slowed in arriving at the Correct Goal in Ambiguous as compared to Unambiguous instructions.

*2.2.2.4. Masking of Incorrect Goal looks in 2-Referent contexts.* Perhaps the most compelling demonstration that Incorrect Goal looks are being masked under 2-Referent conditions comes from examining the eye movements to this region on error trials – trials on which participants were actually moving an animal to the Incorrect Goal. Fig. 7 re-plots the proportion of looks to the Incorrect Goal on correct Ambiguous trials and compares them with looks to the Incorrect Goal on error trials. When listeners incorrectly pick the Goal analysis (i.e., error trials), they rapidly look to the Incorrect Goal in the 1-Referent Scenes, but not in the 2-Referent Scenes. In fact, they only look to the Incorrect Goal in 2-Referent Scenes very late in the trial, when they are starting to carry out their action (which typically involves this object). This pattern most likely arises because in 2-Referent Scenes, partial consideration of the Goal analysis results in spending more time looking at the two frogs, and hence delays (even masks) any looks to the Incorrect Goal. In the 1-Referent Scenes, by contrast, they know which frog is being referred to (because there is only one frog) and thus look right away to the Incorrect Goal.

It is important to note that on correct trials, eye position is also drawn away from the Incorrect Goal in part because people are spending a prolonged amount of time inspecting the Target (see above). People are having some difficulty establishing the correct referent and dwell on the Target longer. This is important because Spivey

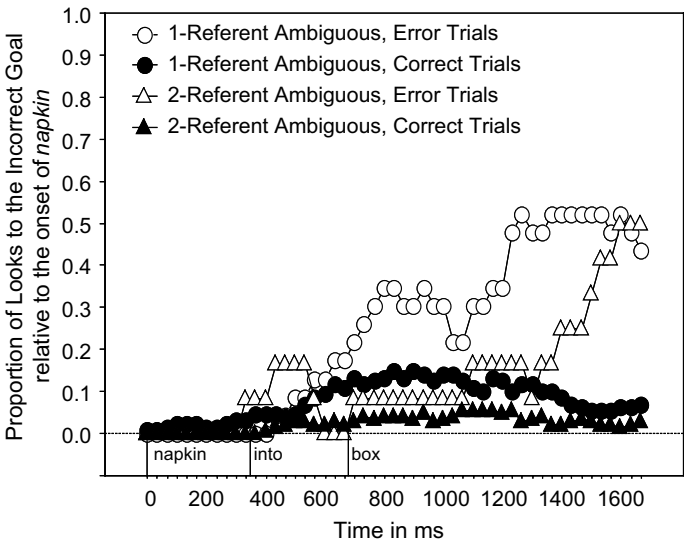


Fig. 7. Proportion of looks to the Incorrect Goal over time as the Ambiguous phrase unfolds (relative to the onset of *napkin*), split by correct and incorrect action trials.

et al. (2002) were also concerned about the masking of Incorrect Goal looks in 2-Referent Scenes, though they focused on how referential ambiguity would attract looks to the Competitor. To rule out this explanation of their data, they developed an ingenuous control condition, called the 3 & 1 condition. Three Competitors were placed on the Competitor platform (e.g., three apples on a plate) and the Target was a singleton (one apple on a napkin). When participants heard *Put the apple on the napkin into the box*, they looked at the Target only, not the Competitors, but still did not look at the Incorrect Goal. They argued that this ruled out the possibility that Incorrect Goal looks were masked because individuals were drawn to the Competitor. Although correct, our own analyses suggest that this may not always be an adequate control: participants may look to the Target for a prolonged period on Ambiguous trials, also drawing looks away from the Incorrect Goal. Indeed, Spivey et al. note the same numerical pattern as we report here. Specifically, they say about their own results: "For the 2-Referent context. . .proportion of fixations on the target referent peaked about 500 ms after the end of the goal word (e.g., "box") for the ambiguous condition and about 300 ms after the goal word for the unambiguous instruction. There were relatively few fixations of the Incorrect Goal in either condition. Proportion of fixations on the Correct Goal exceeded proportion of fixations on the target referent 1133 ms after the end of the sentence in the unambiguous condition and 1266 ms after the end of the sentence in the ambiguous condition. This may suggest a mild ambiguity effect in the 2-Referent contexts".[7]

*2.2.2.5. Correlation in processing difficulty across 1- and 2-Referent contexts.* The implicit assumption behind the discussion of our data is that consideration of the Competitor in 2-Referent Ambiguous conditions is an indication that listeners are having difficulty arriving at the Modifier interpretation of *on the napkin* because they are partially considering the Goal analysis as well. There are, however, essentially no looks to the Incorrect Goal in this condition on correct-action trials, which we believe is because the listener's attention has been drawn to the Competitor animal (and later to the Target animal). These measures suggest, counter-intuitively, that looks to the Incorrect Goal in 2-Referent Scenes may have been blocked for reasons having to do with considering the Goal analysis: partial consideration of the Goal analysis resulted in referential uncertainty over which animal was being referred to.

If this is the case, we might expect that consideration of the Competitor animal in the 2-Referent Scenes would be correlated positively on a subject-by-subject basis with consideration of the Incorrect Goal in 1-Referent Ambiguous condi-

[7] It is also important to note here one puzzling yet potentially relevant data pattern observed in earlier studies, which is that there were sometimes fewer looks to the Incorrect Goal in the 2-Referent Ambiguous condition compared to the 2-Referent Unambiguous condition, though this difference was never significant (e.g., Spivey et al., 2002, Fig. 3B, p. 468). Nevertheless, this occasional trend of reduced looks to the Incorrect Goal in 2-Referent Ambiguous compared to Unambiguous trials might be explained (as in the current study) by these looks having been drawn away from the Incorrect Goal and toward the Target and Competitor objects instead, due to uncertainty as to which referent is the intended one. Consequently, such a pattern could possibly be reflecting consideration of the Goal interpretation, a pattern consistent with what is observed (and directly tested) in the present experiment (see also Appendix A).

tions. This prediction was confirmed for correct action trials. For each subject we calculated the proportion of time spent looking at the Competitor animal in 2-Referent Ambiguous conditions (arcsin transformed) for a 1-s time window beginning 200 ms after the onset of the preposition *on*. We also computed the proportion of time spent looking at the Incorrect Goal in 1-Referent Ambiguous conditions (arcsin transformed) for a 1-s time window beginning 200 ms after the onset of *napkin*. These two measures had a reliable positive correlation ($r = .57$; $p < .0005$) (Fig. 8). In other words, under Ambiguous conditions, those participants who were most tempted to take the Competitor animal as the referent in the 2-Referent Scenes were also those who were tempted to consider the empty napkin as the Goal in 1-Referent Scenes. This correlation suggests that looks to the Competitor animal are indeed reflecting difficulty resolving the syntactic ambiguity associated with *on the napkin* in 2-Referent Scenes.

### 2.3. Summary of Experiment 1

Prior visual-world studies using similar *put* materials, like those in example 2 above, found that the presence of two potential referents (e.g., 2 frogs) guided listeners toward a Modifier interpretation of *Put the frog on the napkin...*, despite strong lexical biases that supported a Goal interpretation of the ambiguity. This pattern is

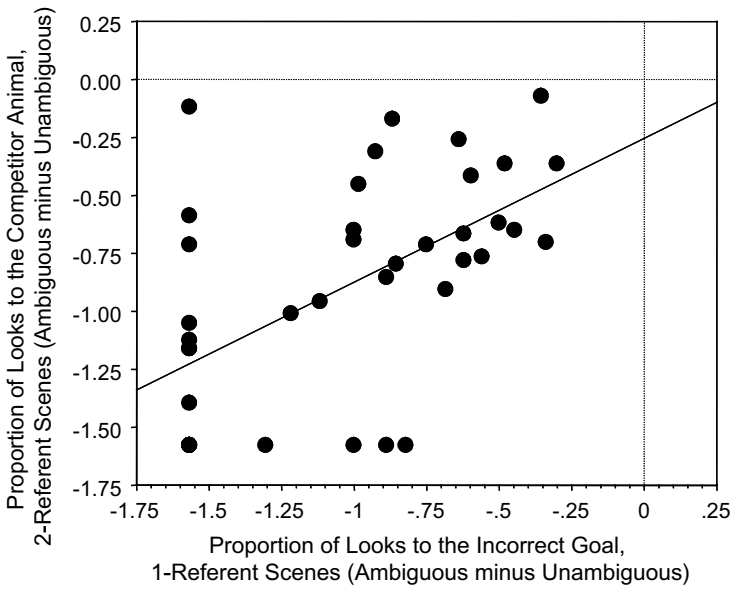


Fig. 8. Significant positive correlation ($r = .57$; $p < .0005$; $y = -.253 + .619 * x$) across conditions in the visual-world task. The dependent variable is the Ambiguity Effect from 1-Referent Scenes (Proportion of Looks to the Incorrect Goal); the independent variable is the Ambiguity Effect from 2-Referent Scenes (Proportion of Looks to the Competitor animal). Correct actions only.

not expected under constraint-based theories (such as the CBL theory) because local lexical information strongly supports a Goal interpretation and hence conflicts with the contextual evidence. In addition, prior reading studies using *put* materials (Britt, 1994) and other visual-world studies using different materials (Snedeker & Trueswell, 2004) have found evidence that these contextual constraints could not completely override lexical constraints when they conflicted – instead, both sets of constraints were weighed simultaneously.

The present study re-examined the *put* studies by introducing a time-constraint to respond to the instructions, and by introducing a spatial contrast between referents (on vs. in) which would allow for a finer-grain measure of when listeners arrived at the Modifier analysis of *the frog on the napkin.* These changes allowed us to identify clear evidence of difficulty in the 2-Referent Ambiguous conditions. Participants were slower to arrive at the intended Modifier analysis in this condition compared to Unambiguous controls. Interestingly, consideration of the Modifier interpretation in 2-Referent Ambiguous trials was inversely related to consideration of the Goal interpretation in 1-Referent Ambiguous trials, suggesting that a listener's difficulty arriving at the Modifier interpretation was related to consideration of the Goal interpretation – both lexical and contextual constraints were weighed simultaneously in all conditions of this study.

It is important to note that these findings are not a failure to replicate past findings. Error rates reported here are quite similar to past studies; they are quite similar to Trueswell et al. (1999) but also Spivey et al. (2002) and Chambers et al. (2004), once the low number of observations is taken into account. Also, the eye movement patterns on correct trials (e.g., delayed looks to the Correct Goal on 2-Referent Ambiguous as compared to 2-Referent Unambiguous trials) were noted as a suggestive numerical difference in Spivey et al. (2002). Our own findings show that these patterns of difficulty are highly reliable and replicable. Finally, our choice to have a response deadline did not seem to influence the overall speed of eye movements in response to linguistic stimuli. Our spatial contrast allowed us to test differences in interpretation within the ambiguous phrase (*on the napkin*) that were not previously possible, but it is not the case that speeded responses played a role in changing our results as compared to previous findings.

Taken as a whole, the results from this experiment and the past experiments strongly support a Constraint-Based Lexicalist account, and suggest that salient co-present referential information is unlikely to override lexical biases entirely. In what remains of this paper, we consider the extent to which individuals differ in their use of lexical evidence, and how this might account for individual variation in the present experiment.

## 3. Experiment 2: Lexical constraints and syntactic ambiguity resolution in reading

We report here the results of a reading study using the DO/SC materials, which was conducted with the same 40 participants. Section 4 will then compare performance across Experiments 1 and 2 within each individual.

### 3.1. Methods

#### 3.1.1. Procedure

Participants were seated in front of a computer screen and given written instructions. Each trial began with a fixation cross on the left side of the screen indicating where the sentence would begin. The fixation cross lasted for 500 ms before being replaced by the sentence, which remained on the screen until the participant finished reading the sentence and pressed the response button again. Each sentence was followed by a yes/no comprehension question that asked about the general content of the sentence to ensure that participants were attending to the task. Participants answered by pressing YES or NO on the button box. The experiment consisted of 10 practice trials followed by 160 experimental trials. Stimulus presentation and data recording were performed by a laptop computer using E-Prime software (Psychology Software Tools). The experiment lasted approximately 35 min.

#### 3.1.2. Stimuli

Forty test items were developed based on those in Garnsey et al. (1997). These sentences, which contained a DO/SC ambiguity, began with a subject Noun Phrase (NP) and continued with a main verb, followed by a sentence complement, as illustrated in example (5) (see Appendix B).

(5) a. *The gossipy neighbor heard the story had been told to everyone but her.* (Ambiguous)
b. *The gossipy neighbor heard that the story had been told to everyone but her.* (Unambiguous)

In (5a) the NP *the story* is temporarily ambiguous between being a direct object of the verb *heard* or the subject of a sentence complement. In (5b), the complementizer *that* has been inserted, removing this temporary ambiguity and forcing the sentence complement interpretation. Half the test items used a verb that was Direct Object (DO)-biased, such as *heard,* which, based on previous norms (Garnsey et al., 1997), is known to be a verb that permits a sentence complement as an argument but strongly prefers a direct object. The other 20 test items contained verbs that were Sentence Complement (SC)-biased, as in example (6) below. Here the verb permits a direct object but strongly prefers a sentence complement.

(6) a. *The friendly clerk indicated the price would be rising very soon.* (Ambiguous)
b. *The friendly clerk indicated that the price would be rising very soon.* (Unambiguous)

For both DO- and SC-biased items, the NP following the verb was a plausible direct object for that verb (based on thematic role ratings in Garnsey et al., 1997).

Questions on test trials were designed to test for any 'lingering' interpretation of the post-verbal noun phrase as incorrectly being the DO of the verb (see Ferreira, Christianson, & Hollingworth, 2001). In particular, all test items asked about the DO interpretation. For example, the questions for sentences 5 and 6 above were correspondingly:

(7) *Did the neighbor hear the story?*
(8) *Did the clerk indicate the price?*

Answers to target sentences were expected to be correctly answered 'no'. The extent to which subjects inferred from the sentence meaning that the answer may likely instead be 'yes' is controlled for with the Unambiguous sentences. Hence, overall 'yes' responses between DO-bias and SC-bias independent of ambiguity is not of interest; rather the effects of Ambiguity within each Verb Type is what is of concern for measuring lingering ambiguity effects. Filler trials included a variety of comprehension questions whose answers were expected to be correctly answered 'yes' or 'no' and did not test for any persistent incorrect analysis of the prior sentence.

Two presentation lists were designed containing the 40 target trials pseudorandomized among 120 filler trials. Of the target items in any given list, 20 in total were Ambiguous (10 for DO-bias verbs, 10 for SC-bias verbs) and 20 were Unambiguous (10 for DO-bias verbs, 10 for SC-bias verbs) resulting in four total experimental conditions. In order to increase the number of items in the experiment, five SC-bias verbs (*claimed, concluded, assumed, suggested, worried*) and two DO-bias verbs (*maintained, established*) appeared twice in each list. They were treated as separate test items, however, because (a) they never appeared in the same condition within a list (Ambiguous, Unambiguous); and (b) they always appeared in a different sentence within a list. Each target item was rotated across lists through the two possible Ambiguity conditions for each Verb Type. Each participant was assigned to one of the two lists. Lists were run in both forward and reverse trial orders, alternating between participants. Filler sentences were similar to target items but were of different constructions intended to detract from the salience of experimental stimuli and to reduce the possibility that participants would develop expectation-driven strategies.

### 3.2. Results

#### 3.2.1. Reading times to target sentences

To minimize the impact of string length variability, reading times were adjusted for the number of characters in a sentence. First, raw reading times beyond 3SD above or below a participant's mean reading time across all conditions were replaced with the 3SD cutoff value, affecting 1.6% of the data. Then, for each participant, a regression equation was derived that predicted reading times from total sentence length (see Ferreira & Clifton, 1986; Trueswell et al., 1994). The statistical analyses reported here were performed on these trimmed residual times, though for the sake of clarity these analyses are graphed in terms of trimmed raw (i.e., whole-sentence, not residual) reading times.

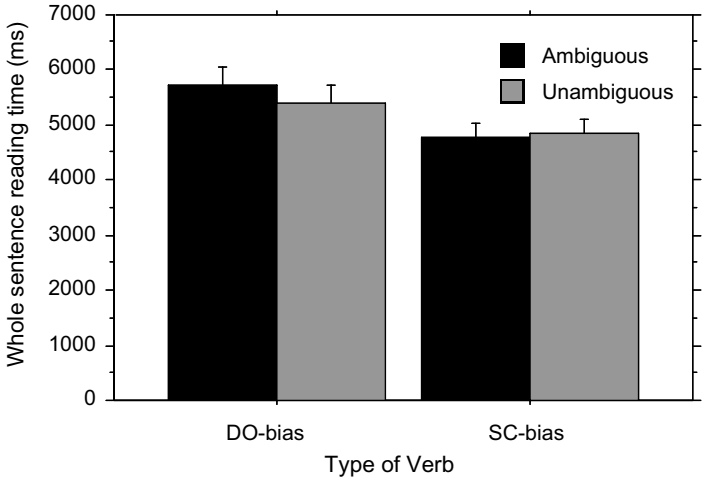


Fig. 9. Reading time (ms) by condition. Main effects of Verb Type and Ambiguity were observed ($ps < .01$) as well as a reliable interaction between them ($p < .05$). Error bars reflect the 95% confidence interval.

The reading time pattern replicated previous findings using these materials (e.g., Garnsey et al., 1997; Holmes et al., 1989). In particular, longer reading times were observed overall for DO-bias verbs when they appeared in SC contexts (377 ms vs. −377 ms; raw RTs: 5567 ms vs. 4815 ms for SC-bias verbs) [$F1(1,38) = 35.5$, $p < .01$]. A main effect of Ambiguity was also replicated: Ambiguous sentences took longer to read than Unambiguous sentences (165 ms vs. 165 ms; raw RTs: 5244 ms vs. 5137 ms) [$F1(1,38) = 10.5$, $p < .01$; $F2(1,36) = 5.3$, $p < .05$]. A reliable Verb Type by Ambiguity interaction also emerged, again replicating previous reports using these stimuli [$F1(1,38) = 5.4$, $p < .05$; $F2(1,36) = 4.5$, $p < .05$]. Fig. 9 illustrates these effects by condition.

### 3.2.2. Accuracy and response times to comprehension questions

Averaging across all four conditions, readers' accuracy on the comprehension questions was 86%. Participants were reliably more accurate on sentences containing SC-bias verbs (88% vs. 84% for DO-bias verbs) [$F1(1,37) = 4.66$, $p < .05$]. No effect of Ambiguity was found (87% for Ambiguous; 86% for Unambiguous) [$F1(1,37) = .14$, $p = .7$; $F2(1,36) = 1.92$, $p < .2$], but a marginal Verb Type by Ambiguity interaction was observed by subject analysis only [$F1(1,37) = 3.3$, $p = .07$; $F2(1,36) = .25$, $p = .6$]. This interaction appears to be the result of the relatively large accuracy difference between SC-Ambiguous and -Unambiguous conditions (90% vs. 86%, respectively) in a direction that was opposite to that observed for DO-bias verbs (84% accurate for DO-Ambiguous vs. 85% accurate for DO-Unambiguous conditions). Residual reading times to correctly answered comprehension questions did not reveal reliable differences in Ambiguity or Verb Type by either Subject or Item analyses (all $Fs < 2$; all $ps > .1$).

### 3.3. Discussion of Experiment 2

Consistent with prior reading studies, participants experienced the greatest Ambiguity effect measured by reading times for sentences containing DO-bias verbs: readers as a whole pursued the interpretation predicted by highly regular lexical preferences. That is, for these materials, readers' processing of a temporary ambiguity depends on the structural preferences of the verb. By contrast, Unambiguous sentences were read reliably faster. Also as expected, reading times for sentences containing SC-bias verbs within SC constructions were read significantly faster, suggesting comparatively little processing difficulty. This is because the syntactic environment in which the verb appeared was consistent with its reliability-based patterns.

## 4. Using reading time measures to predict behavior in the visual-world *put* task

If there exists a range in the normal ability to resolve temporary syntactic ambiguities, and if these abilities are reflected in an individual's consistent sensitivity to lexical cues, then we should expect the sentence processing measures from Experiments 1 and 2 to show correspondences concerning the use of lexical information – independent of the modality of the linguistic input (speech or text) or the type of syntactic ambiguity that a reader or a listener encounters (PP-attachment or DO/SC ambiguity). We have, for instance, assumed thus far that the difficulty that individuals have at arriving at the correct Modifier interpretation of the ambiguous phrase in the *put* task arises from the presence of countervailing lexical evidence (*put* wants a Goal). If this is the case, then we should expect that the individuals who have particular difficulty in the *put* task are, counter-intuitively, the same individuals who are extremely good at using verb evidence to guide parsing choices in the reading task (Experiment 2): "lexically sensitive" individuals are expected to get into trouble when the verb information leads them astray. However, as we discussed earlier, one might expect that individuals who are highly tuned to fine-grain lexical–syntactic contingencies to be more sensitive to lexical–contextual contingencies as well. For instance, within the visual-world study, such individuals should be more sensitive to the lexico-semantic differences between *on* and *in* under 2-Referent conditions, resulting in a greater Target Advantage in the Ambiguous cases compared to those who are not so sensitive to lexical cues. Crucially though, we would not expect such lexically adept individuals to be able to eliminate all difficulty associated with 2-Referent Ambiguous materials as compared to 2-Referent Unambiguous materials: their knowledge of the syntactic preferences of *put* should subtly interfere with their ability to converge on the intended Modifier interpretation.

To test these hypotheses, we used our reading time study as a diagnostic tool to identify lexically sensitive individuals, i.e., those individuals who use verb information to guide parsing commitments. This was done by calculating the Verb Type x Ambiguity interaction term from each individual's mean residual reading times (i.e., the difference between an individual's Ambiguity effect for DO-bias and SC-bias

Verb Types ((DO-Ambiguous RT minus DO-Unambiguous RT) minus (SC-Ambiguous RT minus SC-Unambiguous RT))). Individuals who showed a positive interaction term and a positive Ambiguity effect for DO-bias items alone are henceforth labeled as having *High Lexical Sensitivity* (HLS; $N = 25$), because positive values here mean that such readers were using Verb Type strongly to drive parsing choices, which increased processing difficulty in DO-Ambiguous conditions only. By contrast, all other individuals were labeled as having *Low Lexical Sensitivity* (LLS; $N = 15$) (i.e., individuals not showing a great deal of lexical sensitivity) because their pattern did not yield signs of consistently using verb information. Fig. 10 plots the reading time measure from Experiment 2 separately for HLS and LLS individuals. As can be seen, HLS individuals show a cost for DO-Ambiguous and a slight benefit for SC-Ambiguous. In contrast, LLS individuals show no interpretable pattern.

The predictions are fairly straightforward. First, HLS individuals as compared to LLS individuals should show exaggerated garden-pathing effects in the *put* visual-world study, especially in the 1-Referent contexts. In particular, HLS individuals as compared to LLS individuals should show increased errors on 1-Referent Ambiguous trials and increased time spent looking at the Incorrect Goal in this condition. (The 2-Referent condition is not expected to show interpretable Incorrect Goal looking time effects, for the reasons discussed above.) Second, HLS individuals as compared to LLS individuals should show better lexical sensitivity to the *in*/*on* distinction in 2-Referent contexts, both in Unambiguous and Ambiguous materials. But third, HLS individuals should not be able to completely eliminate the difficulty associated with 2-Referent Ambiguous trials; 2-Referent Unambiguous materials should still show an earlier and more robust Target Advantage as compared to 2-Referent Ambiguous materials. Also, the 2-Referent Unambiguous materials should continue to exhibit an advantage when it comes to looking at the Correct Goal.

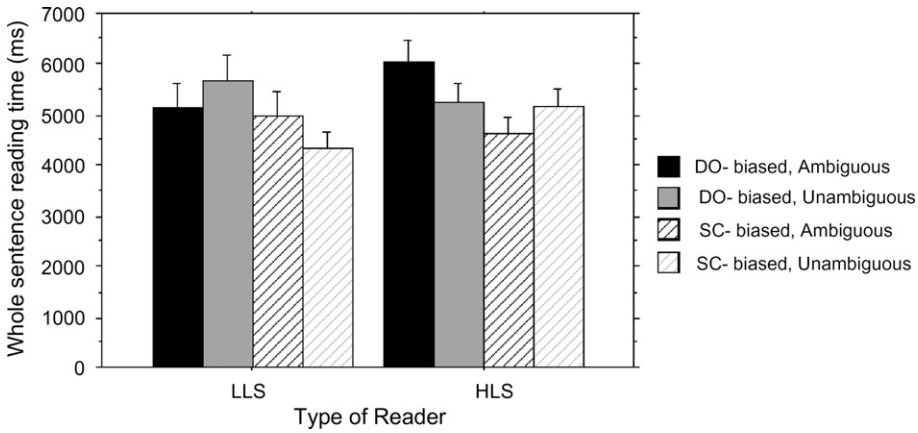


Fig. 10. Reading time (ms) by condition split by Type of Reader. LLS, Low Lexical Sensitivity; HLS, High Lexical Sensitivity.

### 4.1. HLS individuals vs. LLS individuals: Errors in hand actions

HLS individuals should show greater garden-pathing than LLS individuals, especially in 1-Referent contexts. One way to test this prediction is by comparing action errors across the groups. Fig. 11 shows this comparison for the proportion of trials on which participants moved an animal to the Incorrect Goal. As can be seen in the figure, HLS individuals made about three times as many errors of this type as compared to the LLS group (21% vs. 7%). An ANOVA on Subject means was performed using three factors: Subject Type (HLS vs. LLS), Referential Scene, and Ambiguity. This resulted in a marginal Ambiguity by Subject Type interaction [$F(1,38) = 3.39$; $p = .07$] (see Fig. 11). This suggests that HLS individuals are simultaneously relying on the verb information and the contextual information, both of which lead them astray and, in many cases, unable to recover from the temporary ambiguity. It should be noted that this difference is not the result of just a few outlying HLS participants: about half of HLS individuals (12/25) made at least one error in the 1-Referent Ambiguous condition (five of which made errors on at least half of these trials), whereas a quarter (4/15) of the LLS participants made errors in this condition; notably, these four participants made only one error each.

Interestingly however, under 2-Referent Ambiguous conditions, when the visual context supported the Modifier analysis (two frogs) and battled against the lexical constraint associated with *put*, the error rates between HLS individuals and LLS individuals are nearly identical (8% vs. 10%, respectively [$F(1,38) = .21$; $p > .6$]. This suggests that both HLS and LLS groups take relevant referential evidence into account when it comes to their ultimate actions.

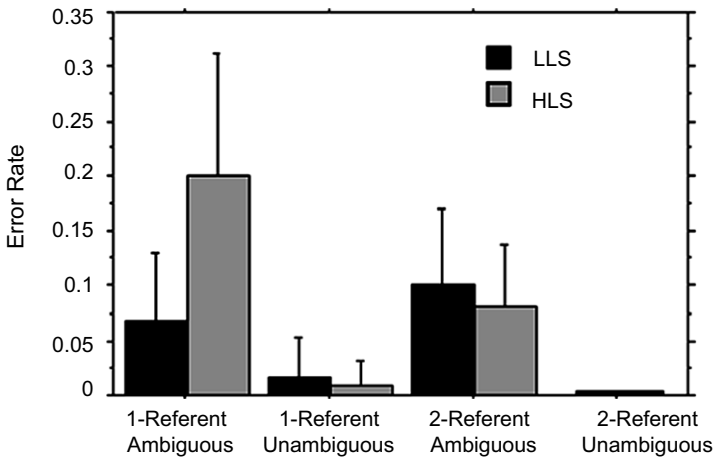


Fig. 11. Proportion of errors involving the Incorrect Goal (Experiment 1) in each condition split by Type of Reader. LLS, Low Lexical Sensitivity; HLS, High Lexical Sensitivity.

### 4.2. HLS individuals vs. LLS individuals: Looks to the Incorrect Goal

If our hypotheses sketched above are correct, then we should expect HLS individuals also to be those who show greater consideration of the Goal interpretation of *on the napkin* in the *put* task within the Ambiguous trials. Within 1-Referent trials, visual inspection of the Incorrect Goal should be early and quite large for HLS individuals. Within 2-Referent trials, a similar effect should in principle be expected,

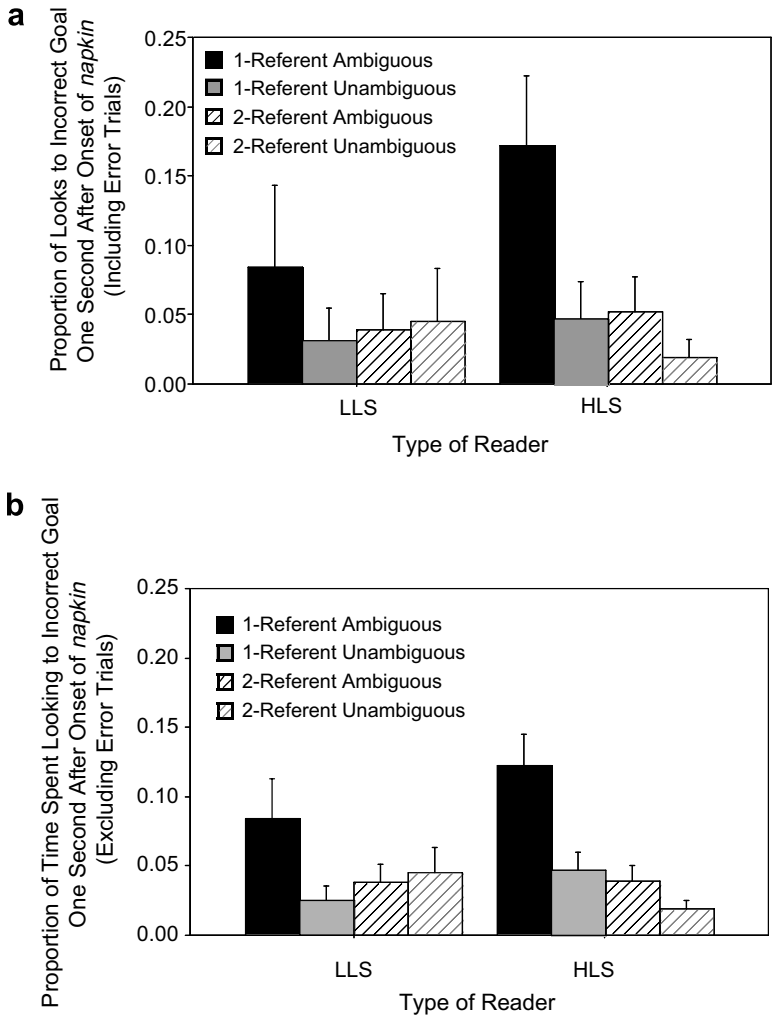


Fig. 12. (a) Proportion of time spent looking to the Incorrect Goal during the first second after the onset of *napkin*, split by Type of Reader, including all trials (correct and incorrect actions). (b) Proportion of time spent looking to the Incorrect Goal during the first second after the onset of *napkin*, split by Type of Reader, correct action trials only. LLS, Low Lexical Sensitivity; HLS, High Lexical Sensitivity.

though, as discussed above, Incorrect Goal looks are difficult to interpret within this context. As can be seen in Fig. 12, HLS individuals do show a greater proportion of time spent looking at the Incorrect Goal for the first second after hearing the onset of *napkin.* (Panel (a) presents this for all trials including errors, whereas (b) presents the data excluding error trials.) For all trials, an ANOVA on arcsin-transformed proportions was run on subject means entering Ambiguity, Context, and Subject Type (i.e., Type of Reader; HLS vs. LLS) as factors, and revealed that Subject Type interacted significantly with Ambiguity $[F(1,38) = 7.38;\ p < .01]$ and with Context

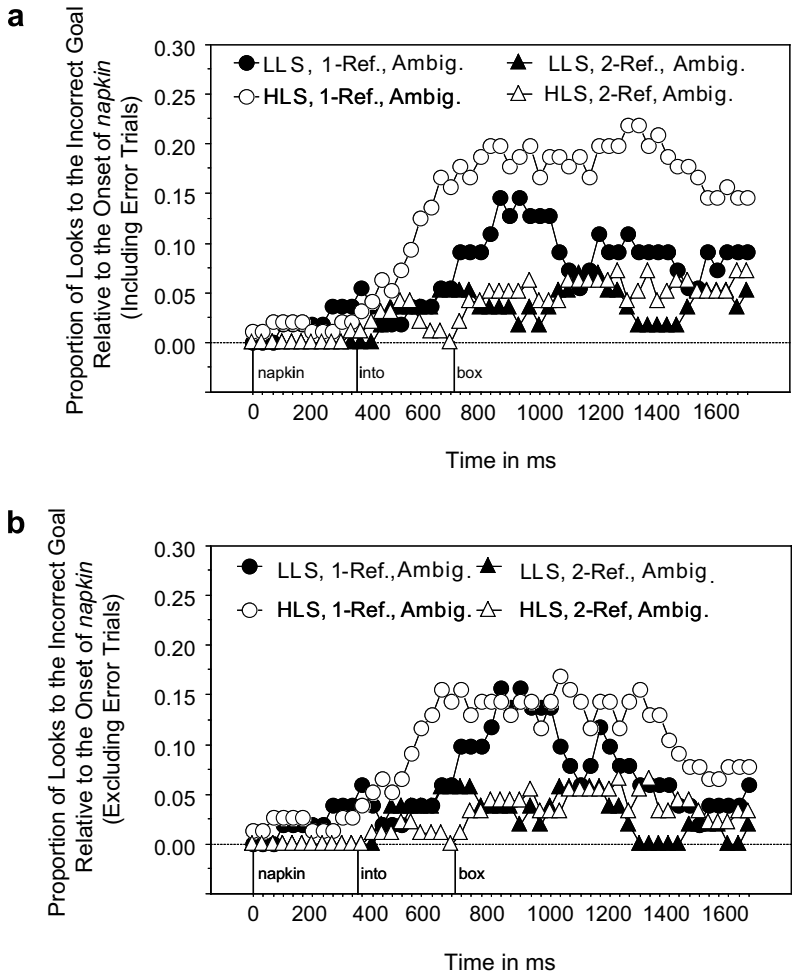


Fig. 13. (a) Proportion of looks to the Incorrect Goal over time as the Ambiguous phrase unfolds (relative to the onset of *napkin*), split by Type of Reader, including all trials (correct and incorrect actions). (b) Proportion of looks to the Incorrect Goal over time (relative to the onset of *napkin*), split by Type of Reader, correct action trials only. LLS, Low Lexical Sensitivity; HLS, High Lexical Sensitivity.

[$F(1,38) = 7.01$; $p = .01$]. Interestingly, however, no three-way interaction was observed between Subject Type, Ambiguity, and Context [$F(1,38) = 2.79$; $p > .1$], suggesting that HLS individuals show a larger consideration of the Incorrect Goal compared to LLS individuals regardless of context, i.e., in both 1- and 2-Referent Scenes.[8] These interactions with Subject Type however were not significant when we examined correct trials only ($F$s $< 2$).

The timing of consideration of the Incorrect Goal is illustrated in Fig. 13 for both HLS readers and the LLS group (Panel (a) with error trials and (b) excluding error trials). As can be seen in the figure, HLS individuals show earlier and greater temptation to look at the Incorrect Goal in 1-Referent Scenes.

### 4.3. HLS Individuals vs. LLS Individuals: Target Advantage in 2-Referent contexts

As mentioned in the introduction to this paper, it is also quite plausible to expect that individuals who are adept at weighing lexical factors may show a greater Target Advantage in 2-Referent Ambiguous cases because of enhanced sensitivity to the in/on distinction, thereby arriving at the Modifier interpretation with greater facility. If this is observed, our diagnostic test has certainly identified more than just those who rely on verb-syntactic information per se, but rather those who are especially adept at using lexical constraints more generally, showing, for instance, real-time sensitivity to how the semantic restrictions of *on* vs. *in* interact with the spatial properties of objects in the environment.

Indeed, this was confirmed when plotting (in Fig. 14) each group's Target Advantage effect for Ambiguous and Unambiguous conditions under 2-Referent contexts, much like what was done in Section 2 for all participants combined (Target looks minus Competitor looks). (Panel (b) presents the data with error trials excluded; very little differences can be seen between the Panels because so few errors occur in 2-Referent Scenes.) As can be seen in the figure, the HLS individuals show faster convergence on the Target animal in the Ambiguous condition compared to LLS individuals in the Ambiguous condition. This means that HLS individuals tend to arrive at the correct referent more quickly than LLS individuals when the phrase is ambiguous. As expected, however, both groups of individuals show an even greater Target Advantage for Unambiguous materials. That is, HLS individuals and LLS individuals show some difficulty arriving at the Modifier interpretation in the Ambiguous condition. Statistical tests (using all trials) confirmed this interpretation. In particular, within Unambiguous materials, the Target Advantage becomes significantly different from zero between 0 and 200 ms after *on* for HLS individuals [$t(24) = 4.5$; $p < .01$] and between 400 and 600 ms for LLS individuals [$t(14) = 14.83$; $p < .01$]. Within Ambiguous materials, the Target Advantage became significantly different from zero between 200 and 400 ms for HLS individuals [$t(24) = 2.29$; $p < .05$] and between 600 and 800 ms for LLS individuals [$t(14) = 2.77$; $p < .01$].

[8] A correlational analysis including all 40 participants revealed a similar statistical pattern: As an individual's interaction term on the reading study (as defined by the criteria above) increased, the time that he or she spent looking at the Incorrect Goal increased as well ($r = .32$; $p < .05$; $y = .23 + 7.98\text{E}{-5} * x$).

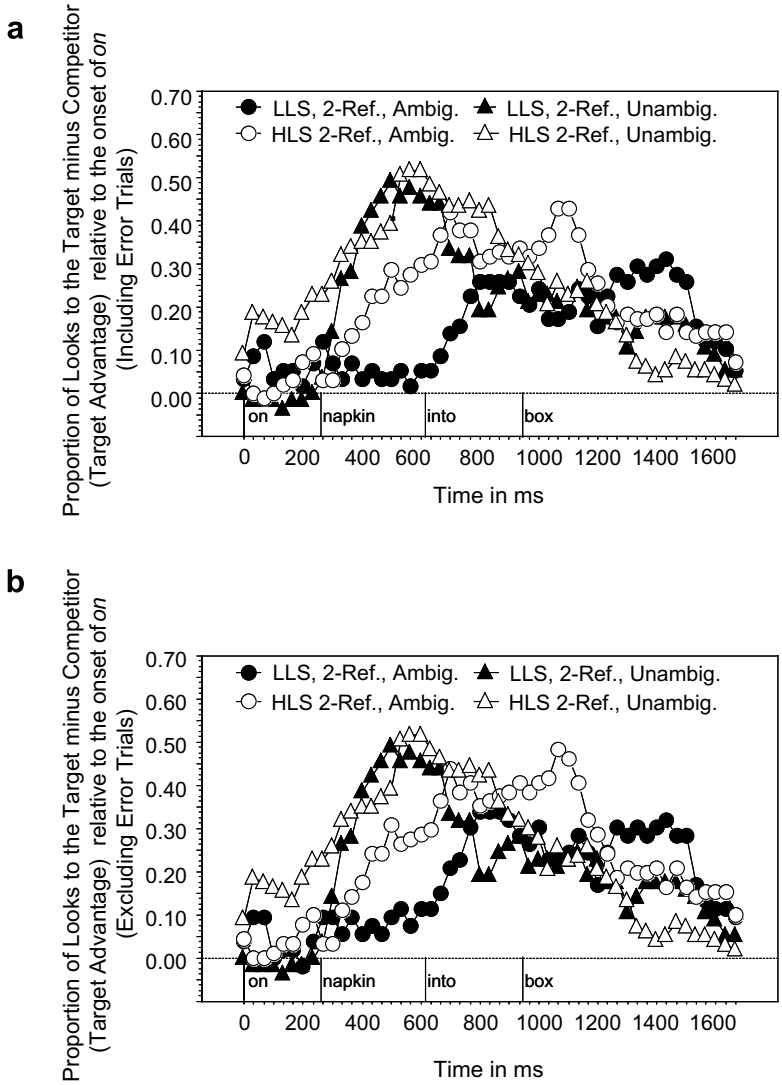


Fig. 14. (a) Probability of fixating the Target animal (e.g., the frog on the napkin) on all trials (i.e., correct and incorrect actions) instead of the Competitor animal (e.g., the frog in the basket) over time for 2-Referent Ambiguous and Unambiguous conditions, split by Type of Reader. 0 ms on the *x*-axis marks the onset of the preposition *on* in the temporarily ambiguous phrase *on the napkin.* (b) Correct trials only. LLS, Low Lexical Sensitivity; HLS, High Lexical Sensitivity.

Here, as expected, when we exclude error trials, a similar pattern holds. In particular, within Unambiguous materials, the Target Advantage becomes significantly different from zero between 0 and 200 ms after *on* for HLS individuals [$t(24) = 2.28$; $p < .05$] and between 600 and 800 ms for LLS individuals

[$t(12) = 4.38$; $p < .01$]. Within Ambiguous materials, the Target Advantage becomes significantly different from zero between 600 and 800 ms for HLS individuals [$t(24) = 4.37$; $p < .01$] and between 800 and 1,000 ms for LLS individuals [$t(14) = 2.93$; $p < .01$].

This greater (and earlier) Target preference for HLS participants under 2-Referent Ambiguous conditions compared to LLS participants suggests that people who readily use lexical constraints to drive their parsing decisions are also highly-tuned to lexico-contextual contingencies as well; they show less difficulty arriving at the Modifier interpretation of the ambiguous PP than LLS participants when the visual context supports the Modifier analysis. Thus, when more constraints are available to inform parsing choices, HLS individuals are in fact multiple-constraint-reliant and therefore more rapidly converge on the intended referent as compared to LLS individuals. However, as expected, contextual support does not completely eliminate processing difficulty associated with Ambiguous materials as compared to Unambiguous materials, even for HLS individuals.

### 4.4. Looks to the Correct Goal

If contextual support facilitates HLS individuals' processing of Ambiguous trials, but does not completely eliminate difficulty with these items as compared to Unambiguous controls, we would expect to see that the advantage to look at the Correct Goal (the box) more quickly on Unambiguous trials as compared to Ambiguous trials should continue to be present for both HLS and LLS groups. Fig. 15 re-plots this advantage in 2-Referent Scenes, split by Subject Type. As can be seen in the figure (which includes only correct action trials), HLS individuals are speeded as compared to LLS individuals, but both continue to show an advantage for Unambiguous sentences.

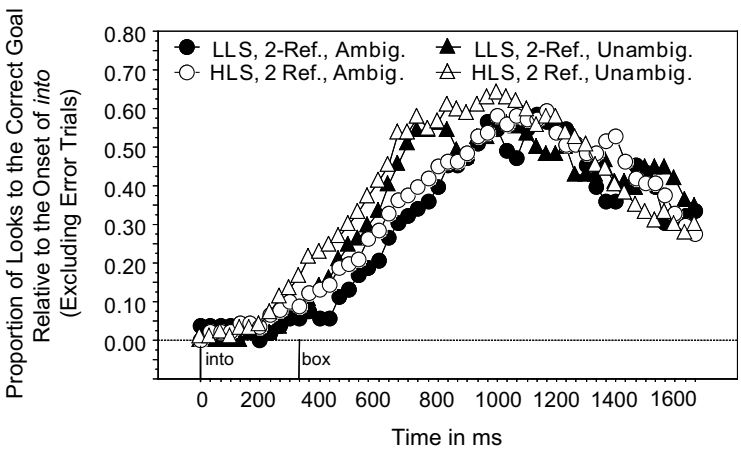


Fig. 15. Proportion of looks to the Correct Goal over time in 2-Referent Scenes, relative to the onset of the second PP (e.g., *into*; marked by 0 ms on the *x*-axis), split by Type of Reader. LLS, Low Lexical Sensitivity; HLS, High Lexical Sensitivity.

During a 1-s window after the onset of the second PP (200–1200 ms after the onset of, e.g., *into*), the proportion of looks to the Correct Goal were computed. Subject means for 2-Referent trials were entered into an ANOVA with two factors: Subject Type and Ambiguity. This revealed a main effect of Ambiguity ($F1(1,38) = 10.67$, $p < .01$), no effect of Subject Type ($F(1,38) = 1.36$), and no interaction ($F(1,38) = 0.01$). Thus, by this test, HLS participants were not reliably looking at the Correct Goal more than LLS participants. Crucially though, there is an Ambiguity effect such that Unambiguous sentences are eliciting more looks to the Correct Goal in this time window than Ambiguous sentences. Thus, both HLS and LLS participants are having some difficulty with Ambiguous items.

### 4.5. Reading span and other properties of HLS individuals

One reasonable interpretation to assign to these individual differences is that our diagnostic tool has not captured one's ability to use lexical and other available constraints per se, but rather more general abilities reflecting differences in working memory capacity that relate to parsing abilities. However, this was not the case. Each of our participants also completed the reading span task (Daneman & Carpenter, 1980), and no differences in reading span were found between HLS individuals and LLS individuals [$t(37) = .58$; $p > .50$] (Fig. 16a). This suggests that working memory capacity and/or reading experience alone does not account for individual

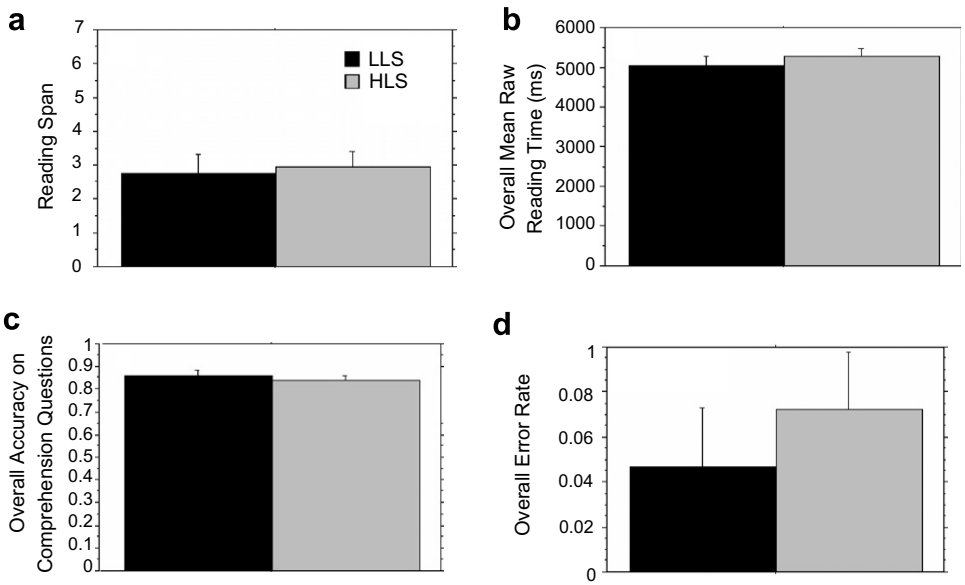


Fig. 16. (a) Reading span performance by Type of Reader (ns); (b) Overall reading time by Type of Reader (ns); (c) Overall comprehension accuracy from Experiment 2 by Type of Reader (ns); (d) Overall proportion of errors involving the Incorrect Goal, by Type of Reader (ns). LLS, Low Lexical Sensitivity; HLS, High Lexical Sensitivity.

variation in the use of lexico-syntactic and visuo-contextual constraints to inform parsing choices. In other words, HLS individuals are not (merely) high-span people.

It is also the case that overall reading time, regardless of condition, does not predict whether an individual is HLS (and, by extension, multiple-constraint-reliant). Averaging across all four conditions of Experiment 2, no differences in raw reading times were observed between the two groups [$t(1598) = 1.56$; $p > .10$] (Fig. 16b). Further, no significant difference was found between HLS and LLS participants concerning their overall accuracy on the comprehension questions [$t(1598) = 1.23$; $p > .20$] (Fig. 16c). Last, as discussed earlier, HLS participants showed a slight, albeit non-significant tendency, to make more action errors involving the Incorrect Goal in Experiment 1 (Fig. 16d). Thus, both groups make approximately the same number of incorrect actions, but as shown earlier in Fig. 11, their distribution of errors across conditions did differ; HLS individuals showed more errors in the 1-Referent Ambiguous condition.

Taken together, the most plausible and parsimonious interpretation is that our diagnostic tool has selected in particular those participants who rely heavily on lexical information to drive parsing commitments, and critically did not select by overall reading, working memory, or comprehension ability as such.

### *4.6. Correlations in verb use across Experiments 1 and 2*

We have thus far talked about lexical sensitivity (and multiple constraint sensitivity) as being a categorical trait of individuals. Obviously though, this sensitivity should fall along a continuum. That is, correlated variation should also exist across subject means in our measure of processing difficulty from the reading task and fixation times to certain objects in the scene from the visual-world task: signs of processing difficulty (elevated reading times) in the DO-bias materials of Experiment 2 should positively correlate with signs of listeners committing to a Goal analysis in the 1-Referent Scene materials of Experiment 1 (proportion of time spent fixating an Incorrect Goal). A comparison of reading time measures with the eye movement measures in the *put*-study also allows us to cross-validate certain linking assumptions associated with each task. In particular, effect sizes from both studies should correlate with each other on a subject-by-subject basis, even though the class of measures is completely different, the structures that are tested are different, and the modality of presentation is different. Increased reading times for Ambiguous compared to Unambiguous sentences are assumed to reflect (in large part) processing difficulty associated with erroneously committing to a Direct Object analysis and the difficulty associated with recovering the correct parse of the temporarily ambiguous sentence. Likewise, increased looks to the Incorrect Goal in the Ambiguous compared to the Unambiguous versions of the *put*-study materials are assumed to reflect erroneous commitment to the Goal interpretation of *on the napkin,* and hence a measure of the degree of reliance on lexical cues in the visual-world study.

To test this prediction, difference scores were computed from individual subject means (as was done for previously reported correlations) as a measure of the magnitude of difficulty or commitments to a particular interpretation in the reading and visual-world experiments, respectively. Magnitude of difficulty in the reading study was com-

puted by subtracting an individual's mean reading time for Unambiguous sentences from his or her reading times for Ambiguous sentences for the DO-bias verb type. This difference score was derived from residual whole-sentence reading times to partial out variance related to string length. For the listening study, the Ambiguity Effect measure was calculated from the eye movement data: the difference between the proportion of time spent looking at the Incorrect Goal (e.g., empty napkin) in Ambiguous and Unambiguous conditions for 1-Referent Scene types (from *napkin* onset), during the 200- to 1200-ms after the onset of *napkin*. Only the 1-Referent conditions were used here because this is where effects of Ambiguity were seen (again because context mitigates the effect when two referents are present). The magnitude of these effects (Ambiguous minus Unambiguous) was computed for the arcsin transformed data.

As predicted, participants' processing measures across reading and listening studies correlated positively with each other (Fig. 17). In particular, when including error trials (Panel a), increased looks to the Incorrect Goal (the empty napkin) in 1-Referent Scenes (Ambiguous minus Unambiguous) positively correlated with elevated reading times for sentences containing DO-biased items (Ambiguous minus Unambiguous) ($r = 0.52$; $p < .001$).[9] When excluding error trials (Panel b), a similar correlation holds ($r = 0.38$, $p < .05$).

Thus, measures that on the surface might seem to be unrelated, i.e., looks to a napkin during listening vs. button press times during reading, are actually highly correlated because in this case the measures were designed to tap an individual's ability to override initial misinterpretation of a temporary ambiguity that was driven by lexical constraints on structure. It is important to note that this effect was driven specifically by the ambiguity present in each of the studies: the Ambiguous conditions themselves are correlated across modality (with error trials: $r = 0.41$, $p < .05$; without errors trials: $r = 0.36$; $p < .05$), whereas the Unambiguous conditions are not (with error trials: $r = .003$; $p > .7$; without error trials: $r = .06$, $p > .7$).[10] Thus, individuals who are using verb-bias information to incorrectly pursue the DO interpretation in the reading study are also using verb-bias information to incorrectly pursue the Goal analysis in the visual-world study.

### 4.7. Discussion

We used the reading task of Experiment 2 as a diagnostic tool to determine whether some participants used lexical constraints to a greater degree than others to drive their

[9] It should be noted that all reported significant correlations in this paper are not merely the result of outliers. For example, the reader might be concerned that in Fig. 17, the participant with a residual RT in the reading study of 3849 ms (upper right-hand corner of the graph) is an outlier and thus driving the correlation. However, removing this participant from the analysis also yielded a significant effect ($r = .4$; $p < .01$). In addition, removing the participant with a residual RT of −490 ms and a −.5 proportion of looks to the Incorrect Goal (arcsin transformed) in the listening study (the lowest data point on the *y*-axis) still resulted in a significant correlation ($r = .42$; $p < .01$).

[10] The difference between these correlations is also significant when all trials (i.e., correct and error trials) are included: $(r_{\text{Ambig}} - r_{\text{Unambig}}) = .407$; $p < .05$. Without error trials, the difference between the correlations is marginally significant: $(r_{\text{Ambig}} - r_{\text{Unambig}}) = .3$; $p < .09$.

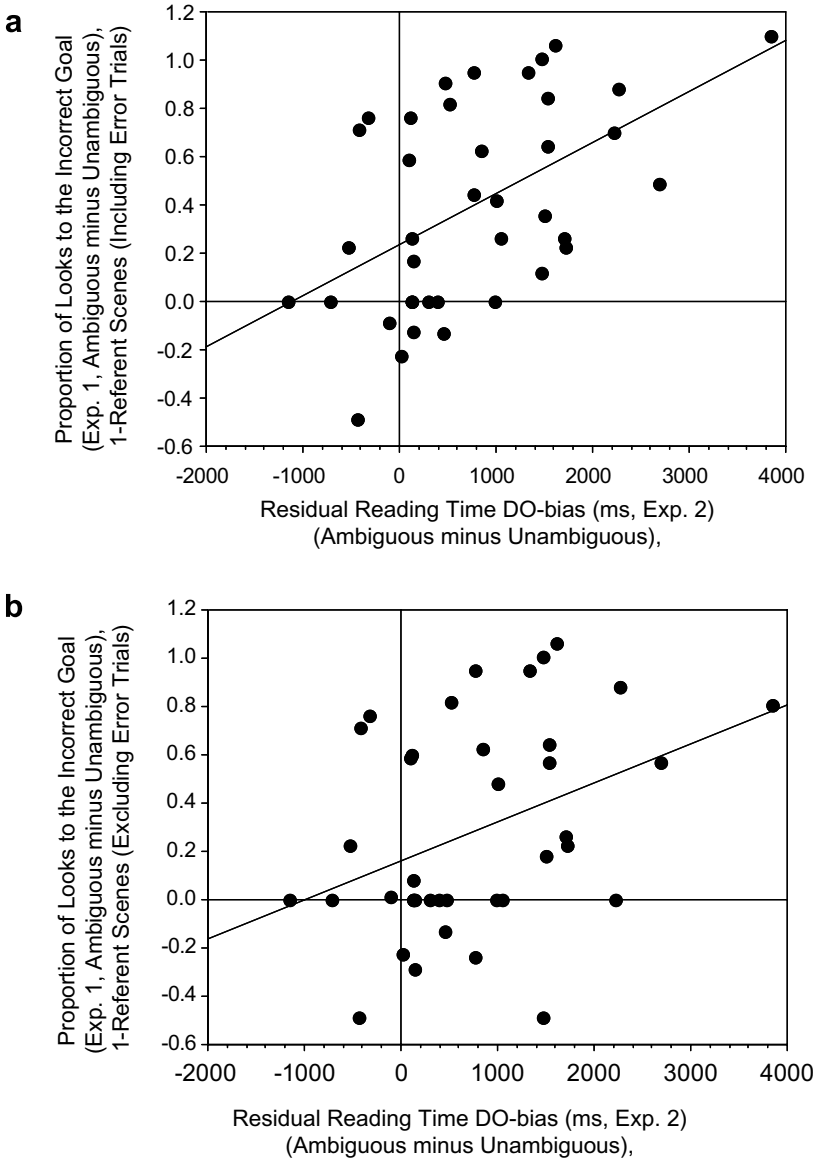


Fig. 17. (a) Significant positive correlation (errors included; $r = .52$; $p < .001$; $y = .236 + 2.11\text{E}{-4} * x$) between the Ambiguity Effect from the visual-world task in Experiment 1 (dependent variable; Proportion of Looks to the Incorrect Goal in 1-Referent Scenes, onset *napkin* to 1 s thereafter) and the Ambiguity Effect in the Reading task in Experiment 2 (independent variable; residual Reading Times in ms). (b) Significant positive correlation (errors excluded); $r = .38$; $p < .05$; $y = .163 + 1.61\text{E}{-4} * x$) between the Ambiguity Effect in Experiment 1 (dependent variable) and the Ambiguity Effect in Experiment 2 (independent variable).

parsing decisions. Doing this enabled us to determine that characterizing people as Highly Lexically Sensitive (HLS) in the reading modality also predicted the extent to which these same individuals relied on lexical constraints to inform their parsing decisions while comprehending spoken language, regardless of visual context. This strongly suggests that the difficulty we observed in Experiment 1 for listeners arriving at the Modifier analysis of *on the napkin...*, even in the presence of a 2-frog context, was due primarily to the lexical constraints of *put* having to battle against the contextual constraints of the visual scene. Moreover, we observed that HLS individuals were not only more apt to use lexical constraints, but also more apt to use all available evidential constraints, including non-linguistic evidence such as visual context, to help them structure the input in real-time in order to arrive at the intended interpretation. On the basis of multiple measures, these individuals do not appear to be faster readers or more accurate comprehenders, nor do they appear to have an advantage due to larger working memory capacities as portrayed by reading span performance. Rather, these individuals simply appear to be very good at (and very reliant on) coordinating multiple available lexico-syntactic and lexico-contextual contingencies to inform their parsing and interpretation choices.

Furthermore, constraint-use does not appear to be a categorical trait of individuals. Correlated variation was found to exist across subject means in our measure of processing difficulty from the reading task and fixation times to certain objects in the scene from the visual-world task, suggesting that comprehenders' use of constraints, regardless of modality, falls somewhere along a continuum of good- and bad-constraint use. This pattern of findings is important because it suggests that individuals do not differ categorically, but rather along a continuum expected from a constraint-based system.

In addition, these findings cross-validate the measures used in reading and visual-world paradigms that assess parsing processes in quite different ways. One possible outcome was that no reliable correspondence would be found between the traditional measure of processing difficulty in the reading task (elevated reading times) and the measure of processing commitments in the visual-world *put* studies (increased looks to an Incorrect Goal). Because elevated reading times evaluate comprehension difficulty with ambiguous sentences, it could have been that increased looks to a certain (predicted) object in the environment in the visual-world studies simply do not reflect processing difficulty in the same way that increased reading times do, and they therefore do not capture measurable ongoing linguistic processes that can be interpreted in any important or revealing way. However, a strong correlation between these measures was found, despite differences in task (pressing buttons vs. moving toys around a table), the type of ambiguous construction, and, of course, modality. This suggests that the behavioral measures used in the visual-world paradigm to assess rapid processing decisions during spoken ambiguities are both relevant and highly comparable to those used in reading techniques. Furthermore, the correspondence also suggests the reality of a domain-general lexicalist parsing mechanism that is responsible at least for guiding syntactic and semantic comprehension processes, but also perhaps for the recovery from temporary ambiguity. Although the standard assumption in psycholinguistics has been that there exists a

single parser that extends its processing capabilities across reading and auditory domains, the correlation effect found here confirmed this assumption empirically.

## 5. General discussion

### 5.1. Revisiting lexical and contextual effects in the put task

In the visual-world listening task (Experiment 1), we observed that 2-Referent Scenes did not completely eliminate uncertainty associated with the syntactic ambiguity despite contextual support for the Modifier analysis. Even though two referents (e.g., two toy frogs) were visually present, listeners showed difficulty settling on the intended Modifier analysis of *on the napkin* in the Ambiguous structure as compared to the Unambiguous control. We were able to observe this difference because our experimental setup contrasted Target and Competitor animals spatially (one was on a flat surface whereas the other was in a container). Typically, a listener hearing a preposition like *on* in such situations can use the semantics of the preposition to guide eye movements to the appropriate referent and exclude inappropriate referents (Chambers et al., 2004). Indeed, we observed such anticipatory eye movement patterns in our syntactically unambiguous materials (*Put the frog that's on the*...). Individuals rapidly and reliably converged on the spatially appropriate referent (i.e., the frog that was on a flat-surfaced object) prior to the point in time when reference to that flat object was phonologically realized (*napkin*).

This same anticipatory pattern was not observed, however, for the syntactically Ambiguous version of these materials (*Put the frog on the*...). In this condition, participants needed to hear *napkin,* and perhaps even the second disambiguating PP *into the box,* before reliably converging on the Target referent (the frog on the napkin, not the frog in the bowl). This pattern indicates that there was some uncertainty in the mind of each listener regarding the appropriate structural analysis of *on the napkin,* namely as a restrictive NP Modifier, even though the visual context supported that analysis; listeners did not initially realize that *on the napkin* was providing information about which frog was being referred to. This uncertainty arose because the verb *put* supported an alternative analysis of *on the napkin* as a Goal.

This pattern strongly indicates that 2-Referent contextual evidence was not able to completely block listeners' interpretation of *on the napkin* as the Goal of *put.* Rather, this contextual evidence showed signs of having to battle against the highly reliable bottom-up lexical evidence. This result is consistent with the informativeness and availability of the lexical cues: *put* almost always appears with a Goal phrase throughout the language, whereas speakers can modify NP referents in any number of different ways (Trueswell & Gleitman, 2004; Trueswell et al., in press).

This pattern also provides new consistency with reading studies that have also shown signs of brief processing difficulty under these same 2-Referent conditions when using countervailing verb information (Britt, 1994). It is also consistent with the recent findings of Snedeker & Trueswell (2004), who showed that 2-Referent Scenes were not sufficient to guide listeners toward a Modifier interpretation of globally ambiguous

sentences (*Feel the frog with the feather*). In that study, when the verb strongly supported an Instrument interpretation, Instrument actions were carried out over 68% of the time by subjects, even though there were two referents present (two frogs, one of which held a feather, thus supporting a Modifier analysis). The percentage of Instrument actions in the 2-Referent cases dropped to essentially zero, however, when the verb was changed to be Equi- or Modifier-biased (e.g., *choose*). 1-Referent Scenes increased Instrument actions for all Verb Types. A finding in the *put*-task that shows brief, albeit attenuated consideration of the Goal analysis in 2-Referent Scenes is consistent with the Snedeker & Trueswell (2004) findings because the *put* materials are temporarily ambiguous rather than globally ambiguous; hearing *into the box* strongly favors a Modifier interpretation of the previous (ambiguous) PP *on the napkin.*

This pattern is predicted by the Constraint-Based Lexicalist (CBL) theory of parsing discussed at the start of this paper (see also MacDonald et al., 1994; Trueswell & Tanenhaus, 1994). Under this account, temporary ambiguity – which arises as a natural byproduct of incremental input – is quickly resolved on the basis of partially independent but interacting representational subsystems (phonological, syntactic, semantic). These modular subsystems are accessed as the input is perceived, and are rapidly coordinated such that probabilistic constraints from each representation conspire to limit the number of ways a sentence is likely to unfold. That is, syntactic information alone does not guide the parser toward a viable interpretation of the input (nor does phonological or semantic information), but it interacts dynamically with evidence from these other subsystems such that constraints at multiple levels are satisfied.

### 5.2. Lexical effects across reading and listening

Even though we did not manipulate verb types in the visual-world experiment (Experiment 1), we did find evidence that parsing preferences in this experiment were related to reliance on lexical tendencies. In particular, individual reliance on lexical constraints correlated positively between reading and listening modalities and even between two different types of syntactically ambiguous sentences, the PP-Attachment Ambiguity and the DO/SC Ambiguity.

Such a pattern is expected if individual differences exist in the ability to rely on detailed lexical information to guide sentence processing decisions (whether they are syntactic or referential decisions). Indeed, it is important to note that HLS individuals (as labeled in this paper) also appear better able to use multiple constraints like lexico-contextual cues to inform their parsing decisions compared to LLS individuals. These same individuals were quite good at arriving at the Modifier interpretation in 2-Referent Scenes, and were even quite good at using the in/on contrast to identify the Target referent (though they did still show some uncertainty as compared to Unambiguous controls, as expected under an account in which multiple constraints were at work). 1-Referent Scenes on the other hand especially tripped these people up, showing, as compared to LLS individuals, increased looks to the Incorrect Goal and even marginal increases in incorrect actions involving this false goal. Again, such a pattern is expected since converging misleading constraints should be particularly misleading to those who are highly tuned to multiple constraints.

### 5.3. Parsing traits or parsing states?

This interpretation of our data suggests that we have identified stable properties of individuals regarding their parsing abilities. However, like most studies of individual differences, another possibility exists, namely that we have identified individuals who for whatever reason came to the laboratory that day performing more efficiently and more effectively than others. Individuals were, after all, tested on all tasks during a single visit. Although such an interpretation is possible, we do not see it as plausible because HLS individuals did not perform better on tasks across the board; they did not do better on the reading span task for instance, and again, counter-intuitively, those who use verb-information in the reading study are three times more likely to make action *errors* in 1-Referent Ambiguous contexts in the listening task. All of this suggests that people came into the laboratory relying on lexical contingencies to a greater or lesser degree. Such a pattern seems less likely to reflect a person's current mood or general alertness and instead reflects properties of these individuals regarding their general parsing abilities. But this is an empirical question; future work could explore whether these individual difference effects replicate when testing is done over several days or weeks.

Future research must also determine what makes some individuals better tuned constraint-users (and in particular highly tuned lexical-constraint users). One possibility worth considering is that these differences reflect individual differences in learning. If differences exist in crucial learning parameters (e.g., related to the precision of hypotheses postulated about lexical items) one might end up with slightly different stable parsing systems in the adult state. We do not expect however that these individual differences are causally related to differences in reading experience, which has been proposed for explaining reading span differences (MacDonald & Christiansen, 2002). Our identified groups do not differ on reading span; moreover our effects are independent of modality. However, it is possible that differences in linguistic experience (particularly spoken linguistic experience) may be contributing to our findings. Some individuals may in fact differ in the amount of linguistic interaction they routinely engage in. Greater linguistic experience of this sort might result in the ability to weigh more linguistic and nonlinguistic evidence, particularly less frequently occurring syntactic contingencies – what might be the cause of these experiential differences in spoken language would at this point be pure speculation.

Finally, it also worth exploring whether HLS individuals commit to interpretive analyses more quickly and more strongly in the face of ambiguity. The patterns indicate that they use lexical contingencies more, but they also suggest that they 'get into trouble' more by doing this (e.g., their errors in actions in Experiment 1). Clearly, all of these findings are fertile ground for modeling efforts (e.g., see Farmer et al., 2007), which might disentangle these dimensions more systematically than what can be done via experimental results alone.

### 5.4. Reading and listening, and cross-validation of popular experimental methods

This finding also addresses the previously untested assumption that there exists a modality-independent mechanism for dealing with syntactic ambiguity

in sentence processing. Despite substantial differences in the measures (reading time vs. looks to an empty napkin) and in the linguistic material, a commonality was hypothesized and confirmed, namely that these measures tap a single comprehension trait pertaining to syntactic ambiguity and its real-time resolution by comprehenders irrespective of modality. Such a finding also provides important cross-validation of these sentence processing methods. Our findings demonstrate that the visual-world method, when used properly and with appropriate linking assumptions, does indeed provide a valuable measure of processes related to sentence parsing. Likewise, reading measures, which only indirectly assess interpretive commitments (via a measure of processing load) appear to be related to eye-movement-during-listening measures, which more directly reflect interpretive commitments of listeners.

### 5.5. Closing remarks

In sum, we view our results as supporting a constraint-based interactive theory of sentence processing in which multiple levels of linguistic and extra linguistic constraints work together to reliably inform the parser in real-time of how interpretation should proceed. Processing moves forward in this fashion regardless of the modality in which the input is encountered, the type of linguistic information that is presented to the parser, or the comprehension environment in which a reader or listener finds herself. However, sensitivity to these constraints differs across individuals; some individuals appear to be highly tuned to lexical contingencies – be they syntactic or referential in nature.

## Acknowledgements

This research was supported by NIH Grant R01-HD37507 (awarded to J.C.T.) and R01-MH67008 (awarded to S.T.S.) and was part of the first author's dissertation. A portion of this work was presented at the 10th Annual Conference on Architectures and Mechanisms for Language Processing in Aix-en-Provence, France, and we thank the conference participants for their feedback. We gratefully acknowledge Susan Garnsey for providing the sentence materials in Experiment 2, and David January for helping run participants and coding eye movement data. We also thank Ted Gibson, Michael Tanenhaus, Julie Sedivy, and an anonymous reviewer for very important suggestions on earlier drafts of this manuscript.

## Appendix A. Contingent eye movement analyses (Experiment 1)

Fig. 5b above plots the proportion of looks to the Target and Competitor animals from the onset of the first preposition *in* or *on*. As can be seen in the figure, there was a baseline difference at the onset of the preposition such that there were more Target looks in the 2-Referent Unambiguous condition as compared with the other condi-

tions. This early difference was not expected, and may be contributing to some of the processing advantage we see for Unambiguous items as time unfolds.

Based on a reviewer's suggestion, we explored this difference further by conducting several contingent eye movement analyses, in which data are divided up based on which region the listener was looking at the onset of the preposition: (1) the Target animal; (2) the Competitor animal; or (3) the Central Fixation (the smiley face). We describe the results of these analyses here. In all cases below, analyses were done on correct action trials only.

### *A.1. Target animal*

Fig. 18 below plots looks to the Target and Competitor animal for those trials on which the listener happened to be fixating the Target animal at the onset of the first preposition. The data are plotted separately for 2-Referent Ambiguous and 2-Referent Unambiguous trials.

We see a striking difference between the Ambiguous and Unambiguous materials. For Unambiguous sentences, listeners were more likely to hold their gaze on the Target animal rather than shifting over to the Competitor. However, for the Ambiguous items, participants frequently look over the Competitor animal. In fact, the initial Target preference briefly becomes a Competitor preference, but reverses back to a Target preference upon hearing *into the box*.

This difference between Ambiguous and Unambiguous items makes sense if, as we have proposed, listeners were partially considering a Goal interpretation of *on the*

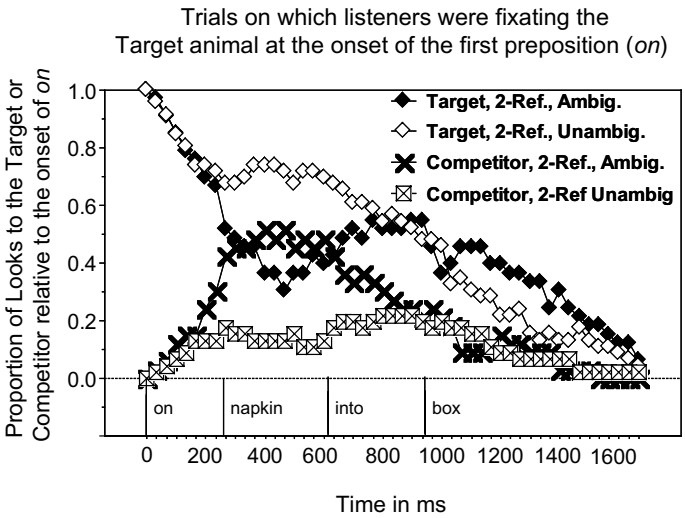


Fig. 18. Proportion of looks to the Target and Competitor animal over time for those trials on which the listener happened to be fixating the Target animal at the onset of the first preposition. The data are plotted separately for 2-Referent Ambiguous and 2-Referent Unambiguous trials.

*napkin* in the Ambiguous condition. If *on the napkin* were a Goal phrase, then there is referential uncertainty as to which animal (which frog) is being referred to, causing listeners to look over to the Competitor animal. Hearing *into the box* (in combination with the contextual constraints) rules out the Goal interpretation and supports the Modifier interpretation, resulting in shifts back to the Target, but at the cost of having to gaze at the Target animal slightly longer in this condition as compared to the Unambiguous condition.

### A.2. Competitor animal

Fig. 19 plots looks to the Target and Competitor animals for those trials on which the listener happened to be fixating the Competitor animal at the onset of the first preposition. The data are plotted separately for 2-Referent Ambiguous and 2-Referent Unambiguous trials.

Here we see that participants quickly shift over to the Target animal upon hearing *on the napkin* in both Ambiguous and Unambiguous trials. However, given the data just discussed, this shift from Competitor to Target is likely to be occurring for different reasons in the Ambiguous and Unambiguous conditions. In the Unambiguous, listeners are likely shifting to the Target because *that's on the napkin* denotes unambiguous reference to the Target. However, for the Ambiguous, at least some proportion of the listeners are likely shifting from the Competitor to the Target because they are uncertain about which animal is being referred to. Indeed, like in the previous graph, we see the characteristic cost of this uncertainty: looks to the Target are prolonged in the Ambiguous, but not in the Unambiguous contexts.

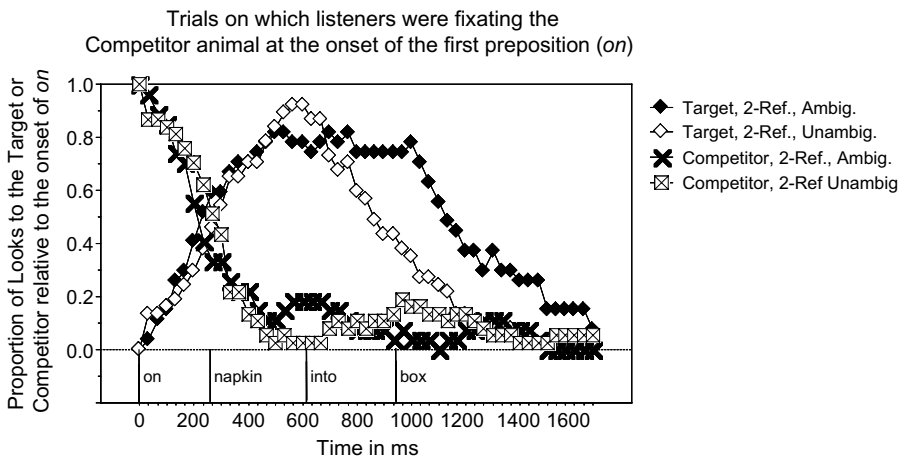


Fig. 19. Proportion of looks to the Target and Competitor animal over time for those trials on which the listener happened to be fixating the Competitor animal at the onset of the first preposition. The data are plotted separately for 2-Referent Ambiguous and 2-Referent Unambiguous trials.

### A.3. Looks to the Central Fixation (the smiley face)

For completeness, graphed below (Fig. 20) are the trials on which participants were still looking at the central fixation at the onset of *on the napkin.*

For these trials, it appears that listeners were relatively unlikely to look at either animal, though there is a clear preference for the Target. Even in this condition though, we see slightly more looks to the Competitor in the Ambiguous as compared to the Unambiguous condition, as well as prolonged looks to the Target late on Ambiguous trials.

The relatively low rates of looking at the Target in this situation could be occurring for several reasons. The most likely explanation is that these trials reflect cases in which participants decided to hold gaze centrally rather than visually inspect the scene further. All of these trials are correct-action trials, so it is unlikely that they reflect severe inattention.

### A.4. Summary and comment on contingent analyses

In sum, division of the data into the different eye position contingencies only serves to bolster our conclusions. Even so, we wish to offer a warning against the use of contingent eye movement analyses. They tend to select small, biased samples of participants, as illustrated in the table below.

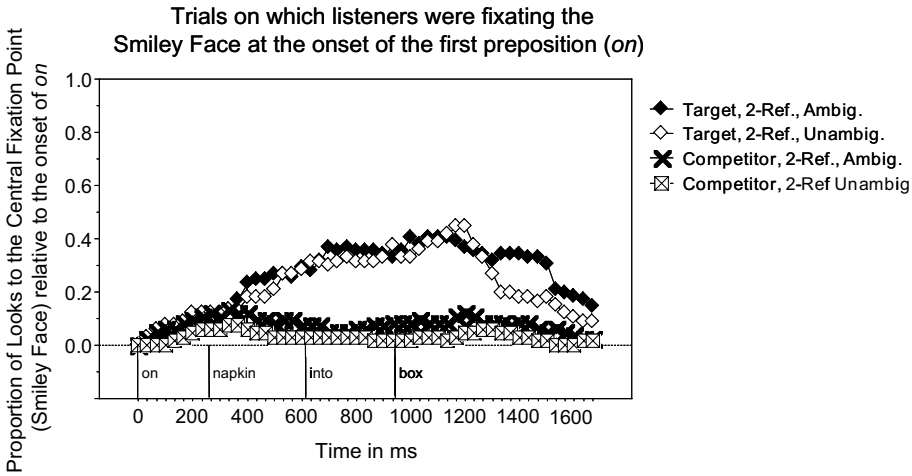


Fig. 20. Proportion of looks to the Target and Competitor animals over time for those trials on which the listener happened to be fixating the Central Fixation Point (the smiley face) at the onset of the first preposition. The data are plotted separately for 2-Referent Ambiguous and 2-Referent Unambiguous trials.

| | Looking at the Target at P1[a] | | Looking at the Competitor at P1 | | Looking Centrally at P1 | |
|---|---|---|---|---|---|---|
| | Ambig. | Unambig. | Ambig. | Unambig. | Ambig. | Unambig. |
| Observations | 33/144 (23%) | 46/154 (30%) | 27/144 (19%) | 37/154 (24%) | 82/144 (57%) | 67/154 (44%) |
| Subjects contributing data | 21/40 (53%) | 27/40 (68%) | 15/40 (38%) | 27/40 (68%) | 31/40 (78%) | 29/40 (73%) |
| Subjects contributing data to both conditions | 19/40 (48%) | | 12/40 (30%) | | 26/40 (65%) | |

[a] Note: P1 = first preposition of target sentence (e.g., *on*).

## Appendix B. Experimental materials from Experiment 2

Ambiguous versions appeared without the complementizer *that;* Unambiguous versions contained the word *that*.

### *B.1. DO-bias items*

1. The talented photographer accepted (that) the money could not be spent yet.
   Q: Did the photographer accept the money?
2. The newspaper editor advocated (that) the truth should not be made public.
   Q: Did the editor advocate the truth?
3. The concerned priest asserted (that) the belief would be hard to justify.
   Q: Did the priest assert the belief?
4. The CIA director confirmed (that) the rumor should have been stopped sooner.
   Q: Did the CIA director confirm the rumor?
5. The scuba diver discovered (that) the wreck was not where he was told it would be.
   Q: Did the scuba diver discover the wreck?
6. The angry father emphasized (that) the problems should be ignored.
   Q: Did the father emphasize the problems?
7. The primary suspect established (that) the alibi had been a total lie.
   Q: Did the subject establish the alibi?
8. The gossipy neighbor heard (that) the story was told to everyone but her.
   Q: Did the neighbor hear the story?
9. The confident engineer maintained (that) the machinery would be hard to destroy.
   Q: Did the engineer maintain the machinery?
10. The journal editor printed (that) the quote had been slanderous to him.
    Q: Did the editor print the quote?
11. The young teacher wrote (that) the essay had been very poorly revised.
    Q: Did the teacher write the essay?
12. The surgical nurses protested (that) the policy was not a priority.
    Q: Did the nurses protest the policy?
13. The alert consumer understood (that) the label had been removed from the jar.
    Q: Did the consumer understand the label?

14. The trained referees warned (that) the spectators would probably get too rowdy.
    Q: Did the referees warn the spectators?
15. The city planners proposed (that) the strategy did not include land development.
    Q: Did the planners propose the strategy?
16. The French instructor repeated (that) the poem should be finished by Friday.
    Q: Did the teacher repeat the poem?
17. The chemistry student learned (that) the equations could have made measurement more precise.
    Q: Did the student learn the equations?
18. The substitute locksmith forgot (that) the key was made of copper.
    Q: Did the locksmith forget the key?
19. The school board established (that) the policy would get a proper hearing.
    Q: Did the board establish the policy?
20. The devoted caretaker maintained (that) the garden was causing his chronic allergies.
    Q: Did the caretaker maintain the garden?

### *B.2. SC-bias items*

1. The ticket agent admitted (that) the mistake had nothing to do with the price of the fare.
    Q: Did the agent admit the mistake?
2. The district attorney argued (that) the point would be impossible to avoid.
    Q: Did the attorney argue the point?
3. The friendly clerk indicated (that) the price would be rising very soon.
    Q: Did the clerk indicate the price?
4. The job applicant believed (that) the interviewer had been dishonest with her.
    Q: Did the applicant believe the interviewer?
5. The weary traveler claimed (that) the luggage had been stolen in Rome.
    Q: Did the traveler claim the luggage?
6. The reading tutor concluded (that) the lesson could be finished very quickly.
    Q: Did the tutor conclude the lesson?
7. The cab driver assumed (that) the blame belonged to the other driver.
    Q: Did the cab driver assume the blame?
8. The math whiz figured (that) the sum could be calculated very quickly.
    Q: Did the whiz figure the sum?
9. The experienced judge decided (that) the appeal should be started right away.
    Q: Did the judge decide the appeal?
10. The careful scientist proved (that) the theory had not been sufficiently tested.
    Q: Did the scientist prove the theory?
11. The travel agent suggested (that) the vacation would just have to wait.
    Q: Did the agent suggest the vacation?
12. The film director suggested (that) the scene should be removed completely.
    Q: Did the director suggest the scene?
13. The factory owner suspected (that) the workers would love the Christmas bonus.
    Q: Did the owner suspect the workers?
14. The bus driver worried (that) the passengers were starting to get annoyed.
    Q: Did the bus driver worry the passengers?
15. The rejected bachelor inferred (that) the reason had never been openly stated.
    Q: Did the bachelor infer the reason?
16. The class president assumed (that) the burden would be shared by everyone.
    Q: Did the president assume the burden?
17. The young babysitter worried (that) the parents would be several hours late.
    Q: Did the babysitter worry the parents?
18. The naive child believed (that) the fable had been written for him.
    Q: Did the child believe the fable?

19. The anonymous informant claimed (that) the reward should not be given to him.
    Q: Did the citizen claim the reward?
20. The account executive concluded (that) the speech should not be delivered.
    Q: Did the executive conclude the speech?

## References


Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*(3), 191–238.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33*, 251–283.

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*, 30–49.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30*, 687–696.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107.

Crain, S., & Steedman, M. J. (1982). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives. ACL studies in natural language processing* (pp. 320–358). Cambridge: Cambridge Univeristy Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Elman, J., Hare, M., & McRae, K. (2004). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 111–138). Mahwah, NJ: Lawrence Erlbaum.

Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57*(4), 570–595.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Ferreira, F., & Henderson, J. M. (1990). The use of verb information in syntactic parsing: A comparison of evidence from eye movements and segment-by-segment self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 555–568.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalysis of garden- path sentences. *Journal of Memory and Language, 30*, 725–745.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of reanalysis. *Journal of Psycholinguistic Research, 30*, 3–20.

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*, 58–93.

Green, M. J., & Mitchell, D. C. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 55*, 1–17.

Holmes, V. M., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language, 28*, 668–689.

Kim, A., Srinivas, B., & Trueswell, J. C. (2002). The convergence of lexicalist perspectives in psycholinguistics and computational linguistics. In P. Merlo & S. Stevenson (Eds.), *Sentence processing and the lexicon: Formal, computational, and experimental perspectives* (pp. 109–135). Philadelphia, PA: John Benjamins Publishing.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35–54.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics, 53*, 372–380.

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*, 134–157.

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 432–460.

Meroni, L., & Crain, S. (2003). On not being led down the garden path. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Boston university conference on language development*. Somerville, MA: Cascadilla Press.

Novick, J. M., Kim, A., & Trueswell, J. C. (2003). Studying the grammatical aspects of word recognition: Lexical priming, parsing, and syntactic ambiguity resolution. *Journal of Psycholinguistic Research, 32*, 57–75.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 940–961.

Sedivy, J., & Spivey-Knowlton, M. (1994). The use of structural, lexical, and pragmatic information in parsing attachment ambiguities. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 389–413). Hillsdale, NJ: Erlbaum.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238–299.

Spivey, M., & Tanenhaus, M. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1521–1543.

Spivey, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447–481.

Spivey-Knowlton, M. J., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*(3), 227–267.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Parsing reduced relative clauses. *Canadian Journal of Psychology. Special Issue. Reading and Language Processing, 47*(2), 276–309.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

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.

Trueswell, J. C., & Gleitman, L. R. (2004). Children's eye movements during listening: Evidence for a constraint-based theory of parsing and word learning. In J. M. Henderson & F. Ferreira (Eds.), *Interface of language, vision, and action: Eye movements and the visual world*. NY: Psychology Press.

Trueswell, J. C., & Kim, A. (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language, 39*(1), 102–123.

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

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*(3), 528–553.

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*, 285–3 18.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*, 89–134.

Trueswell, J. C., Papafragou, A. & Choi, Y., (in press). Syntactic and referential processes: What develops? In E. Gibson and N. Pearlmutter (Eds.), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 52*, 284–307.