ABSTRACT

Title of Dissertation:	Understanding and remembering pragmatic inferences
	Alix Kowalski, Doctor of Philosophy, 2018
Dissertation directed by:	Assistant Professor Yi Ting Huang, Hearing and Speech Sciences

This dissertation examines the extent to which sentence interpretations are incrementally encoded in memory. While traditional models of sentence processing assume that comprehension results in a single interpretation, evidence from syntactic parsing indicates that initial misinterpretations are sometimes maintained in memory along with their revised counterparts (e.g., Christianson, Hollingworth, Halliwell & Ferreira, 2001). However, this evidence has largely come from experiments featuring sentences that are presented in isolation and words that are biased toward incorrect syntactic analyses. Because there is typically enough sentential context in natural speech to avoid the incorrect analysis (Roland, Elman, & Ferreira, 2006), it is unclear whether initial interpretations are incrementally encoded in memory when there is sufficient context. The scalar term “some” provides a test case where context is necessary to select between two interpretations, one based on semantics (some and possibly all) and one based on pragmatic inference (some but not all) (Horn, 1989). Although listeners strongly prefer the pragmatic interpretation (e.g., Van Tiel, Van

Miltenburg, Zevakhina, & Geurts, 2016), prior research suggests that the semantic meaning is considered before the inference is adopted (Rips, 1975; Noveck & Posada, 2003; Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; De Neys & Schaeken, 2007; Huang & Snedeker, 2009, 2011). I used a word-learning and recall task to show that there is evidence of the semantic meaning in the memory representation of sentences featuring "some," even when the pragmatic interpretation is ultimately adopted. This raises two possibilities: first, the memory representation was of poor quality because both interpretations were available during encoding, or the semantic meaning was computed and encoded first and lingered even after the pragmatic interpretation was computed and encoded. Data from a conflict-adaptation experiment revealed a facilitating effect of cognitive control engagement. However, there was still a delay before the pragmatic inference was adopted. This suggests that only the semantic meaning is available initially and the system failed to override it in memory when the pragmatic interpretation was computed. Taken together, these findings demonstrate the incrementality of memory encoding during sentence processing.

UNDERSTANDING AND REMEMBERING PRAGMATIC INFERENCES

by

Alix Kowalski

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Assistant Professor Yi Ting Huang, Chair
Professor Jan Edwards
Assistant Professor Jared Novick
Professor Colin Phillips
Associate Professor L. Robert Slevc

Dedication

To my dad. I love you big time, forever and ever and two weeks.

Acknowledgements

I would like to sincerely thank my dissertation committee for their continued advice and encouragement throughout the research process. I am grateful to Yi Ting Huang, my advisor and committee chair, for always pushing me to do my best. To Jared Novick, who always made me feel supported. To Bob Slevc, who showed me the fun in research and let me draw cartoons for our research poster and study. To Jan Edwards, who gave me valuable advice on my data analysis. And finally, to Colin Phillips, who wrote me an extremely welcoming email when I was first accepted to Maryland, which ended up being a true representation of the kind, encouraging, and ever approachable Language Science Community at Maryland.

Coming to grad school, I was extremely lucky to befriend and live with Zoe Schlueter for five years. I know I would not have been able to get through this program without her constant support and willingness to watch an ungodly amount of British murder mysteries. She was always there to help me get through the difficult moments and I'm glad we could celebrate the victories together as well. I miss you terribly, and so do the cats (even Bruce).

I am also grateful to Rachel Adler, who was my soundboard and lunch date for years. Along with Alia and Zeke Lancaster, Mike Fetters, Eric Pelzl, Peter Enns (even though he went on to better things) and Chris Heffner, Rachel was (and still is) a constant source of love and friendship. Thank you also to everyone whom I shared classes with, the rest of the Hearing and Speech Science Department, and the Language Science Community. A big thanks also to the administrators who made my life easier at many turns, especially Valerie Brownlee in HESP and Tess Wood of the Language Science Center.

Table of Contents

List of Tables

List of Figures

Chapter 1

1.1 Overview

"Time flies like the wind. Fruit flies like a banana." This classic line attributed to comedian Groucho Marx exploits a fundamental characteristic of the language processing system: its incrementality. Rather than wait until the end of a sentence in (1), listeners generate hypotheses about what the speaker means as the sentence is unfolding. Groucho's set-up biases listeners toward one interpretation of the sentence (2a) before the punchline forces a reinterpretation (2b). Humor is thought to arise when listeners contrast the old and new interpretations (Attardo & Raskin, 1991; Attardo, 1997; Nerlich & Clarke, 2001; Ritchie, 1999; Dynel, 2009ab).

(1) "Fruit flies like a banana."

a. Initial interpretation: Fruit flies in a similar way to a banana

b. Revised interpretation: Fruit flies enjoy eating bananas

Recognizing this contrast would require a language processing system that builds and maintains both interpretations in memory, even after one is proven to be "correct." While there is evidence from syntactic processing that readers sometimes maintain initial interpretations after reanalysis (Christianson, Hollingworth, Halliwell & Ferreira, 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, Christianson, & Hollingworth, 2001; Patson, Darowski, Moon, & Ferreira, 2009; Christianson, 2008; Slattery, Sturt, Christanson, Yoshida & Ferreira, 2013; Malyutina & den Ouden, 2016), the sentences in these studies are known as "garden-paths" because they guide comprehenders toward the incorrect interpretation. For example, “drinks” is typically a transitive verb, leading readers to initially assume that “wine” is its object in “Because Bill drinks wine is never kept in the house” (Ferreira & Henderson, 1991). However, speakers are usually cooperative and provide as much information as is required to convey their meaning (Grice, 1975). Indeed, corpus analyses show that there is typically enough syntactic and semantic information from a garden-path sentence’s context to disambiguate its meaning (Roland, Elman, & Ferreira, 2006). Thus, it may be the case that incorrect interpretations only persist in the unique situations where word biases and a lack of context encourage comprehenders to misanalyse a sentence, like during linguistic experiments or joke-telling. It may be that during typical language comprehension, when sufficient context is provided, only a single interpretation is maintained in memory.

One way to test whether multiple interpretations are maintained when context is available is to examine a case where context is necessary to select one meaning of a word over another. For example, in sentences like (3), the semantic meaning of “some” is consistent with any quantity greater than none (4a). However, a listener could infer that because the speaker did not use the stronger term “all,” it must not apply in this context (4b) (Horn, 1972, 1989; Gadzar, 1979). Although listeners strongly prefer the pragmatic meaning of “some” (Noveck & Posada, 2003; Bott & Noveck, 2004; Geurts & Pouscoulous, 2009; Van Tiel, Van Miltenburg, Zevakhina, & Geurts, 2016), prior research suggests that its semantic meaning is considered before the inference is made (Rips, 1975; Noveck & Posada, 2003; Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; De Neys & Schaeken, 2007; Huang & Snedeker, 2009, 2011).

(3) "The girl ate some of the cookies"

(4) a. Semantic: The girl ate some and possibly all of the cookies

b. Pragmatic: The girl ate some but not all of the cookies

Thus, at the point at which comprehenders accept the pragmatic inference they have entertained both the semantic and pragmatic meanings of "some." Is the semantic meaning included in the interpretation that is carried forward in memory? Or is it replaced by the pragmatic inference? One possibility is that during real-time comprehension, meaning is encoded in memory after a final interpretation is reached. If this is the case, then there should be minimal traces of the semantic meaning when the pragmatic inference overrides initial semantic analysis. Another possibility is that during comprehension, meaning is incrementally encoded in memory, even before a final interpretation is reached. If this is the case, then the semantic meaning may persist in memory and possibly emerge in later recall.

In the remainder of this Introduction, I first discuss seemingly conflicting evidence about the generation of multiple interpretations during word recognition and syntactic parsing. On the one hand, interpretations based on initial misanalyses are sometimes encoded in memory along with their updated counterparts. However, other data suggest that context is quickly used to select a single interpretation to be encoded in memory. Next, I discuss past accounts and how they characterize processing when multiple interpretations are available. Then I describe an alternative account of the existing data that includes predictions about pragmatic inferencing. Finally, I provide an outline of the dissertation.

1.2 Interpretation building in word recognition and syntactic parsing

Evidence of the incrementality of interpretation building has come from word recognition and syntactic parsing. For example, phono-semantic priming suggests that hearing the start of a word activates multiple lexical representations. When presented with "Pick up the beaker," listeners were more likely to initially fixate on a phonological competitor (e.g., a beetle) than an object with a phonologically unrelated name (e.g., a carriage). This suggests that all phonologically consistent lexical representations are triggered as a word unfolds (Allopenna, Magnuson, & Tanenhaus, 1998). Moreover, at the point of word onset (e.g., "Point to the lo-"), listeners look to depictions of the semantic relations of phonologically related words (e.g., looks to KEY which is related to "lock") (Yee & Sedivy, 2006). This suggests that the meanings, not just the form, of words that fit the phonological context (e.g., "lock") and words that could not possibly be the target (e.g., "key") are incrementally activated as the word unfolds.

Evidence from syntactic processing suggests that the activation of multiple syntactic representations can lead to the generation of multiple interpretations. A series of studies by Ferreira and colleagues have demonstrated that readers often maintain initial misinterpretations, even after sentence completion (Christianson et al., 2001; Christianson et al., 2006; Ferreira et al., 2001; Patson et al., 2009; Christianson, 2008; Slattery et al., 2013; Malyutina & den Ouden, 2016). For example, after reading temporarily ambiguous sentences like “While Bill hunted the deer that was brown and graceful ran into the woods,” participants incorrectly answered, “yes” to “Did Bill hunt the deer?” 56% of the time. Critically, they correctly answered, “yes” to “Did the deer run into the woods?” 91% of the time, bringing the total percentage of “yes” answers to the two questions to well over 100% (Christianson et al., 2001). This suggests that, initially, “the deer” was interpreted as the thing being hunted, and this misanalysis persisted even when the correct interpretation was adopted, leading to the encoding of both interpretations in memory. Also, when presented with three pictures and asked to select the one that matched the sentence “While the man hunted the deer ran into the woods,” listeners correctly identified the depiction consistent with a reanalyzed interpretation (e.g., a man hunting a bird and a deer running into the woods) 84% of the time. Notably, 69% of incorrect responses were based on a “blended interpretation” between the initial and revised analyses (e.g., a man hunting a deer as the deer runs into the woods) (Malyutina & den Ouden, 2016). These results suggest that the initial and revised interpretations both persisted in the memory of the utterance. This may occur when there is prolonged uncertainty about which syntactic analysis is correct. Worse performance on comprehension questions about sentences that trigger uncertainty (e.g. “The coach smiled at the player tossed the Frisbee”) versus sentences that do not (e.g., “The coach smiled at the player thrown the Frisbee”) suggest that uncertainty can be maintained throughout the sentence (Levy, Bicknell, Slattery, & Rayner, 2009). This further suggests that multiple interpretations can be built and maintained in memory, one for each of the competing syntactic analyses.

However, prior studies have also shown that context can be used to rapidly constrain the set of interpretations under consideration, which may allow the system to avoid generating multiple interpretations altogether. For example, evidence from word recognition studies suggests that listeners often activate only contextually consistent lexical representations, rather than all phonologically possible forms (Becker, 1980; Goldinger, Luce, & Pisoni, 1989; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Swinney, 1979). For example, when presented with a constraining context like "Never before climbed a go-…" Dutch-speaking listeners looked to the contextually appropriate picture of a goat ("bok") and not to a word that shared the same onset, e.g., "bot" (bone) (Dahan & Tanenhaus, 2004).

In syntactic processing, visual context is often used to activate only contextually appropriate representations (Tanenhaus et al., 1995; Altmann & Kamide, 1999; Kamide, Scheepers, & Altmann, 2003; Kamide, Altmann, & Haywood, 2003; Chambers, Tanenhaus, & Magnuson, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). For example, when listeners heard "Put the apple on the towel into the box" while viewing a display of an apple on a towel and a pencil, their eye movements to an empty towel suggested that they initially interpreted "on the towel" as indicating a destination for the apple. However, when the pencil was exchanged for a second apple, participants immediately interpreted the phrase as a modifier and looked to the box before hearing “on the box” (Tanenhaus et al., 1995).

Context can be so constraining that listeners are able to predict upcoming linguistic material. For example, when presented with sentences like “The boy will eat the…” while looking at a display of a boy, a cake, and several non-edible items, listeners looked to the cake before the start of the word “cake.” This suggests that listeners used the semantics of the verb “eat” to predict the upcoming noun. Also, when presented with mini-stories like “The burglar had no trouble finding the family safe. Of course, it was situated behind a…” Dutch readers spent a longer time reading adjectives (e.g., big) that did not match the syntactic gender of the noun predicted by the discourse (e.g., painting) (vanBerkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). This suggests that readers were anticipating the upcoming noun and experienced processing difficulty when the adjective did not match the noun they predicted. These findings provide empirical support for the view that the language processing system uses contextual information to quickly resolve competition between lower-level representations and selects a single representation for interpretation.

In sum, part of the empirical evidence suggests that initial interpretations can be built and encoded in memory before a final interpretation is reached (Christianson et al., 2001; Christianson et al., 2006; Ferreira et al., 2001; Patson et al., 2009; Christianson, 2008; Slattery et al., 2013; Malyutina & den Ouden, 2016). However, other evidence suggests that the processing system is rapidly influenced by prior context and generates interpretations only after the competition between phonological, lexical, and syntactic representations has been resolved (Frazier & Fodor, 1978; MacDonald et al., 1994; Trueswell et al., 1993). In the following section, I will describe how these accounts deal with situations in which multiple representations are consistent with the linguistic input. This can give us a hint as to when the models would predict an interpretation is encoded in memory. Must competition between representations be resolved before the system moves on to new linguistic input or can multiple representations continue in the interpretive process?

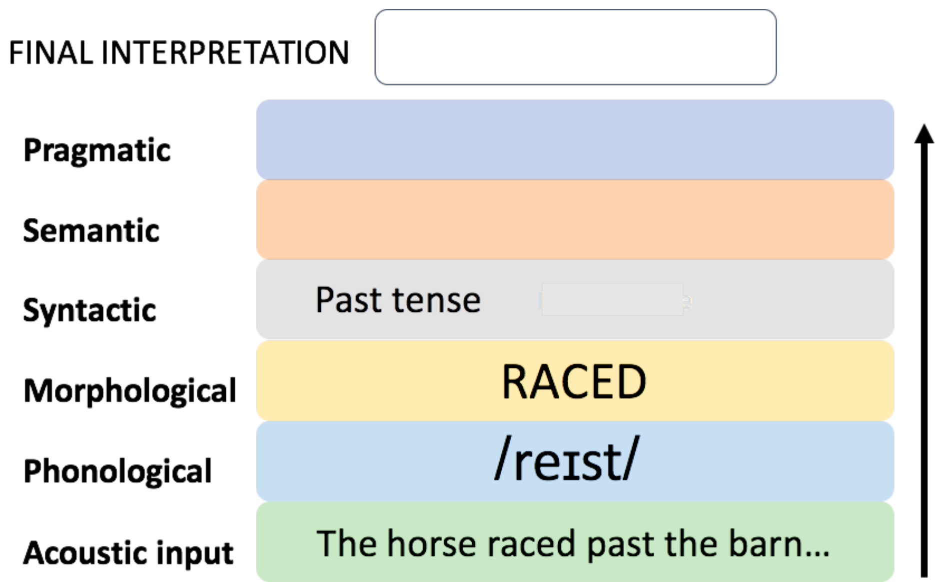
1.3 Models of sentence processing that predict single interpretations

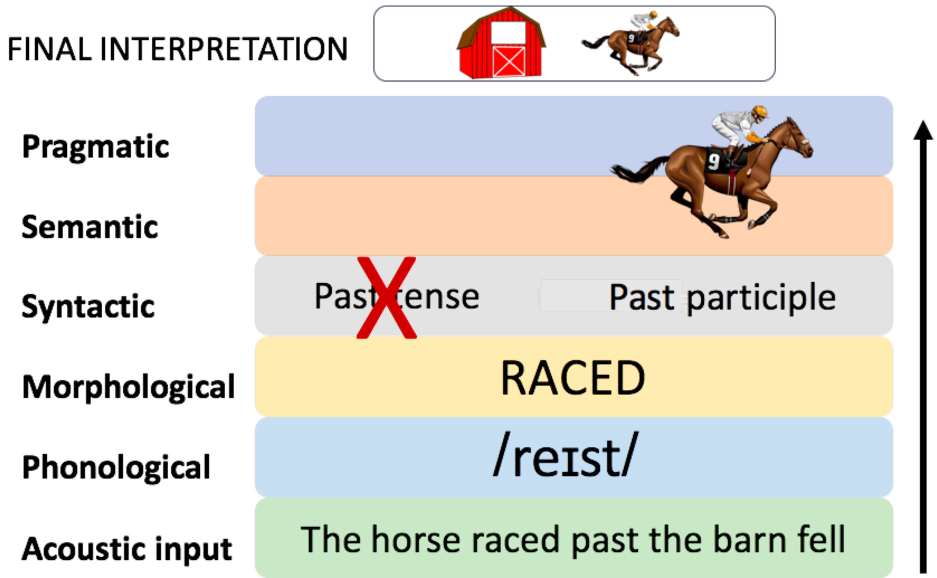
While past accounts of sentence processing were developed with different goals in mind, these models can make predictions as to when the final interpretation of a sentence is encoded in memory. Classic models of parsing like two-stage (Frazier & Fodor, 1978; Frazier, 1979; Clifton & Frazier, 1989) and constraint-based models (MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell, Tanenhaus, & Kello, 1993; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) were designed to explain how the processing system deals with temporary syntactic ambiguities. Baysian approaches like surprisal theory and noisy-channel models describe how a system deals with errors in the language input, either due to a noisy environment, speaker errors, or perceiver errors (Hale, 2001; Levy, 2008; Levy et al., 2009; Gibson, Bergen, & Piantadosi, 2013). The Good Enough model challenges the assumption that the principal aim of the language comprehension system is to derive a complete, accurate, and detailed representation of the speaker's utterance (Ferreira, Bailey, & Ferraro, 2002; Ferreira et al., 2001; Ferreira & Patson, 2007).

One example where the linguistic input is consistent with two different representations is in the case of temporary syntactic ambiguity. For example, in (5), the word "raced" is temporarily ambiguous: prior to the word "fell" a comprehender may treat "raced" as a past tense verb, rather than the past participle introducing a reduced relative clause.

(5) The horse raced past the barn fell.

According to two-stage models, the system computes syntactic analyses in two stages (Frazier & Fodor, 1978; Frazier, 1979; Clifton & Frazier, 1989). This is depicted in the diagram in Figure 1. In the first stage, an initial analysis is built using only syntactic information. Because this first stage parser has limited working memory capacity, it cannot access what comes later in sentence, beyond the phrase it is currently analyzing. Moreover, the system favors analyses which cost the least amount of effort. Thus, for the sentence in (5), the first stage parser would encounter "raced" and treat it as a past tense verb since a simple declarative sentence is a less complicated structure than a reduced relative clause. In the second stage, other sources of information can be used, like semantic or contextual information. The second stage parser also has unlimited working memory capacity and it combines multiple phrases together. Thus, it is not until the second stage that the parser would attempt to reconcile the word "fell" with the rest of the sentence. Critically, only one structure is considered at a time. Thus, reanalyzing the word "raced" in the second stage would completely replace its syntactic representation as a past tense verb with a representation as a past participle. This suggests that, within a two-stage model, an initially incorrect parse that was built in the first stage is lost after it is shunted to the second stage parser and reanalyzed. The result is a single interpretation of the utterance.

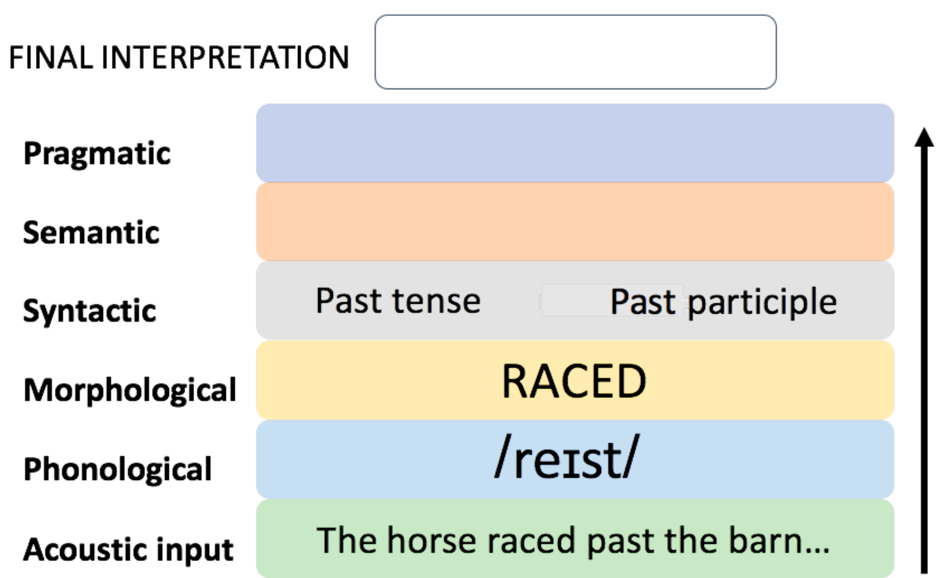


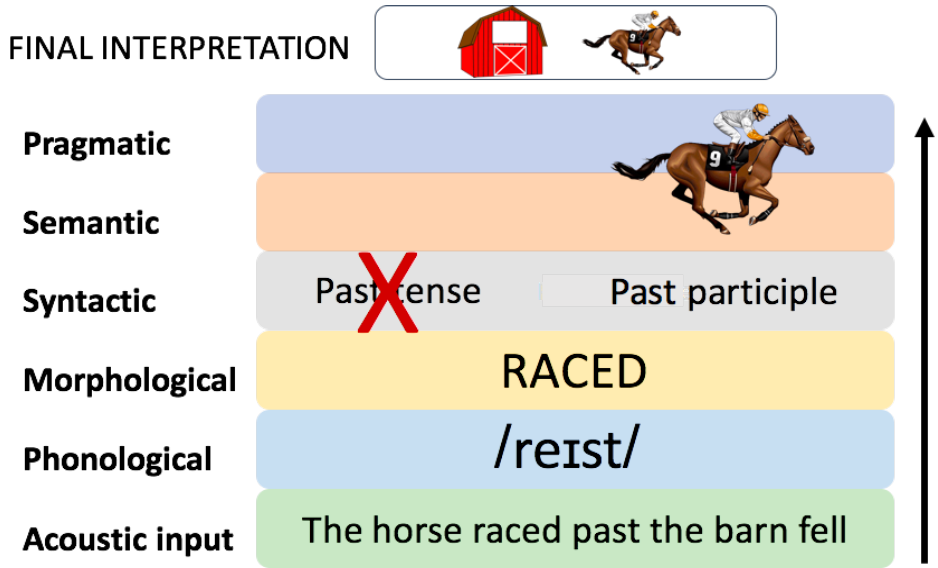


b.

Figure 1. Two-stage model of sentence processing. In (a), based on the speech segment "The horse raced past the barn...," the processing system has completed speech sound segmentation, word retrieval, and analyzed the word "raced" as a transitive verb. In (b), the system has encountered the verb "fell," triggering a re-analysis of "raced" as a past participle. The initial syntactic analysis is completely discarded and a final interpretation is reached.

In contrast, constraint-based systems consider multiple syntactic structures in parallel, but only interpret the one that has the most contextual support (MacDonald et al., 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell et al., 1993; Tanenhaus et al., 1995). This is depicted in the diagram in Figure 2. Unlike earlier views that assumed syntactic processing involved constructing one structure at a time, constraint-based models hypothesize that multiple structures can be partially activated. The degree of activation depends on the amount of evidence in favor of a particular structure. This evidence can come from any informational source, including prior sentence context and extra-linguistic knowledge like the visual scene, as soon as it is available. Disambiguation occurs when one structure is selected and the system inhibits all alternatives. Importantly, constraint-based models assume that disambiguation must be achieved at all levels of representation. Thus, the system quickly rules out context-incompatible structures and any initially incorrect hypotheses are actively inhibited by the selection of an alternative. This suggests that, within the constraint-based system, only a single representation is encoded in memory, *after* competition between representations is resolved.





b.

Figure 2. Constraint-based model of sentence processing. In (a), based on the speech segment "The horse raced past the barn...," the processing system has completed speech sound segmentation, word retrieval, and "raced" has been analyzed as a transitive verb and as a past participle. In (b), the system has encountered the verb "fell," eliminating the representation of "raced" as a past participle. The incorrect syntactic analysis is completely discarded and a final interpretation is reached.

1.4 Models of sentence processing that predict multiple interpretations

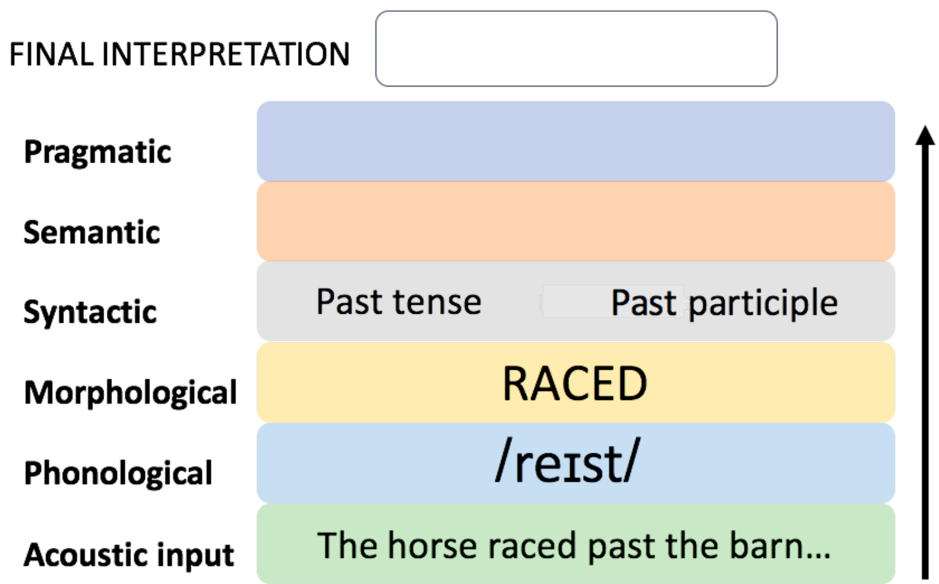
Both the two-stage and constraint-based accounts of language processing predict that an interpretation is encoded in memory after competition between representations has been resolved or an initial misanalysis has been revised. This results in a single interpretation of the sentence in question. In contrast, more recent language processing models describe systems where multiple interpretations may persist in memory. In particular, the Good Enough model developed by Ferreira and colleagues grew out of the observation that sentence comprehension does not always result in a single, detailed representation of a sentence. Instead, The Good Enough model assumes that the goal of comprehension is to quickly build a sensible meaning for linguistic input, not necessarily a completely accurate representation (Ferreira et al., 2001; Ferreira et al., 2002; Ferreira & Patson, 2007). This means that sentences are often incorrectly interpreted or incompletely revised. When this happens, misinterpretations can persist in memory even when there is evidence that a new, revised interpretation was generated. Support for this assumption comes from evidence that the interpretation based on an incorrect syntactic analysis is maintained in memory even after the sentence is reanalyzed and a new interpretation is made (Christianson et al., 2001; Christianson et al., 2006; Ferreira, 2003; Ferreira et al., 2001; Slattery et al., 2013). This can lead to a final interpretation that is only partially reanalyzed and not necessarily consistent with the linguistic input. However, this theory claims that interpretations "persist in memory" but is not specific about what type of memory is involved (e.g., working memory or long-term memory).

The possibility of the system activating and maintaining multiple interpretations is also alluded to by noisy-channel models. Noisy-channel models have proposed a highly rational system that simultaneously infers likely sentences based on speakers' intended meaning (Levy, 2008; Levy et al., 2009; Gibson et al., 2013). Within this system, the final interpretation is not completely determined by a sentence's syntax. Rather than only using meaning to guide interpretation during temporary syntactic ambiguity, semantic and contextual information are used to generate expectations about likely sentences throughout the interpretation process. Notably, within these rational models, the processor maintains uncertainty about word identities and constantly reassesses based on grammatical information from subsequent input. This suggests that, under conditions of uncertainty, multiple word meanings can be maintained, at least until later in the sentence (Levy et al., 2009). However, the noisy-channel model does not explicitly claim that full sentence interpretations are built and maintained in memory, or predict how long multiple word meanings can persist. Again, these prior accounts of sentence processing do not explicitly describe what information is carried forward in memory.

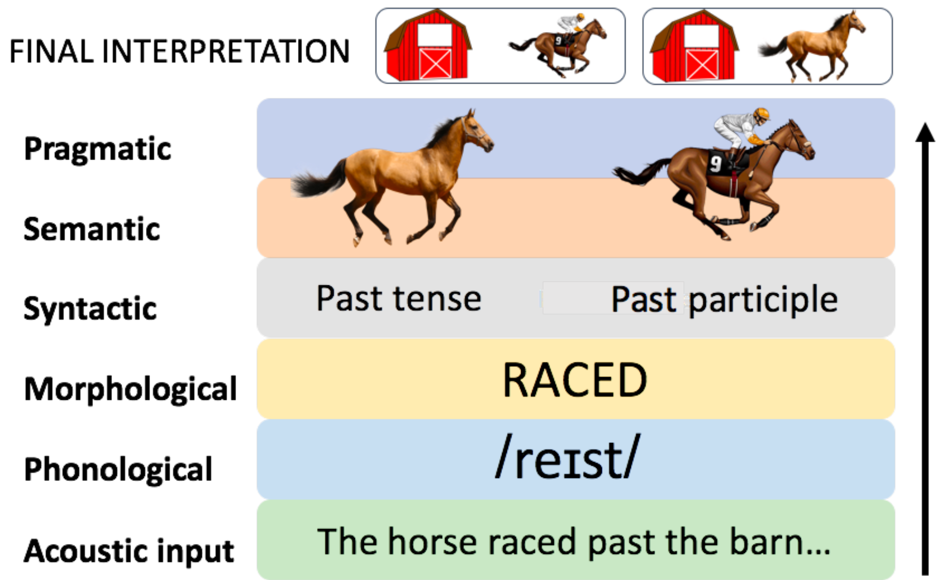
1.5 A model of memory encoding during interpretation building

A model that specifically addresses when the language processing system encodes an interpretation in memory can be developed with these prior accounts in mind. One such account is depicted in the diagram in Figure 3. The defining feature of this model is that multiple interpretations can be incrementally committed to memory as a sentence is processed. One way this may be accomplished is through memory consolidation similar to the "Chunk-and-Pass" system described by Christiansen and Chater (2016). Incoming linguistic material is recoded into chunks and passed to progressively higher levels of linguistic representation (e.g., lexical, syntactic). This input can be chunked in more than one way simultaneously. For example, the lexical item “raced” in Figure 3 may be recoded with the previous item “the horse” at the syntactic level as both a transitive structure and a reduced relative clause. Unlike the Chunk-and-Pass framework in Christiansen and Chater (2016), the present model allows these competing lower-level codes to be recoded into higher-level chunks that are activated in parallel. At the highest levels, chunks are integrated with pragmatic context to form discourse-level structures. The “Final Interpretation” box in Figure 3 signifies the output(s) of this integration.

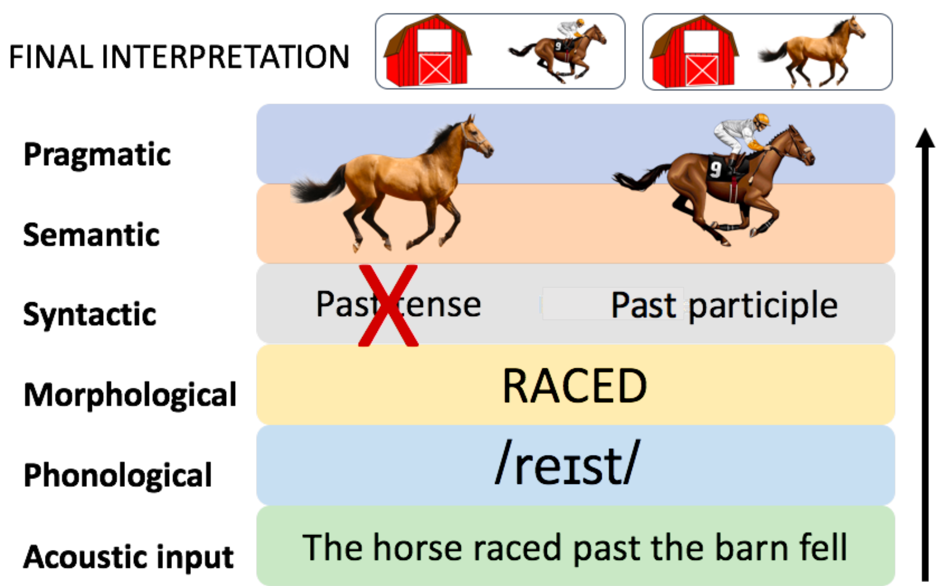
Prior work in syntactic processing suggests that multiple interpretations can persist in memory long enough to interfere with recall well after a sentence has completed (e.g., when answering comprehension questions). This suggests that interpretations are maintained in memory longer than a limited capacity system like working memory would allow. Thus, while the chunking process occurs in a limited capacity system, final interpretations are ultimately committed to a longer-term memory store. Within a classic model of memory (e.g., Baddeley & Hitch, 1974) this would require a transfer from working memory to long-term memory. However, more recent theories of memory in language comprehension draw a distinction between *limited capacity buffers* that retrieve representations from *declarative memory*, which can include short- and long-term memory (e.g., Lewis & Vasishth, 2005). These representations are bundles of features (e.g. singular noun) that have assigned activation values based on prior usage and decay over time. Chapter 5, Section 2 describes how multiple interpretations may persist in an activation-based memory framework like this. However, the experiments presented here were designed to probe for the existence of multiple interpretations in memory and not distinguish between these memory architectures. So, to simplify terminology for the question at hand, an interpretation will be called "encoded" when it is out of working memory/the focus of attention and in long-term/short-term/declarative memory. In the case of scalar inferences, this model of processing proposes that two interpretations of the word "some" will persist in memory. When "some" is encountered, the semantic meaning is retrieved and integrated into an interpretation which is then encoded in memory. When the pragmatic inference is generated, it is also committed to memory.



a.



b.



c.

Figure 3. An alternative account. In (a), based on the speech segment "The horse raced past the barn...," the processing system has completed speech sound segmentation, word retrieval, and "raced" has been analyzed as a transitive verb and as a past participle. In (b), interpretations based on both syntactic analyses are built.

In (c), the system has encountered the verb "fell," eliminating the syntactic representation of "raced" as a past participle. However, both interpretations persist in memory.

In order to test this account, it is necessary to incorporate two characteristics that have been lacking from prior studies that examined persisting interpretations in memory. First, the evidence to date has largely focused on syntactic parsing, which relies on representations that are internal to linguistic processing. In these cases, competition between syntactic representations is ultimately resolved by linguistic disambiguation (e.g., hearing "…ran into the woods" after "While Bill hunted the deer…" tells you that Bill did not hunt the deer). Knowing that a sentence will ultimately be disambiguated may decrease a comprehender's motivation to immediately select a single interpretation. Critically, comprehension also involves higher-level analyses like pragmatic inferencing, which operate over meaning-based representations and offer no linguistic disambiguation.

Second, the evidence that multiple interpretations are formed comes from cases where limited context is available to select a single representation. Unlike naturalistic speech, garden-path sentences in studies probing for persisting memory interpretations are typically presented in isolation (Christianson et al., 2001; Christianson et al., 2006; Ferreira et al., 2001; Patson et al., 2009; Slattery et al., 2013; Malyutina & den Ouden, 2016; cf. Christianson, 2008). However, corpus analyses show that there is generally sufficient sentence information to adopt the correct interpretation of an ambiguous verb (Roland et al., 2006). Thus, the activation of an initially incorrect syntactic analysis may be avoided by a strongly constraining context, leading to only a single interpretation in memory.

The use of scalar inferences addresses both of these concerns. First, there is no linguistic cue to disambiguate the final interpretation. For example, sentences like (6) are globally ambiguous between the semantic meaning in (7a) and the pragmatic meaning in (7b). Second, a rich context is necessary to generate the inference. If the language processing system typically relies on context to override or avoid initially incorrect interpretations, the semantic meaning of "some" should not persist in memory.

(6) "The girl ate some of the cookies"

(7) a. Semantic: The girl ate some and possibly all of the cookies

b. Pragmatic: The girl ate some but not all of the cookies

1.6 Outline of dissertation

This dissertation features three experiments that explore the time-course and mechanisms supporting the encoding of a sentence's interpretation (or interpretations) in memory. The goal of Chapter 2 is to examine the extent to which initial interpretations are encoded in memory. Specifically, it asks whether the semantic meaning of "some" is maintained in memory following a pragmatic inference. To answer this question, I used a task in which participants must generate a pragmatic inference ("some") or use exact semantics ("all/two/three") to learn the meaning of a novel word. Later, participants were asked to recall the novel word. If it is the case that the semantic meaning of “some” is interpreted and maintained in memory, then participants may recall the wrong novel object. Thus, recall of the novel words that were learned by pragmatic inference should be lower than recall for words learned by exact semantics. However, if the pragmatic inference overrides the semantic meaning of “some” and only one interpretation is encoded in memory, recall for the novel words should be similar regardless of which quantifier was featured during word learning.

The goal of Chapter 3 is to investigate the possible role that visual attention plays in the paradigm used in Chapter 2. Specifically, it asks whether there is still memory interference when “some” is used but no pragmatic inference is made. To answer this question, “some” was presented in a downward-entailing context, which blocks the pragmatic inference. Cues were used to ensure that the same amount of time was spent inspecting the novel objects as in Experiment 1. If simply attending to the objects during the word-learning task is enough to encode their visual representation in memory, then recall of the novel words learned using “some” should be lower than recall of words learned using other quantifiers. However, if attention to the objects does not encode them in memory then recall should be similar across all quantifiers.

The findings from Experiments 1 and 2 raise two possibilities: first, the memory representation may have been of poor quality because both interpretations were available during encoding, or the semantic meaning was computed and encoded first and lingered even after the pragmatic interpretation was computed and encoded. Chapter 4 explores the extent to which both semantic and pragmatic interpretations are available prior to the adoption of the pragmatic inference. Experiment 3 tests for the presence of competition between interpretations prior to the generation of a pragmatic inference by engaging cognitive control, the mechanism responsible for selecting between competing representations. Prior work on cognitive control has shown that engagement of this mechanism occurs when multiple syntactic representations are in conflict (Novick, Trueswell, & Thompson-Schill, 2005, 2010; January, Thompson-Schill, Trueswell, 2009; Ye & Zhou, 2009; Hsu & Novick, 2016). If accessing the word "some" activates both the semantic and pragmatic meanings of the word, there should be evidence of competition prior to the adoption of the pragmatic inference. However, if semantic analysis precedes the inference and only one interpretation is available prior to the pragmatic inference, then there should be no evidence of competition during this time.

In the final chapter, I compare the findings of the experiments included in this dissertation. I discuss the results within the context of current work on scalar inferences. Finally, I examine the broader implications of this research for current models of language processing and outline future directions.

Chapter 2

2.1 Overview

In this chapter, I report the first of three experiments that investigate the nature of sentence interpretation encoding using scalar inferences. Scalar inferences are used to see whether multiple interpretations persist in memory even when there is contextual support for one over the other. First, I discuss where scalar inferences lie on either side of Grice's distinction between Generalized and Particularized Conversational Implicatures (Grice, 1975). Then I summarize contrasting theories of how scalar inferences are calculated in real-time and discuss prior work investigating scalar inference processing (Noveck & Posada, 2003; Bott & Noveck, 2004; Breheny et al., 2006; De Neys & Schaeken, 2007; Huang & Snedeker, 2009, 2011, in press; Grodner, Klein, Carbary & Tanenhaus, 2010; Degen & Tanenhaus, 2015). In addition, I describe an experimental paradigm that provides an unbiased measure of memory recall. Finally, I describe an experiment that uses the unbiased memory paradigm to study scalar inferences.

2.2 Scalar inferences

Listeners must often infer an intended meaning that goes beyond what a speaker explicitly states. Often, the inferred meaning is triggered by context. Grice (1975) drew a distinction between inferences that arise based on the specific features of the context and those that arise unless they are blocked by context. Take, for example, the question and answer in (8):

(8) Speaker A: Did Alix submit her dissertation?

Speaker B: She wrote some of the chapters.

a. Alix did not submit her dissertation

b. Alix did not write all of the chapters

One inference that could be drawn from this short dialogue is that Alix has not yet submitted her dissertation. Grice called this type of inference a Particularized Conversational Implicature because it relies heavily on the context and the question under discussion. For example, if the response in (8) had been to the question of whether Alix wrote Chapter 2, the inference that she did not submit her dissertation would not have arisen. A second inference that could be drawn from this dialogue is that Alix did not write all of the chapters. Grice characterized this type of inference as a Generalized Conversational Implicature, which is thought to arise relatively independently of the context. The inference in (8b) is an example of a scalar inference.

A scalar inference arises when the use of an informationally weaker term like “some” is inferred to mean the negation of an informationally stronger term like “all.” Instead of the response in (8) the second speaker could have said “Alix wrote all of the chapters,” which would have been more informative. “All” is an informationally stronger scalar term than “some” because “some” can be used in any situation where “all” is true, but “all” is not true in all cases where “some” is true. Because the listener assumes that the speaker is a cooperative interlocutor by being as informative as possible (the Competence Assumption, Sauerland, 2004; van Rooij & Schulz, 2004), the listener infers that Alix wrote some but not all of the chapters (8b). The pragmatic interpretation of "some" is sometimes referred to as an "upper-bounded" reading because it excludes referents compatible with the stronger term (all). In contrast, the semantic meaning of "some" has a "lower-bounded" reading that includes the stronger term.

2.3 Accounts of scalar inference processing

There are opposing accounts of how and when comprehenders compute scalar inferences as an utterance unfolds. One set of accounts describe inferences as immediate, either because the pragmatic interpretation is the default interpretation (Levinson, 2000; Chierchia, 2004), or because the inference is rapidly calculated based on current experience (Sperber & Wilson, 1995; Degen & Tanenhaus, 2015). Other accounts assume that computing the scalar inference takes time and requires initial access to the semantic meaning (Huang & Snedeker, 2009, 2011, in press).

Early work on this subject supports the latter account, based on response times during sentence judgment tasks (Rips, 1975; Noveck & Posada, 2003; Bott & Noveck, 2004; De Neys & Schaeken, 2007). For example, when presented with under-informative statements like "Some elephants are mammals," readers were slower to respond when they evaluated the statement as "False" (consistent with the pragmatic meaning of "some") compared to "True" (consistent with the semantic meaning of "some") (Bott & Noveck, 2004). Longer response times for "False" statements suggest that computing the scalar inference takes more processing time than accessing the semantic meaning of "some." Eye-tracking studies suggest that this extra processing time is required to access the semantic meaning before the pragmatic inference (Huang & Snedeker, 2009, 2011, in press). Huang and Snedeker (2009) presented listeners with sentences like "Point to the girl that has all/some of the socks/soccer balls," while their eye-movements were measured to a total-set of soccer balls and a subset of socks. After the onset of "all," listeners quickly looked to the total-set. However, after the onset of "some," they were equally likely to look at the subset and total-set. These findings support a theory of scalar inference processing where the semantic meaning of "some" is retrieved before the inference is adopted.

In contrast, Default accounts of scalar inferencing assume that scalar inferences are computed automatically by a default mechanism (Levinson, 2000) or generated by semantic rules (Chierchia, 2004). According to Levinson (2000), the pragmatic meaning of "some" is stored in the lexicon and can be immediately and effortlessly accessed. Evidence that the pragmatic inference is generated rapidly has come from visual world experiments (Grodner et al., 2010; Breheny, Ferguson, & Katsos, 2013; Degen & Tanenhaus, 2015). For example, Grodner et al. (2010) presented listeners with sentences like "Click on the girl who has summa the balls" while their eye movements were measured to a total-set of balloons and a subset of balls. Participants looked to the subset compatible with the pragmatic meaning of “some” and avoided a competitor compatible with the semantic meaning of “some” following the onset of “summa.” Critically, looks to the target were as fast for “summa” as for the quantifiers “nunna” and “alla.” This suggests that the scalar inference was automatically triggered when participants encountered “summa.”

Evidence for the Default model also comes from a text comprehension experiment featuring number terms by Bezuidenhout and Cutting (2002). Many linguists have suggested that numbers have lower-bounded semantics (“two” means at least two) and upper-bounded interpretations (“two” means exactly two) arise through scalar inferences (Horn, 1972, 1989; Gadzar, 1979; Levinson, 2000). In Bezuidenhout and Cutting (2002), number terms were presented either upper- or lower-bound contexts, and the time taken to read the terms was longer in the lower-bound context. This suggests that the pragmatic inference was made by default and cancelling this inference took extra processing time.

Degen and Tanenhaus (2015) discuss a constraint-based account of scalar inferencing that tries to account for the contradicting evidence from visual world paradigms. In this account, the more probabilistic support from cues in the sentence means it is more likely that the inference will be made. If there is less support, listeners will take longer to arrive at the inference (Degen & Tanenhaus, 2015). They argue that the inference was delayed in Huang and Snedeker’s experiment because number terms were included in their stimuli. Number terms are more natural labels for set sizes (Grodner et al., 2010; Degen & Tanenhaus, 2011; cf. Huang & Snedeker, in press) and the use of “some” instead of a number may have led to the delay in processing. Experimental support for this theory comes from evidence that the use of "some" is less natural for some set sizes than others. For example, when presented with a gumball machine with an upper and lower chamber, listeners were less likely to generate the scalar inference when the number of gumballs in the lower chamber was a small set (1-3 of 13 gumballs) compared to an intermediate set (6-8 gumballs). The Default account would not predict any difference in naturalness for different set sizes.

Further support for a context-based account comes from Breheny and colleagues, who replicated the text comprehension experiment by Bezuidenhout and Cutting (2002) using the conjunction "or." A statement like "Mary dated John or Bill" is true even if Mary dated John *and* Bill, but the implication is that Mary did not date both. Contrary to Bezuidenhout and Cutting (2002), Breheny and colleagues found that reading times of the region including "or" were longer in contexts that supported the inference versus contexts that supported the literal interpretation. Longer reading times at this point suggest that the inference was only drawn in the inference-supporting context. They concluded that scalar inferences are not generated by default, but only when the context warrants them. Bergen and Grodner (2012) came to a similar conclusion when they tested the impact of speaker knowledge on inference computation. Regions of sentences including "some" were read more slowly when a preceding context established that the speaker was likely to know whether the stronger alternative ("all") was true versus when it was only possible that the speaker knew whether it was true. This suggests that the inference was drawn when the reader thought the speaker was fully knowledgeable, further supporting the theory that scalar inferences are context-dependent.

2.4 An unbiased measure of memory recall

Importantly, prior work leaves open the question of whether the semantic meaning of "some" is encoded in memory before a pragmatic inference is made. One challenge to investigating what is encoded in memory is that the tasks commonly used to probe interpretations may themselves introduce bias for meanings. Recall that responses to verification questions and picture-matching tasks suggest that interpretations based on initial and revised syntactic analyses are maintained in memory (Christianson et al., 2001; Malyutina & den Ouden, 2016). However, it remains unclear whether readers encoded both interpretations in memory, or whether the presence of verification questions and pictures themselves generate a bias to recall a misinterpretation. For example, asking a comprehension question may re-activate the interpretation that the question is about, even if it had been successfully revised (Tabor, Galantucci, & Richardson, 2004).

In contrast, Experiment 1 indirectly probes interpretation and memory by using a word-learning and recall task. Words link sound to meaning: when sounds are perceived, they activate corresponding meanings in memory. These sound-to-meaning mappings can be learned by using cues within sentences (Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1990; Naigles, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Snedeker & Gleitman, 2004). These facts are useful for our present purposes because listeners could learn a novel word by generating an interpretation based on a scalar inference. Then, this interpretation can be probed by asking the listeners to recall that novel word.

In recent work on syntactic parsing, Huang and Arnold (2016) examined how children's learning and memory of novel-word meanings varied with the need to revise. During each trial of a word-learning task, adults and 5-year-olds heard a novel word as the first or second noun embedded in a passive sentence (e.g. "The blicket/seal will be eaten by the seal/blicket") or active sentence (e.g., "The blicket/seal ate the seal/blicket"), and were asked to select the object that corresponded to that noun. Following the word-learning task, participants were asked again for the referent of the novel word (e.g., "Which one is the blicket?"). During the word-learning task when the novel word was first in the sentence, eye movements suggest that both adults and children initially considered an interpretation where the novel word was the agent of the action. After encountering the passive marker ("-en by"), this interpretation was revised. When syntactic revision was required, both adults and children took longer to disambiguate the sentence. Importantly, even when children *correctly* mapped word meanings during the word-learning task (e.g., selected "blicket" to refer to the thing eaten), they often misremembered the meaning during the recall task (e.g., choosing the eater). This suggests that, during recall, initially incorrect interpretations interfered with the retrieval of the novel word meanings. Moreover, adults who incorrectly responded to the word-learning task sill correctly selected the Target during the recall task 57% of the time. This suggests that the sentences were partially revised and the correct syntactic representation persisted in memory to be accessed at recall. These findings demonstrate that the word-learning and recall task can be used to probe participants' memory for the presence of multiple interpretations.

Experiment 1 extends this task to scalar inferences in adults. During the word-learning task, participants were given instructions like "Click on the girl that has some/all/two/three of the blickets" while their eye-movements were recorded to a display depicting novel objects divided among characters (Figure 4). Similar to Huang and Snedeker (2009), one critical character had a subset of objects (subset), and a gender-matched character had a total-set of objects (total-set). Thus, both girls are consistent with the semantic meaning of "*some*" but only the subset is consistent with the pragmatic meaning of "some." Based on prior research (Huang & Snedeker, 2009, 2011, in press), fixations following the onset of "some" are expected to be split between the subset and total-set. Moreover, based on prior research (Rips 1975; Noveck & Posada, 2003; Bott & Noveck, 2004; Breheny et al., 2006; De Neys & Schaeken, 2007; Huang & Snedeker, 2009, 2011), participants should make the scalar inference and select the subset.

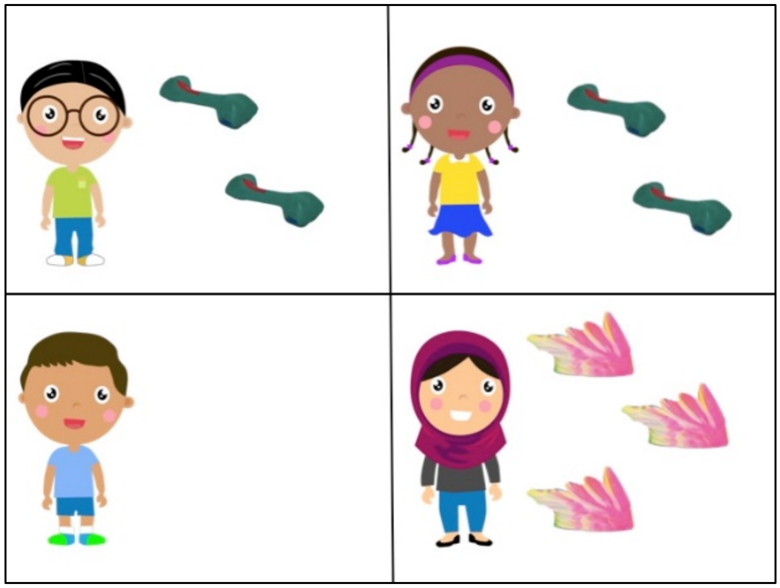


Figure 4. Sample display from the word-learning task. The girl in upper-right quadrant has a subset of one object and the girl in the lower-right quadrant has all of a second object.

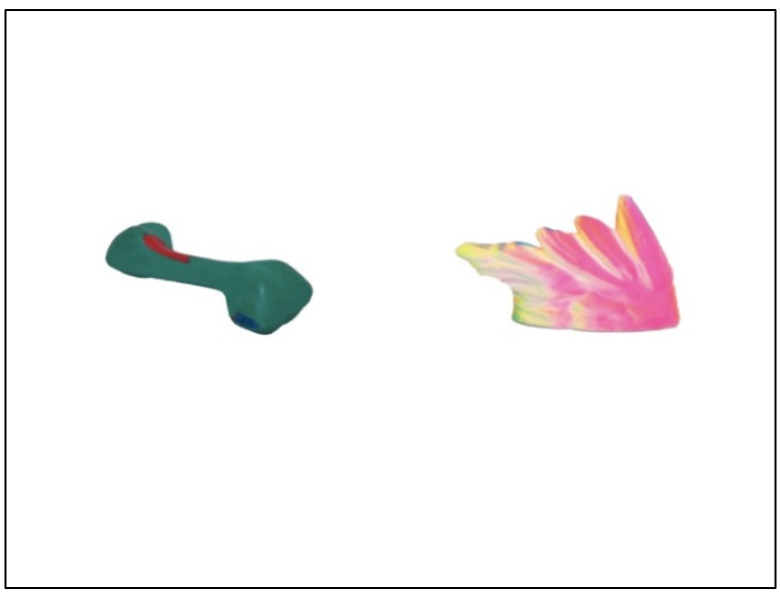


Figure 5. Sample display from the recall task. The subset object is on the left and the total-set object is on the right.

Nevertheless, it remains unknown the extent to which the semantic meaning is encoded in memory before the pragmatic inference is made. To address this, participants were also presented with a separate recall task, after the completion of the word-learning task. In the recall task, they were presented with the same novel items and were asked to, "Click on the blicket" (Figure 5). If listeners generate an interpretation of the utterance using the semantic meaning of "some", the novel word could be mapped to either the subset or total-set object. However, if listeners generate an interpretation of the utterance using the pragmatic meaning of "some", the novel word would be mapped exclusively to the subset object. Assessing the degree to which object selection in the word-learning task matches object selection in the recall task can test whether semantic meaning persists in memory even after a pragmatic inference in generated. If memory encoding relies on the resolution of competition at lower levels of analysis, then only the interpretation based on the pragmatic meaning of "some" will be encoded during word-learning trials, leading to a high recall rate. However, if representations are encoded incrementally, interpretations based on the pragmatic *as well as* semantic meanings will be encoded in memory, leading to interference and a low recall rate.

2.5 Experiment 1

2.5.1 Participants

Forty undergraduate students from the University of Maryland were recruited for this study. They received either $5 or course credit for their participation. All participants were native English speakers.

2.5.2 Procedures

The study involved two tasks, presented in fixed order. Participants completed a word-learning task followed by a recall task.

Word-learning task. Participants sat in front of a computer monitor, and their eye-movements were measured to the display using an Eyelink 1000 desktop eye-tracker (SR Research). Prior to the experiment, participants were told that they would see a display of four children with strange objects, and they should pay attention to these objects' names. Each trial featured two phases. During the familiarization phase, participants were presented with a display of four children. Objects appeared next to the boy and girl in top two quadrants, followed by the boy and girl in the bottom two quadrants. In the test phase, participants heard the critical instructions to click on one of the children, as in (9). When the participant did this, the trial ended and the next trial began.

(9) "Click on the girl that has <u>some/two/all/three</u> of the blickets."

Recall task. After completing 16 trials in the word-learning task, the recall task began. On each trial, participants were presented with two objects from the word-learning task and asked to select the object corresponding to the novel word (e.g., “Click on the blicket”). Once the participant did this, the trial ended and new novel objects were displayed. Participants were asked to recall all 16 of the novel words.

2.5.3 Materials

Word-learning task. Across all trials, the four children appeared on the display in the same configuration. The vertically adjacent children matched in gender while the horizontally adjacent children did not. A set of four novel objects (subset object) were split between a horizontally adjacent boy and girl and a set of three novel objects (total-set object) were given to one of the remaining children. This ensured that among the critical characters, one had a subset and the other had a total-set.

The instructions featured four quantifiers that constitute four conditions varying along two factors. In addition to "some," the scalar term "all" was included to establish a baseline for the time it takes to show a preference for the Target when the subset object is not consistent with the semantic meaning of the quantifier. In addition to the scalar terms, two terms from the number scale ("two" and "three") were included to ensure that differences between scalar trials was not due to preference to look at larger quantities. Like "all,” number terms do not require an inference to specify quantities. Importantly, the meaning of “two” rules out the same competitor as “some” when the inference is made. Comparing these trials will reveal whether there is a temporal difference between reference restriction via semantic meaning and reference restriction via pragmatic inference. The first factor, quantifier type, contrasts terms from the Gricean scale ("some" and "all") with terms from the number scale ("two" and "three"). The second factor, quantifier strength, contrasts weaker quantifiers ("some" and "two") with stronger ones ("all" and "three"). These four terms were paired with four target instructions, as in (9).

The target instructions within each condition were identical except for the gender of the target child and the novel word. The gender of the requested child was linked to the display (e.g., if the set of three objects is given to a girl, then a girl was requested). The requested child was labeled as the Target and their gender-matched counterpart was the Distractor. Thus, in the "some/two" trials, the Target was the subset character and the Distractor was the total-set character. In the "all/three" trials, the Target was the total-set character and the Distractor was the subset character. Across 16 trials, there were 32 novel objects: 16 presented as part of a subset and 16 presented as part of a total-set. Four versions of each instruction were used to create four presentation lists such that each list contained four instructions in each condition. Each pair of objects appeared just once in each list for a total of 16 items per list. These word-learning trials were randomly displayed.

Recall task. The recall task featured 16 trials. Within each trial, a display depicted a pair of novel objects from a trial in the word-learning task, arranged side-by-side. The trials in the recall task were randomly displayed.

2.5.4 Results

Eye-movements were examined to determine whether there were online processing differences between the four quantifier conditions. Based on prior research, preference for the Target referent was expected to occur later in the "some" trials compared to "all," "two," and "three" trials (Huang & Snedeker, 2009, 2011, in press). The proportion of fixations to the Target was calculated for two regions over the duration of the quantifier and up to the onset of the novel word:

> Early quantifier region: This region starts at the onset of the quantifier and continues for 650ms. Looks to the Target should increase during this region in the conditions where only the Target object is consistent with the quantifier ("two/all/three"). Because generating a pragmatic inference is preceded by semantic analysis, looks should remain equivocal between the Target and Distractor in the "some" condition.
>
> Late quantifier region: This region starts 650ms after the onset of the quantifier and continues for 650ms, ending after "*the*" (e.g., "…some/two/all/three of the…"). Looks to the Target should increase during this region in all conditions.

For each time window, the primary dependent measure examined Target preference, calculated by taking the ratio of looks to the Target over looks to the Target and Distractor (see Huang & Snedeker (2009, 2011) for similar approaches). Looks to the other characters were rare following the onset of the gender cue, accounting for 6% of all time points. For this reason, these fixations were not included in the analysis. Thus, values ranged from 0 to 1 where 1 indicates exclusive looks to the Target and 0 indicates exclusive looks to the Distractor. Target preference was analyzed in a linear mixed-effects model using the lme4 software package in R (Bates, Maechler, Bolker, & Walker, 2014). Quantifier type (scalar vs. number) and quantifier strength (weak vs. strong) were included as fixed effects. Models were created with the “maximal” random-effects structure for both subjects and items (Barr, Levy, Scheepers, & Tily, 2013).

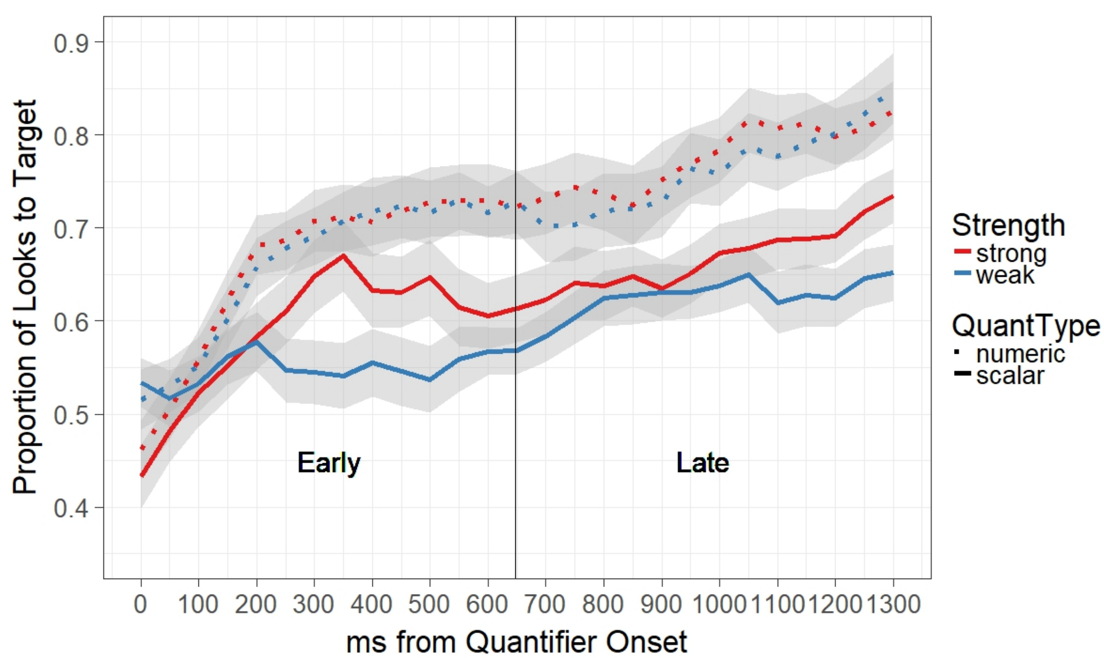


Figure 6. In Experiment 1, Target preference fixations during the Early and Late quantifier regions within the instruction. Error bars represent standard errors.

	Estimate	Std. Error	t-value
Intercept	0.0658	0.0031	20.97
Quantifier Type	-0.0072	0.0031	-2.17
Quantifier Strength	-0.0030	0.0016	-2.10
Quantifier Type x Quantifier Strength	-0.0037	0.0017	-2.08

Table 1. Parameter values of linear effects model in Early quantifier region of the word learning instructions, Experiment 1. R syntax:

lmer(TargProp ~ 1 + QuantType * Strength + (1 + QuantType * Strength|Subject) + (1 + QuantType * Strength|Item), data = early, REML=FALSE)

	Estimate	Std. Error	t-value
Intercept	0.7507	0.0352	21.33
Quantifier Type	-0.002	0.0027	-0.06
Quantifier Strength	-0.0276	0.0487	-0.57
Quantifier Type x Quantifier Strength	0.0101	0.0812	0.12

Table 2. Parameter values of linear effects model in Late quantifier region of the word learning instructions, Experiment 1. R syntax:

lmer(TargProp ~ 1 + QuantType * Strength + (1 + QuantType * Strength|Subject) + (1 + QuantType * Strength|Item), data = late, REML=FALSE)

Figure 6 illustrates that in the Early quantifier region, there was a main effect of quantifier type ($F(1) = 8.05$, $p < 0.01$). There was no main effect of quantifier strength ($p = 0.44$). Critically, looks to the Target during the "all" (59%), "two" (66%), and "three" (66%) trials increased while looks to the Target during the "some" trials remained around chance (55%), leading to an interaction between quantifier type and quantifier strength ($F(1) = 4.30$, $p < 0.05$). Planned comparisons revealed, within the scalar terms, there was no significant difference between Agent preference during "some" versus "all" trials ($p = 0.28$). Within the numeric terms, there was no significant difference between "two" and "three" trials ($p = 0.93$). Within the weak terms, analysis revealed that Agent preference during "two" trials was significantly greater than during "some" trials ($t(15) = 3.42$, $p < 0.001$). Within the strong terms, Agent preference during "three" trials was significantly greater than during "all" trials ($t(15) = 3.08$, $p < 0.001$). This is likely because the exact semantics of number words isolates the domain of quantification to the basic level, and generates a clear expectation that the up-coming novel word will distinguish the objects. The quantifier terms refer to relationships between individuals within a set, so listeners might entertain the possibility that the novel word is a superordinate category that refers to both object kinds. This may have delayed processing in the "all" compared to the "three" case.

In the Late quantifier region, the main effect of quantifier type ($F(1) = 12.72$, $p < 0.001$) continued. However, looks to the Target increased in all conditions (some: 61%, all: 64%, two: 72%, three: 75%), and no main effect of quantifier strength ($p = 0.55$) and no interaction between quantifier type and quantifier strength was found ($p = 0.90$). This suggests that the pragmatic inference was generated during the Late quantifier region. This confirms that reference restriction based on a scalar implicature takes longer than restriction via semantic analysis (Huang & Snedeker, 2009, 2011, in press).

Participants' actions were coded for accuracy on the word-learning task. A response was coded as accurate if the participant selected the Target. A response was coded as inaccurate if the participant selected any non-Target. Accuracy on the word-learning task was high across all conditions, shown in Figure 7. Participants only selected either the Target or Distractor, suggesting they correctly interpreted the gender cue and constrained their interpretation appropriately. Because the participants were selecting one of two characters/objects, the likelihood of correct responses was compared to 50%. These analyses confirmed that accuracy was above chance for "some" trials (84%; $t(15) = 11.60$, $p < 0.001$), "all" trials (96%; $t(15) = 27.60$, $p < 0.001$), and "two" trials (99%; $t(15) = 55.82$, $p < 0.001$). Accuracy for "three" trials was 100%. This confirms that participants had no trouble identifying the correct character based on the quantifier they heard. Also, participants were generating the inference on most of the "some" trials and selected the subset/Target accordingly.

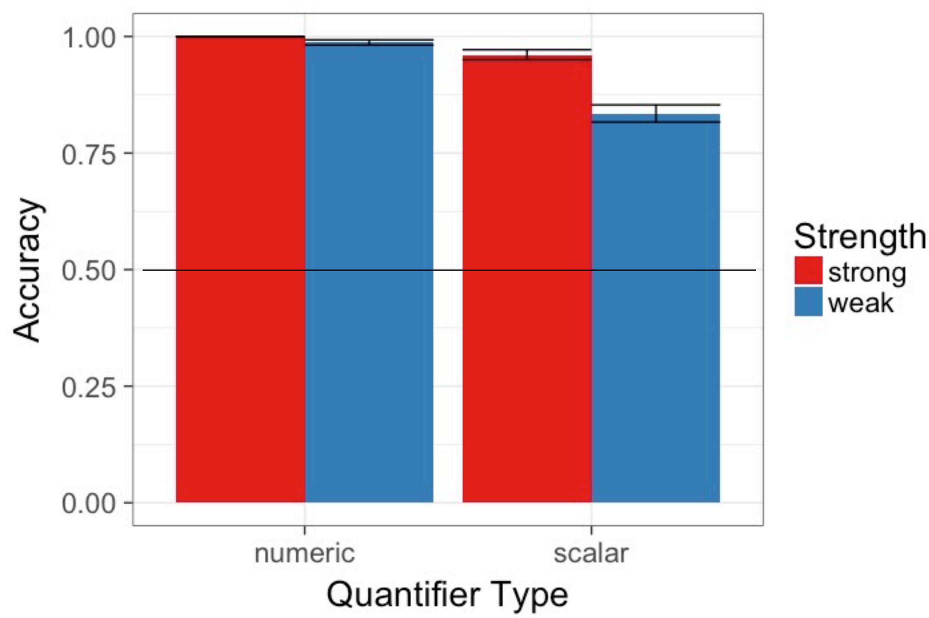


Figure 7. Accuracy on the word-learning task, Experiment 1. Error bars represent standard errors.

The dependent measure of the recall task was the proportion of matches between responses on the word-learning and recall tasks. Only the matches for accurate word-learning trials were analyzed, to ensure that interpretations on “some” trials were made via pragmatic inference. A response was coded as a match if the same object was selected in both word learning and recall. A response was coded as not a match if a different object was selected. The proportion of matches was calculated within each condition as the number of matches over the number of trials. Overall, there were 71% matches between responses on corresponding word-learning and recall trials. Figure 8 illustrates the proportion of matches per condition.

Match proportion was analyzed in a mixed multinomial logistic model using the lme4 software package in R (Bates, Maechler, Bolker, & Walker, 2014). Quantifier type (scalar vs. number) and quantifier strength (weak vs. strong) were included as fixed effects. Models were created with the "maximal" random-effects structure for both subjects and items (Barr, Levy, Scheepers, & Tily, 2013). However, when the model failed to converge within 50,000 iterations, the model was estimated without random intercepts. Analyses revealed no significant main effect of quantifier type ($p = 0.90$) or quantifier strength ($p = 0.66$). Critically, there was a significant interaction between quantifier type and quantifier strength ($F(1) = 3.73$, $p < 0.05$). Planned comparisons revealed that, within the scalar terms, the proportion of matches for "some" trials was significantly lower than "all" trials ($t(15) = 2.63$, $p < 0.05$). Within the numeric terms, the proportion of matches was not significantly different between "two" and "three" trials ($p > 0.80$). Within the weak terms, the proportion of matches for "some" trials was significantly lower than "two" trials ($t(15) = -3.19$, $p < 0.01$). Within the strong terms, the proportion of matches was not significantly different between the "all" and "three" trials ($p > 0.40$). The fact that matches for "some" trials were significantly lower than the other conditions suggests that the semantic meaning of "some" was encoded in memory during "some" trials and this led to interference during the recall task.

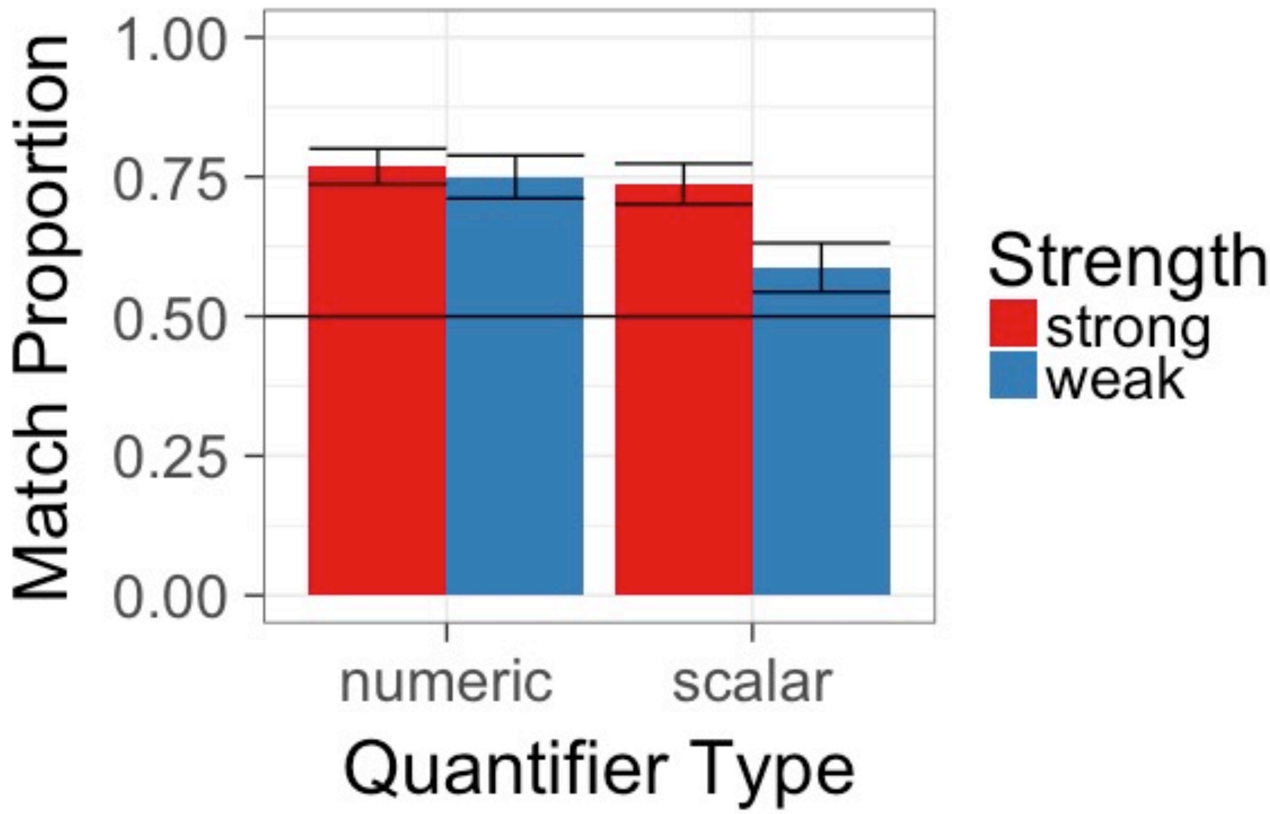


Figure 8. Proportion of matches between word-learning and recall responses in Experiment 1. Error bars represent standard errors.

	Estimate	Std. Error	z-value
Intercept	1.3602	0.2935	4.64
Quantifier Type	-0.0430	0.3301	-0.13
Quantifier Strength	-0.1655	0.3785	-0.44
Quantifier Type x Quantifier Strength	-0.1372	0.0682	-2.73

Table 3. Results of mixed multinomial logistic model of matches between responses to word-learning and recall tasks, Experiment 1.

glmer(match ~ 1 + QuantType * Strength + (0 + QuantType * Strength|Subject) + (0

+ QuantType * Strength|Item), data = matchdata, family = binomial)

Critically, the recall task allows one to link the processing of scalar inferences (measured by eye-movements during the critical instructions) and memory for their interpretations (measured by actions in the recall task). If longer consideration of the semantic meaning of “some” during the word-learning task lead to a stronger memory trace and greater interference at recall, then time spent looking to the Target character may predict recall accuracy. However, if only the pragmatic meaning for “some” is encoded in memory, then time spent looking to the Target character should not predict recall accuracy. A Pearson correlation was computed to assess the relationship between Target preference (during the word-learning instructions) and Match proportion on the recall task. This relationship is plotted in Figure 9. The analysis revealed no significant relationship between Target preference and Match proportion in the “all” ($p = 0.14$), “two” ($p = 0.76$), or “three” ($p = 0.74$) conditions. However, there was a significant correlation in the “some” condition ($r = 0.38$, $p < 0.05$). This suggests that longer consideration of the semantic meaning of “some” led to a stronger memory trace and greater interference at recall.

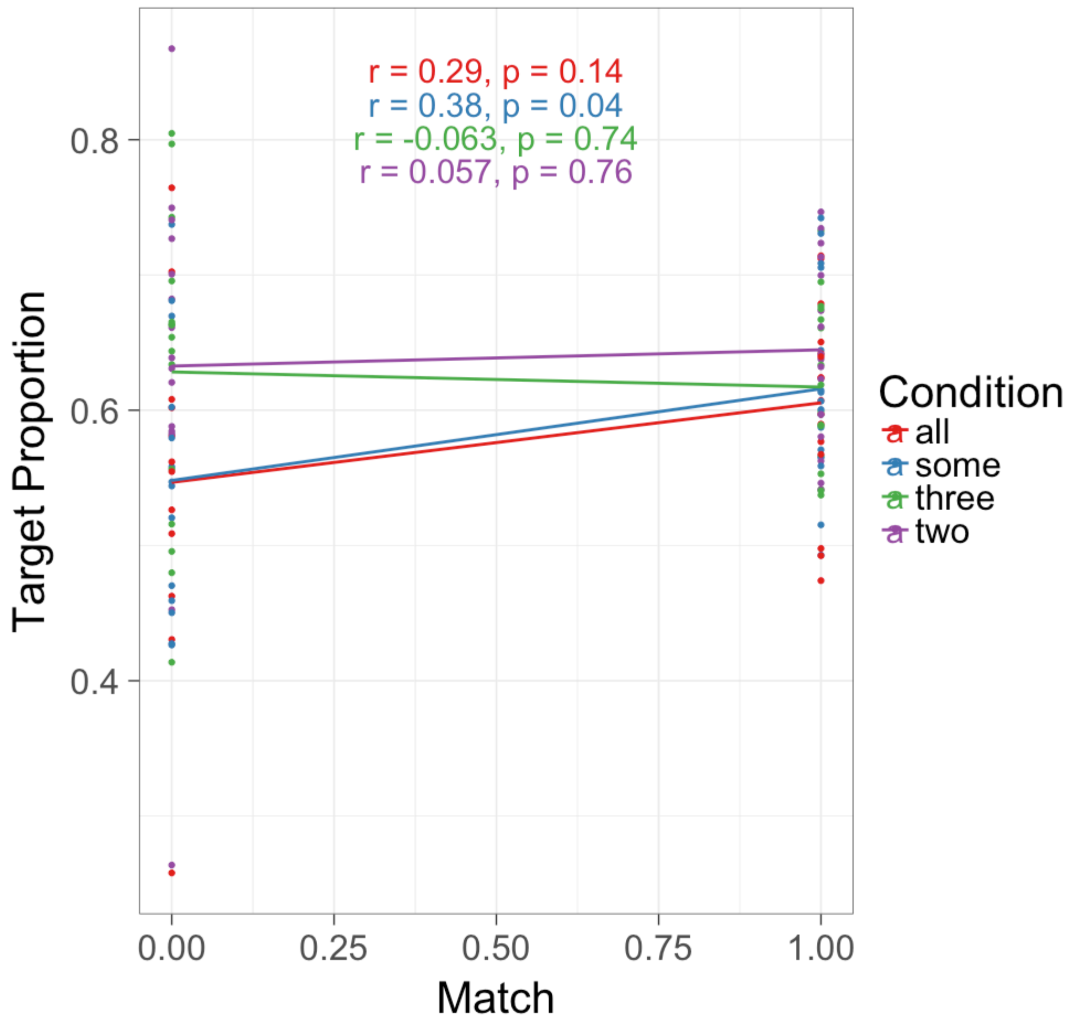


Figure 9. Scatterplot of Target fixations during word-learning and Matches on recall responses in Experiment 1.

2.6 Discussion

The goal of Experiment 1 was to test whether the semantic meaning of "some" is encoded in memory before a pragmatic inference is made. This addresses the extent to which lower-level processes must be completed before the final interpretation of a sentence is encoded in memory. These results suggest that when a scalar inference was generated during “some” trials, interpretations based on the semantic *and* pragmatic meanings were encoded in memory. This lead to interference during recall, resulting in lower match proportions for “some” trials compared to “two/all/three” trials. Moreover, during word learning, longer consideration of the semantic meaning (measured by looks to the Distractor) led to greater interference during recall. This suggests that initial interpretations are encoded in memory prior to the adoption of final interpretations.

These findings support a model of language processing where activated representations are incrementally encoded in memory, before competition between them has been resolved. The results are consistent with the Good Enough processing theory that claims that ambiguity between representations may never be resolved (Ferreira et al., 2001; Ferreira et al., 2002; Ferreira & Patson, 2007). Prior work in support of this theory has come largely from syntactic processing, but these data extends the model to pragmatic processing. This suggests that activated representations at all levels of processing, not just syntactic, may lead to multiple interpretations being encoded in memory.

However, alternative explanations for performance on the recall task must be considered. One possibility is that participants recalled the Distractor object more often for “some” trials because they spent a greater amount of time fixating on the Distractor during word learning, strengthening memory of that object. During word-learning trials featuring “two/all/three,” only the Target object was relevant because only one interpretation of each of these terms was available. However, during “some” trials, both the Target and Distractor objects were possible interpretations. Thus, memory interference may depend on the extent to which the Distractor object was attended to during word learning, whether or not an inference was made. Indeed, prior research in memory encoding provides strong evidence that attention during encoding directly influences recall performance (Baddeley, Lewis, Eldridge, & Thomson, 1984; Fisk & Schneider, 1984; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Mulligan, 1998), by creating more "durable" memory traces (Craik & Lockhart, 1972; Craik & Tulving, 1975) or higher quality representations (Awh & Jonides, 2001; Awh, Dhalival, Christensen, Matsukura, 2001). This alternative explanation will be addressed in the following chapter of this dissertation.

Chapter 3

3.1 Overview

Findings from Experiment 1 revealed that there were fewer matches in the "some" condition compared to "two/all/three" conditions. One possibility is that the semantic meaning of "some" persisted in memory and interfered with recall. This would suggest that lower level analyses do not need to be completed before the interpretation of a sentence is encoded in memory. However, another possibility is that the results were due to task demands. The Distractor object may have been stored in memory during "some" trials simply because it was attended to longer in that condition than the others, where semantic analysis quickly restricted the set of possible targets. Indeed, longer consideration of the Distractor object in the "some" condition led to fewer matches between word learning and recall.

In this chapter, I will flesh out the logic behind this alternative hypothesis by discussing prior work on visual attention and memory, and discuss how this relationship may have led to the results obtained in Experiment 1. Next, I describe an experiment that tests this possibility by replicating the conditions of Experiment 1 but without the inferencing aspect.

3.2 Visual attention and memory

In order to complete the word-learning task, participants had to map novel words to novel objects by interpreting a sentence that featured a scalar term. They did so by visually inspecting a display (reproduced in Figure 10). There was only one object type featured on the display that was consistent with the scalar terms “all,” “two,” or “three.” The results showed that, following the presentation of these scalar terms, visual attention was quickly restricted to the correct object type. However, when participants were presented with a sentence featuring “some,” they looked to both object types (consistent with the semantic meaning of “some”) before restricting their visual attention to only one object type (consistent with the pragmatic meaning of “some”). Thus, only one object type was visually attended to during “all,” “two,” and “three” trials while two object types were visually attended to during “some” trials. Later, recall of the novel word was worse in the “some” condition compared to the “all,” “two,” and “three” conditions. One possibility is that the worse performance in the “some” condition is the result of visual attention being paid to both object types rather than just one.

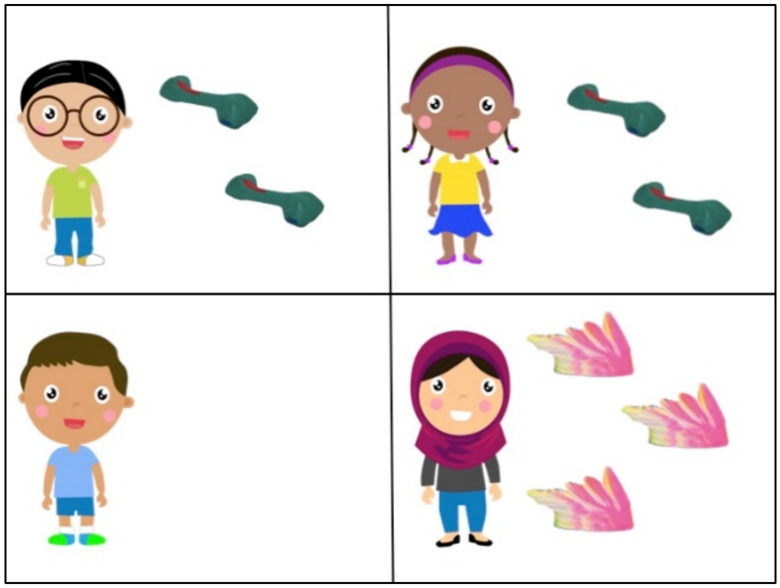


Figure 10. Sample display from the word-learning task. The girl in upper-right quadrant has a subset of one object and the girl in the lower-right quadrant has all of a second object.

Indeed, prior work exploring the relationship between attention and memory retrieval suggests that attention during encoding directly effects recall (Baddeley et al., 1984; Fisk & Schneider, 1984; Craik et al., 1996; Mulligan, 1998). For example, Craik et al. (1996) asked listeners to recall lists of words or word pairs as they completed a concurrent visual reaction time task. When the task was performed during encoding, divided attention led to poorer memory performance. In contrast, when the task was performed during recall, there was no reduction in memory. These results suggest that divided attention during encoding specifically hinders memory task performance. Moreover, Duncan (1984) found that when presented with two objects superimposed in space, observers were better able to report two aspects of one of the objects than one aspect from each of the two objects. The author concluded that visual processing is limited by the number of objects that require attention. Recall that in the “some” trials of the word-learning task, two objects were under consideration, compared to one object for the “all/two/three” trials. Thus, divided attention during encoding as well as the higher number of objects to be discriminated may have led to a weaker memory trace of the Target object.

Experiment 2 tests this possibility by replicating Experiment 1 with a context that blocks the pragmatic inference but retains the amount of attention paid to two objects during “some” trials. Because the inference is blocked, listeners should only consider the semantic meaning of “some” when they hear the instructions. This should lead to equivalent looks to the Target and Distractor objects because their quantities are both consistent with this meaning. Thus, if simply paying visual attention to the Distractor object in “some” trials led to poor recall, there should be a similar result when the same attention is paid to that object but the pragmatic interpretation is not considered. However, if the poor recall was due to multiple interpretations competing in memory, there should be no interference when only the semantic meaning of “some” is considered.

3.3 Downward-entailing contexts

In order to block the pragmatic inference, Experiment 2 presents "some" in a downward-entailing context. When one sentence (A) entails another (B) it means that for every situation in which A is true, B is also true. For example, "I ate pizza with anchovies" entails "I ate pizza" (Chierchia, Frazier, & Clifton, 2009). Upward-entailing contexts preserve the relation of semantic strength between a set of expressions. For example, the stimuli in Experiment 1 were presented in an upward-entailing context as in (10):

(10) a. Click on the girl that has some of the blickets.

b. Click on the girl that has all of the blickets

Here, (4a) is a weaker statement than (10b) because instructing someone to click on "some" of the blickets does not instruct them to click on "all" of them. However, listeners typically strengthen (10a) with a pragmatic inference and assume that "some" means "not all." In contrast, downward-entailing contexts like (11) block the pragmatic inference.

(11) a. If the girl has some of the blickets, click on her.

b. If the girl has all of the blickets, click on her.

In (11) the semantics of "all" is consistent with "some." That is, if the girl has all of the blickets then she must have some of them. This makes (11a) stronger than (11b) so a pragmatic inference is not supported. Prior research suggests that people are less likely to make scalar inferences in downward-entailing contexts like negation and conditional expressions (Chierchia, 2004; Chierchia et al., 2009; Schwarz, Clifton &

Frazier, 2010). For example, Chierchia et al. (2009) asked students to complete sentence fragments like “John has two cars…” or “If John has two cars…” Afterward, the students were asked to select one of the following sentences that fit how they interpreted the sentence: *John has two or more cars* (weak) or *John has exactly two cars* (strong). Eighty-nine percent of respondents who completed the simple sentence chose the strengthened interpretation of “two.” In contrast, only 64% of the respondents who completed the conditional sentence chose the strengthened interpretation. Thus, the downward-entailing conditional context led to less strengthening of “two” to mean “exactly two.”

Online measures have revealed a different processing time-course for scalar terms in downward- versus upward-entailment contexts (Panizza, Chierchia, & Clifton, 2009; Panizza, Huang, Chierchia, & Snedeker, 2009). Panizza, Chierchia, and Clifton (2009) found that numerals in upward-entailing contexts were read longer than in downward-entailing contexts. Also, during a following context that required the inference to be made, there were more regressions to the number term when it appeared in a downward- versus upward-entailing context. These findings suggest that the inference was not made when the number was encountered in the downward-entailing context, but it was generated when the following context supported it. These results complement a visual world experiment by Panizza, Huang, and colleagues. Using a paradigm similar to Huang and Snedeker (2009), the authors found that eye movements to a target associated with the pragmatic interpretation of “some” were equivocal prior to a disambiguating noun (Panizza, Huang, et al., 2009). This suggests that participants did not generate the inference to restrict reference to the target prior to disambiguation.

Thus, based on prior work (Chierchia, 2004; Chierchia et al., 2009; Schwarz et al., 2010; Panizza, Chierchia et al., 2009; Panizza, Huang et al., 2009), scalar inferences are not expected to be generated when “some” is featured in a downward-entailing context. Eye movements should show roughly equal looks to the Target and Distractor following the onset of “some” while looks to the Target should increase following “all/two/three.” However, because participants will be cued to ultimately select the Target, accuracy on the word-learning task should be high.

If increased attention to the Distractor object during word learning is enough to encode its visual representation in memory and cause interference during recall, accuracy should be lower for “some” trials compared to “two/all/three” trials on the recall task, replicating the results of Experiment 1. However, if attention to the Distractor object does not impact memory encoding, then recall rates should be similar across all trial types. Since only one interpretation is available for all trial types, memory interference is expected to be limited. Critically, looks to the Distractor should not predict recall rates if simply looking at the object does not commit it to memory.

3.4 Experiment 2

3.4.1 Participants

Forty undergraduate students from the University of Maryland were recruited for this study. They received either $5 or course credit for their participation. All participants were native English speakers.

3.4.2 Procedures and Materials

The procedures and materials were similar to Experiment 1 with two key changes. First, the same four quantifiers featured in Experiment 1 were presented in Experiment 2 in a downward-entailing context, as in (11). Second, the correct answers for word-learning trials were cued by a star appearing next to the Target object. The number of cued trials depended on the number of correct trials for each condition in Experiment 1. For example, Experiment 1 participants correctly selected the Target on 85% of "some" trials, so 85% of "some" trials were cued in Experiment 2. The delay between trial onset and cue presentation on a given trial in Experiment 2 was equivalent to the average reaction time for that trial in Experiment 1. Participants were told in advance that this cue indicates the correct answer and that it would appear on some trials and not others.

3.4.3 Results

As in Experiment 1, eye movements were examined to determine whether there were online processing differences between the four quantifier conditions. The proportion of fixations to the Target was calculated for two regions over the duration of the quantifier and up to the onset of the novel word:

Early quantifier region: This region starts at the onset of the quantifier and continues for 650ms. Looks to the Target should increase during this region in the conditions where only the Target object is consistent with the quantifier ("two/all/three"). Because generating a pragmatic inference is preceded by semantic analysis, looks should remain equivocal between the Target and Distractor in the "some" condition.

Late quantifier region: This region starts 650ms after the onset of the quantifier and continues for 650ms, ending after "*the*" (e.g., "…some/two/all/three of the…"). Looks to the Target should increase during this region in all conditions.

The same analytical strategy used in Experiment 1 was used in Experiment 2.

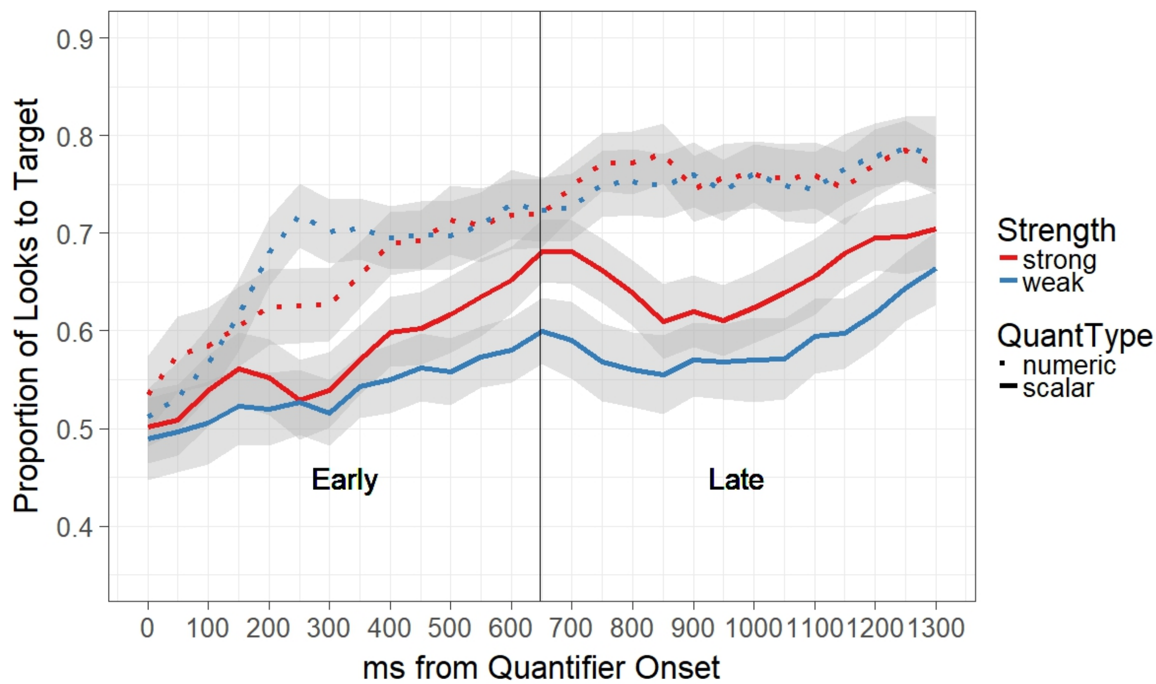


Figure 11. In Experiment 2, Target preference fixations during the Early and Late quantifier regions within the instruction.

	Estimate	Std. Error	t-value
Intercept	0.6447	0.0347	18.56
Quantifier Type	-0.0625	0.0460	-1.39
Quantifier Strength	0.0002	0.0383	-0.01
Quantifier Type x Quantifier Strength	-0.0378	0.0622	-0.61

Table 4. Parameter values of linear effects model in Early quantifier region of the word learning instructions, Experiment 2. R syntax:

lmer(TargProp ~ 1 + QuantType * Strength + (1 + QuantType * Strength|Subject) + (1 + QuantType * Strength|Item), data = early, REML=FALSE)

	Estimate	Std. Error	t-value
Intercept	0.	0.	

Quantifier Type	-0.	0.	
Quantifier Strength	-0.	0.	-
Quantifier Type x Quantifier Strength	0.	0.	0

Table 5. Parameter values of linear effects model in Late quantifier region of the word learning instructions, Experiment 2. R syntax:

lmer(TargProp ~ 1 + QuantType * Strength + (1 + QuantType * Strength|Subject) + (1 + QuantType * Strength|Item), data = late, REML=FALSE)

Figure 11 illustrates that in the Early quantifier region, looks to the Target were higher in the "two" (65%) and "three" (63%) trials than the "some" (53%) and "all" (55%) trials, leading to a main effect of quantifier type ($F(1) = 6.09$, $p < 0.05$). In contrast to Experiment 1, there was no main effect of quantifier strength ($p = 0.48$) or interaction between quantifier type and quantifier strength ($p = 0.55$) during this region. This is because Target looks during the "all" trials were not as high in Experiment 2 during this region (55%) as they were in Experiment 1 (59%), leading to a smaller difference between looks in the "all" versus "some" trials (4% in Exp. 1, 2% in Exp. 2). This delay in looks to the Target in the "all" condition may be because the use of a conditional introduces a third possibility that was not there in Experiment 1: that neither of the two girls has all of an object. In Experiment 1 participants heard "Click on the girl that has all of the blickets," making it clear that one of the girls was the Target. However, in Experiment 2 participants heard "If the girl has all of the blickets, click on her," introducing the possibility that neither of the girls have all of the blickets. This may have caused a prolonged uncertainty that extended the time during which participants were comparing the number of objects belonging to the characters. This did not happen in the "two" and "three" trials, where the exact number quickly identified one of the two girls. The eye movements in the "some" trials of Experiment 2 also pattern closely with Experiment 1, likely because the presence of objects next to each girl meant that they had to have "some" of something.

In the Late quantifier region, looks to the Target increased in the "all" (65%), "two" (75%), and "three" (75%) conditions while looks to the Target in the "some" (57%) condition did not increase as much, leading to a main effect of quantifier type ($F(1) = 15.67$, $p < 0.001$). There was no main effect of quantifier strength ($p = 0.54$) or interaction between quantifier type and quantifier strength ($p = 0.95$). The lack of an interaction between the variables was expected because the downward-entailing context blocked the scalar inference, making the total-set and subset objects equally likely to be the target. Participants' actions were coded for accuracy on the word-learning task. Figure 12 shows that accuracy on the word-learning task was high across all conditions. As in Experiment 1, participants only selected either the Target or Distractor, suggesting they correctly interpreted the gender cue and constrained their interpretation appropriately. Because the participants were selecting one of two characters/objects, the likelihood of correct responses was compared to 50%. These analyses confirmed that accuracy was above chance for "some" trials (84%; $t(15) = 11.60$, $p < 0.001$), "all" trials (96%; $t(15) = 27.60$, $p < 0.001$), and "two" trials (99%;

t(15) = 55.82, p < 0.001). Accuracy for “three” trials was 100%. This confirms that participants were paying attention to the cues and selected the Target accordingly.

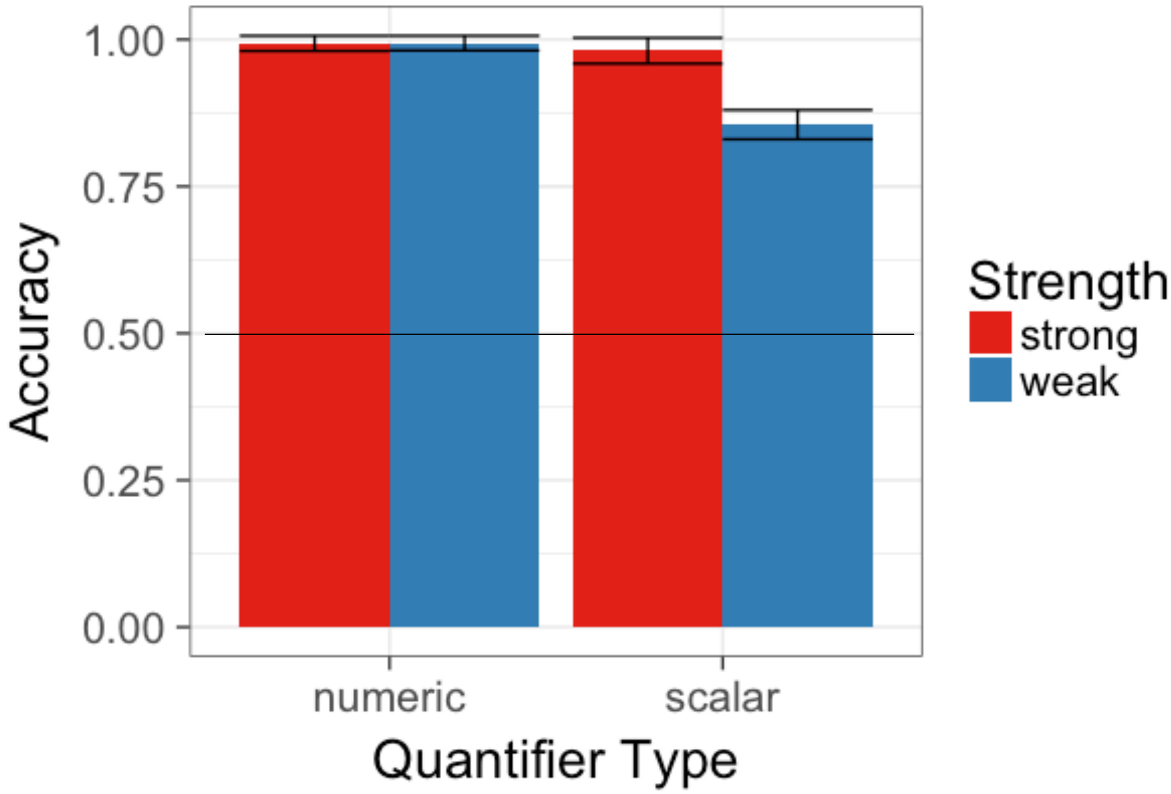


Figure 12. Accuracy on the word-learning task, Experiment 2. Error bars represent standard errors.

Actions during the recall task were analyzed using the same analytic strategy as Experiment 1. Again, the maximal model failed to converge with 50,000 iterations so the model was estimated without random intercepts. Overall, there were 65% matches between responses on corresponding word-learning and recall trials. Figure 13 illustrates the proportion of matches per condition (“some”: 61%; “two”: 70%;

“all”: 64%; “three”: 66%). The match results of Experiment 1 are included for comparison. Analyses revealed no main effect of quantifier type or strength (all p’s > 0.05).

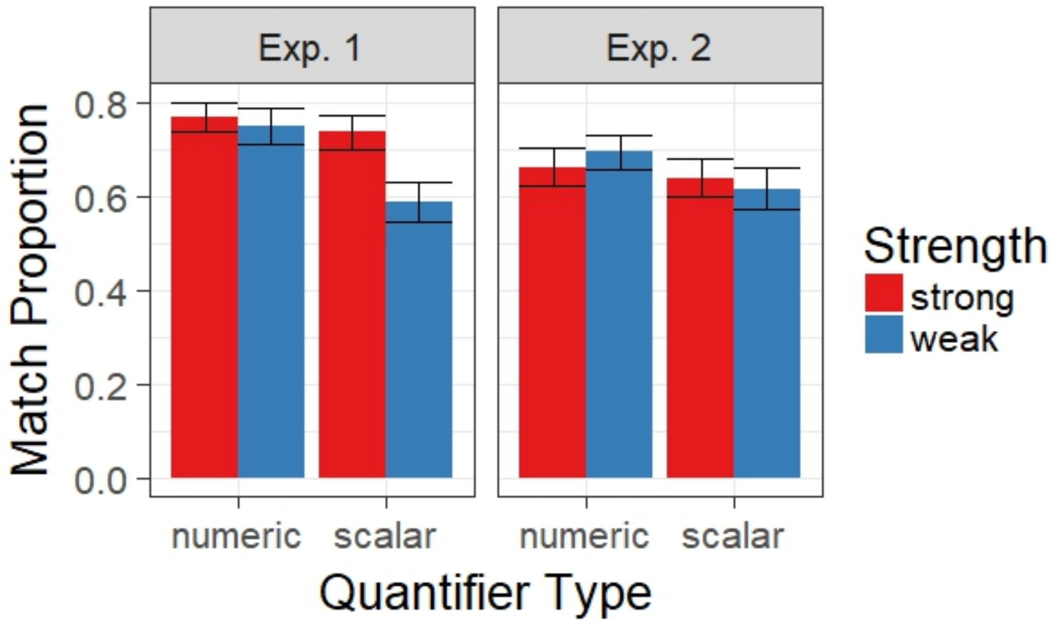


Figure 13. Proportion of matches between word learning and recall responses in Experiments 1 and 2.

	Estimate	Std. Error	z-value
Intercept	1.3367	0.2830	4.72
Quantifier Type	-0.0267	0.3321	-0.08
Quantifier Strength	-0.1453	0.3786	-0.38
Quantifier Type x Quantifier Strength	-0.7632	0.4325	-1.77

Table 6. Results of mixed multinomial logistic model of matches between responses to word-learning and recall tasks, Experiment 2.

glmer(match ~ 1 + QuantType * Strength + (0 + QuantType * Strength|Subject) + (0

+ QuantType * Strength|Item), data = matchdata, family = binomial)

The final analysis examined the relationship between eye movements and recall. If the visual representation of a Distractor object is encoded during word learning regardless of how the instructions are interpreted, a higher proportion of looks to this object should predict recall rates across all trial types. However, if the visual representation of a Distractor object does not interfere with recall, the proportion of looks to this object should not predict recall rates. As in Experiment 1, a Person correlation was computed to examine the relationship between Target preference (during the word-learning instructions) and Match proportion on the recall task. These analyses revealed no significant relationship between these variables for any of the conditions (“some”: $p = 0.92$, “all”: $p = 0.28$; “three”: 0.30; “two”: 0.06). This relationship is depicted in Figure 14.

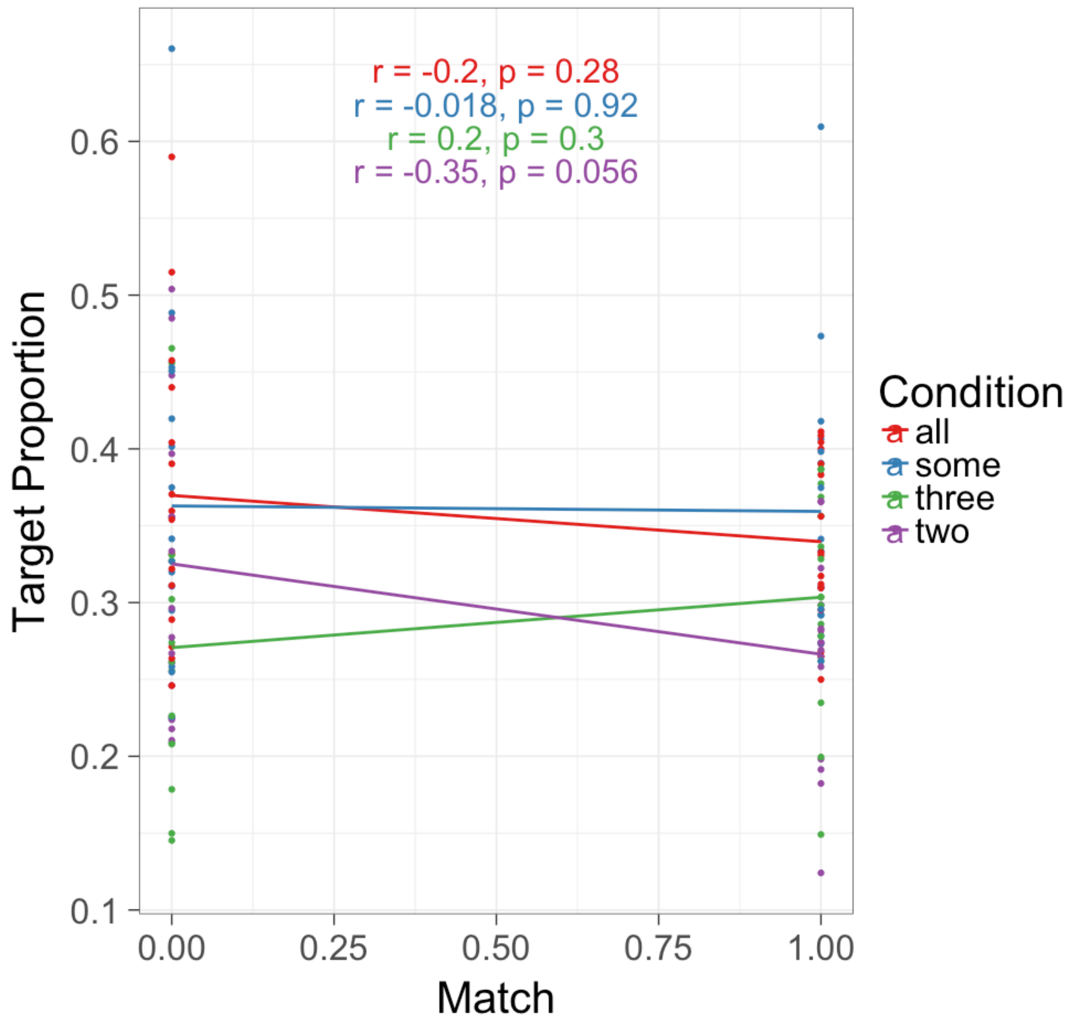


Figure 14. Scatterplot of Target fixations during word-learning and Matches on recall responses in Experiment 2.

3.5 Discussion

The goal of Experiment 2 was to evaluate the validity of the memory task in Experiment 1. Specifically, the experiment tested the extent to which visual attention to the Distractor object during word learning interfered with recall of the Target object. The results of Experiment 2 showed that when the same amount of attention was paid to the Distractor object but there was no scalar inference, recall of the Target object was similar across conditions. This suggests that visual attention to the Distractor was not enough to interfere with recall. Thus, the lower proportion of matches in the “some” condition of Experiment 1 were not due to greater visual attention to the Distractor object during the word-learning task in the “some” condition compared to the “all/two/three” conditions.

These results also address scalar inference processing in downward-entailing contexts. Prior work suggests that scalar inferences are not drawn as frequently in downward-entailing versus upward-entailing contexts (Chierchia, 2004; Chierchia et al., 2009; Schwarz et al., 2010). Eye-movement data during the word-learning instructions support this finding. Analysis showed that Target preference in the “some” condition was significantly below the other conditions even through the Late quantifier region, suggesting that the inference had not been made.

The combined results of Experiments 1 and 2 build on previous work investigating the persistence of multiple sentence interpretations in memory. In previous studies, responses to comprehension questions suggested that initial interpretations based on incorrect syntactic analysis can be accessed even after a sentence is complete (Christianson et al., 2001; Christianson et al., 2006; Ferreira et al., 2001; Patson et al., 2009; Christianson, 2008; Slattery et al., 2013; Malyutina & den Ouden, 2016). Moreover, prior work showed that these interpretations can interfere with the processing of a subsequent sentence (Slattery et al., 2013). The present work goes farther to show that an initial interpretation can be available well after sentence completion, even after hearing intervening sentences. Participants in Experiments 1 and 2 responded to memory probes after completing all 16 word-learning trials.

However, it remains unclear whether the semantic and pragmatic interpretations are both immediately accessed when “some” is encountered. The results of Experiments 1 and 2 are compatible with the possibility that the semantic meaning of “some” is computed and encoded in memory prior to the computation of the pragmatic interpretation. However, it is also possible that both interpretations are initially available and encoded in memory before the pragmatic interpretation is selected. Eye-movement analyses from Experiment 1 showed that listeners made an equal proportion of looks to the Target and Distractor prior to the pragmatic inference. This suggests that the semantic interpretation of “some” was available during this time, but it cannot address whether the pragmatic interpretation was also available since equal looks to the Target and Distractor is consistent with both interpretations. It could be the case that the semantic and pragmatic interpretations are both available prior to the inference and are in competition, similar to competing parses during syntactic processing. Alternatively, it could be the case that only the semantic meaning is available prior to the pragmatic inference, suggesting that there is some amount of semantic analysis that must be computed before the inference is drawn. The next chapter explores these possibilities.

Chapter 4

4.1 Overview

The findings from Experiments 1 and 2 showed that initial interpretations can persist in memory even when contextual information is provided. However, it is still unknown whether both the semantic and pragmatic meanings of “some” are available immediately after the word is encountered, or if only the semantic meaning is initially accessed. According to constraint-based accounts of scalar inferencing, listeners generate probabilistic expectations about possible speaker meanings as a sentence unfolds, which can include either the semantic or pragmatic meaning of “some” (Degen & Tanenhaus, 2015). This suggests that, like syntactic processing, these two meaning-based representations are active at the same time and in direct competition. If both meanings are accessed immediately after “some” is encountered, there should be evidence of competition prior to the adoption of the pragmatic inference. If there is a delay before the pragmatic interpretation is generated, there may be no evidence of competition. In the following sections, I discuss the relationship between cognitive control and competition between representations. Then, I talk about cognitive control’s potential role in pragmatic inferencing. Finally, I describe an experiment that engages cognitive control to probe for the presence of competition between interpretations prior to the generation of the pragmatic inference.

4.2 Cognitive control

One way to investigate whether both semantic and pragmatic meanings are activated immediately following “some” is to look for evidence of cognitive control. Cognitive control is a domain-general mechanism that detects and resolves conflict during information-processing tasks (Botvinick, Braver, Barch, Carter, & Cohen, 2001). It plays a specific role in language processing by facilitating recovery from misanalysis (Novick et al., 2005, 2010). Prior research suggests that cognitive control is triggered when an initial syntactic analysis is contradicted by incoming linguistic input, and inhibits this misanalysis in favor of a competing alternative (Novick et al., 2005, 2010; January et al., 2009; Ye & Zhou, 2009; Hsu & Novick, 2016).

Importantly, cognitive control engagement does not seem to be limited to syntactic processing. Prior research using neuroimaging and patient studies suggests that cognitive control is also engaged when selecting between competing semantic representations (Thompson-Schill, D’Esposito, Aguirre, & Farrah, 1997; Thompson-Schill, Swick, Farah, D’Esposito, Kan, & Knight, 1998; Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Botvinick, Braver, Barch, Carter, & Cohen, 2001). Thus, it may be the case that cognitive control is also engaged during scalar inferencing, when the semantic and pragmatic meanings of “some” are in competition. Recall from Chapter 2 that some accounts of scalar inference processing assume that the inference is preceded by semantic analysis, suggesting that during the time before the inference only the semantic meaning of “some” is being considered (Huang & Snedeker, 2009, 2011, in press). However, there may also be a period of time during which both the semantic and pragmatic meanings of “some” are active and contrasted with one another to select the more likely interpretation (Degen & Tanenhaus, 2015). If this is the case, competition between these meanings may trigger cognitive control.

Empirical support for the engagement of cognitive control during inference processing comes from a dual-task experiment where participants had to complete an executive function task while judging whether sentences featuring scalar terms were true or false. When presented with a sentence like “some oaks are trees” while holding a complex dot pattern in memory, readers made more responses consistent with the semantic meaning compared to the pragmatic meaning of “some” even though the pragmatic meaning is typically preferred (DeNeys & Schaeken, 2007). The authors concluded that the same cognitive control processes that are engaged by the dot pattern task also contribute to scalar inferencing. However, the task they used is better characterized as a test of visuospatial working memory, and does not induce representational conflict like standard cognitive control tasks (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Thus, a more valid measure of cognitive control is required.

4.3 Conflict adaptation paradigm

Evidence of the causal role of cognitive control during syntactic reanalysis has come from “conflict adaptation” studies. Prior work has shown that cognitive control is engaged when conflict is detected, and it lessens the cost of processing later conflict (Gratton, Coles, & Donchin, 1992; Ullsperger, Bylsma, & Botvinick, 2005). For example, the Stroop task (Stroop, 1935) asks participants to name the ink color of a printed color term that either matches (e.g., the word BLUE printed in blue ink) or does not match (e.g., the word BLUE printed in red). When the ink color and word does not match, conflict arises and participants have to inhibit its orthographic representation (BLUE) in favor of its visual representation (RED). Critically, if the non-matching term is preceded by another non-matching term versus a matching term, the processing difficulty associated with the conflicting representations is attenuated (Freitas, Bahar, Yang, & Banai, 2007; Kerns, Cohen, MacDonald, Cho, Stenger, & Carter, 2004). This suggests that cognitive control is triggered by the conflict experienced by the first non-matching term and facilitates processing when conflict re-arises with the second non-matching term.

Hsu & Novick (2016) recently used this paradigm to show that engaging cognitive control leads to successful revision of temporarily ambiguous sentences. Participants were presented with a Stroop task interleaved with a visual world task. The visual world task required participants to correctly analyze a temporarily ambiguous sentence (e.g., "Put the dumpling on the plate into the wok"). Due to the bias of the verb "put," listeners initially assume that "on the plate" is a destination for "the dumpling," and must revise this analysis when later information indicates that it is a modifier. Participants were faster to revise these initial misanalyses when the preceding Stroop trial recruited cognitive control than when it did not. Specifically, looks to the correct destination increased 400ms earlier when cognitive control was engaged. This finding suggests that competition between two syntactic analyses cued the engagement of cognitive control. The authors argue that this led to a decrease in activation of the irrelevant analysis, leading to faster recovery from the initial interpretation.

Experiment 3 extends the findings of Hsu and Novick (2016) to pragmatic processing. Using a conflict-adaptation paradigm, this experiment tests for evidence that cognitive control is recruited during scalar inferencing. Cognitive control may be used to facilitate the adoption of the pragmatic inference by inhibiting the semantic meaning just as cognitive control engagement works to inhibit misanalyses during syntactic processing. The timing of cognitive control engagement is of particular interest in the case of scalar inferences because it remains unclear whether both semantic and pragmatic interpretations are immediately available following “some”. If there is a period of semantic analysis that precedes the generation of the pragmatic inference then there should be a delay in adoption of the pragmatic inference regardless of cognitive control engagement because the pragmatic meaning is unavailable. However, if both interpretations are initially available, cognitive control engagement may have a faciliatory effect in the period immediately following the onset of “some”.

Participants in Experiment 3 completed two interleaved tasks: A Stroop task to engage cognitive control, and the word-learning task from Experiment 1. The critical comparisons are between word-learning trials that featured the pragmatic inference ("some") versus trials that did not ("all"), and those that were preceded by engagement of cognitive control (incongruent Stroop trials) versus no engagement of cognitive control (congruent Stroop trials). Analyses will look for evidence that engaging cognitive control facilitates processing of these scalar terms, leading to either (a) higher accuracy on word learning responses; (b) shorter word-learning response time; and/or (c) greater looks to the Target character during word-learning instructions. Eye-movement analyses will further reveal whether the potential cognitive control effect occurs immediately following the onset of "some" or after a delay. If both the semantic and pragmatic interpretations are both immediately available, then engaging cognitive control should facilitate processing quickly following "some" during word-learning trials preceded by an incongruent Stroop trial compared to a congruent Stroop trial. However, if only the semantic meaning is initially accessed, then cognitive control engagement should have no effect immediately following "some". A second possibility is that cognitive control is recruited only after the pragmatic inference is generated, to help overcome the initial semantic analysis. If this is the case, there may be evidence of cognitive control engagement at a later time period following "some". Finally, cognitive control may have no role in scalar inferencing, in which case engagement should have no effect on how quickly the pragmatic inference is adopted. The results of "some" trials will be compared to trials that feature "all". Because only one interpretation of "all" is considered during processing, there should be no difference in "all" trials preceded by incongruent versus congruent Stroop trials.

4.4 Experiment 3

4.4.1 Participants

Forty undergraduate students from the University of Maryland were recruited for this study. They received either $5 or course credit for their participation. All participants were native English speakers.

4.4.2 Procedures

Experiment 3 was a conflict-adaptation task that interleaved trials of the word-learning task and a Stroop task (Stroop, 1935). During Stroop trials, participants were presented with color terms that either matched in font color (e.g., BLUE printed in blue) or did not match in font color (e.g. BLUE printed in red). The task was to use a three-button mouse to indicate the font color. The response set always consisted of blue, red, and yellow. Sixty of these trials were Congruous (color term matches font color) and 60 were Incongruous (color term does not match font color). Each trial began with a 500ms fixation followed by either a Stroop or word-learning stimulus for 1000ms. After 1000ms or once a selection is made in the word-learning trial, the trial ended and another trial began. Prior to the experimental trials, participants completed 2 word-learning practice trials and 20 Stroop practice trials to become familiar with the tasks. The word-learning task was identical to Experiment 1.

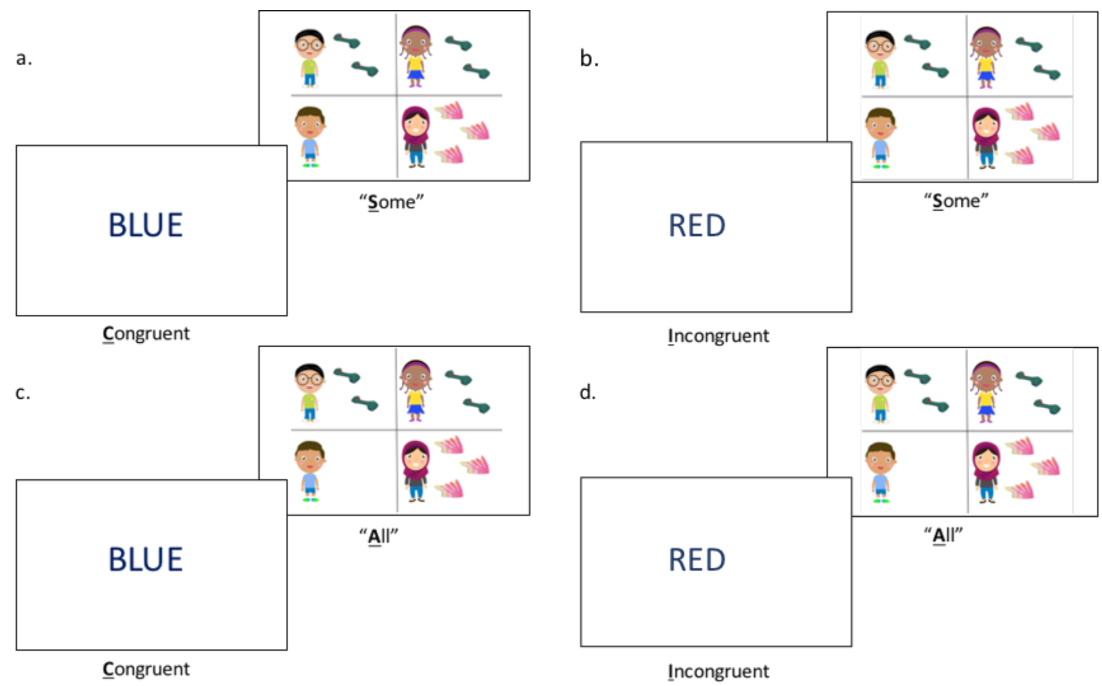


Figure 15. Example of four trial sequences in Experiment 3: (a) Congruent-Some, (b) Incongruent-Some, (c) Congruent-All, (d) Incongruent-All.

4.4.3 Materials

The word-learning trials featured two quantifier instructions from Experiment 1: "Click on the girl who has <u>some/all</u> of the blickets." To test for cross-task conflict adaptation, trial sequences were pseudo-randomized to form four sequence types. There were 12 pairs of critical trials: congruent-all sequences, incongruent-all sequences, congruent-some sequences, and incongruent-some sequences. Thus, within the critical pairs, there were 24 word-learning trials featuring "some," and 24 word-learning trials featuring "all." There were 24 congruent Stroop trials and 24 incongruent Stroop trials. Figure 15 illustrates these four types of trial sequences.

A total of 48 filler visual world trials featuring numeric quantifiers “two/three” and 72 filler Stroop trials were included to ensure that participants could not predict the condition of an upcoming trial. Stroop trials were preceded by other Stroop trials ($n = 45$) and visual world trials ($n = 74$) and visual world trials were preceded by Stroop trials ($n = 27$) and other visual world trials ($n = 21$).

4.4.4 Results

The first set of analyses were conducted to verify that the Stroop task results fit the expected pattern found in previous research. This verifies that incongruent Stroop trials induced conflict and engaged cognitive control, and that the congruent Stroop trials did not. Dependent variables were analyzed using the same strategy as Experiments 1 and 2 except the models included only congruency as a fixed effect.

The Stroop task was analyzed for accuracy and reaction time. A Stroop trial was coded as correct if the participant named the ink color of the word within the 1000ms allowed. A Stroop trial was coded as incorrect if the participant gave a response that was not the ink color of the word or did not respond within 1000ms. The analyses revealed a significant difference between accuracy on congruent Stroop trials and incongruent Stroop trials ($\chi^2(1) = 32.39$, $p < 0.01$) such that accuracy was higher on congruent Stroop trials (89%) than incongruent Stroop trials (86%).

Response time was measured from the onset of the color term presentation to the onset of the participant's verbal response. The analyses revealed a significant difference between reaction time on congruent Stroop trials and incongruent Stroop trials ($\chi^2(1) = 27.04$, $p < 0.01$) such that reaction time was longer for incongruent Stroop trials (891ms) than congruent Stroop trials (804ms). This confirms that participants had more difficulty naming incongruent than congruent Stroop trials, suggesting that the incongruent trials engaged cognitive control.

The remainder of the results are organized around three main questions. When cognitive control is engaged, do listeners (a) make the scalar inference more frequently (fewer action errors); (b) make the scalar inference more quickly (shorter action response time); (c) consider the semantic meaning of "some" to a lesser extent (greater looks to the Target character)?

Participants' actions were coded for accuracy on the word-learning task. As in Experiments 1 and 2, a response was coded as accurate if the participant selected the Target. A response was coded as inaccurate if the participant selected any non-Target. Figure 16 shows that accuracy on the word-learning task was high in both conditions ("some": 95%; "all": 92%). Accuracy was analyzed in a mixed multinomial logistic model using the lme4 software package in R (Bates, Maechler, Bolker, & Walker, 2014). Scalar term (some vs. all) and prior Stroop (congruent vs. incongruent) were included as fixed effects. Models were created with the "maximal" random-effects structure for both subjects and items (Barr, Levy, Scheepers, & Tily, 2013). However, when the model failed to converge within 50,000 iterations, the model was estimated without random intercepts. Analyses revealed no main effect of scalar term (p = 0.50), no main effect of prior Stroop (p = 0.75), and no interaction between the two variables (p = 0.31). These results suggests that cognitive control engagement does not increase the probability that a scalar inference is made.

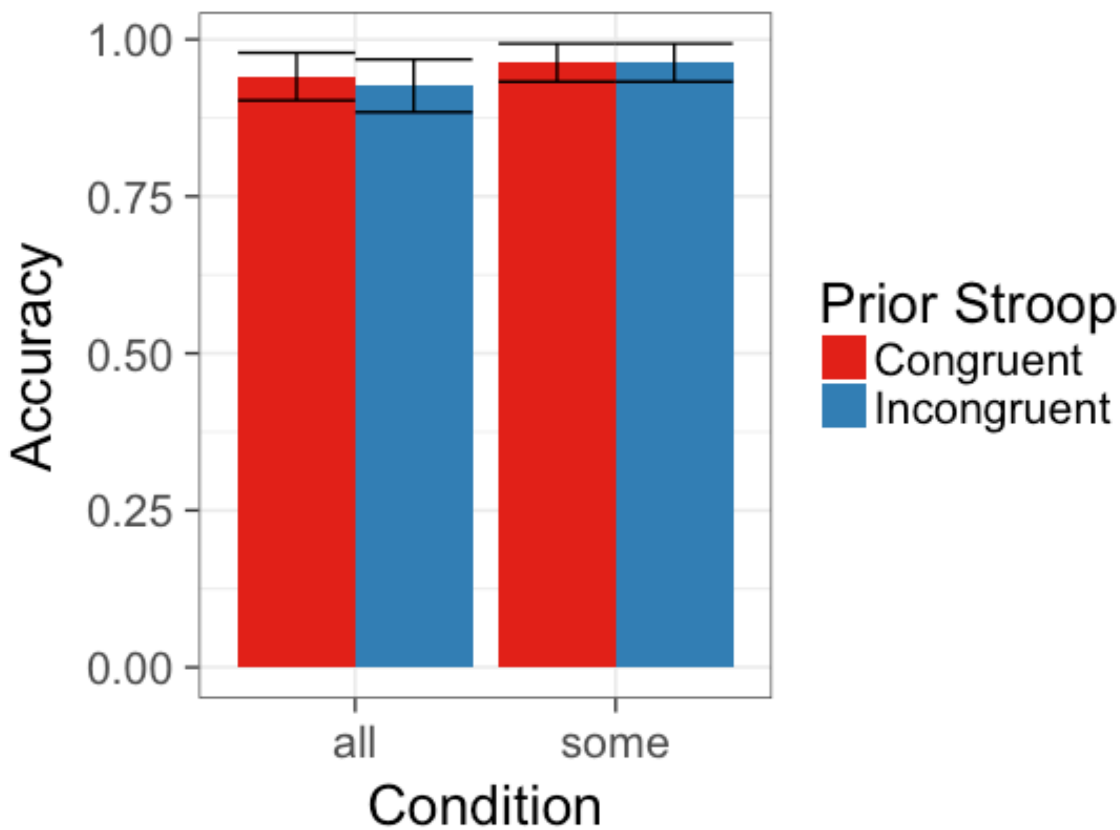


Figure 16. Accuracy on the word-learning task by prior Stroop type, Experiment 3. Error bars represent standard errors.

	Estimate	Std. Error	z-value

Intercept	5.5385	1.1695	4.74
Scalar term	-1.0447	1.5416	-0.68
Prior Stroop	-0.4957	1.5746	-0.32
Scalar term x Prior Stroop	2.9600	2.8885	0.31

Table 7. Results of mixed multinomial logistic model of accuracy on the word-learning task, Experiment 3.

glmer(accuracy ~ 1 + scalar * priorstroop + (0 + Condition * priorstroop |Subject) + (0 + Condition * priorstroop |Item), data = stroopsent, family = binomial)

Participants' actions were also coded for response time. Response time was measured from the end of the instructions to the point at which the participant clicked on a character. Reaction time is depicted in Figure 17. Response time was analyzed in a linear mixed-effects model. Scalar term (some vs. all) and prior Stroop (congruent vs. incongruent) were included as fixed effects. The model was created with the "maximal" random-effects structure for both subjects and items (Barr, Levy, Scheepers, & Tily, 2013). Analyses revealed a main effect of scalar term ($F(1) = 4.24$, $p < 0.05$). However, there was no main effect of prior Stroop ($p = 0.41$) or interaction between scalar term and prior Stroop ($p = 0.22$). These results suggest that cognitive control engagement does not speed up the generation of a scalar inference.

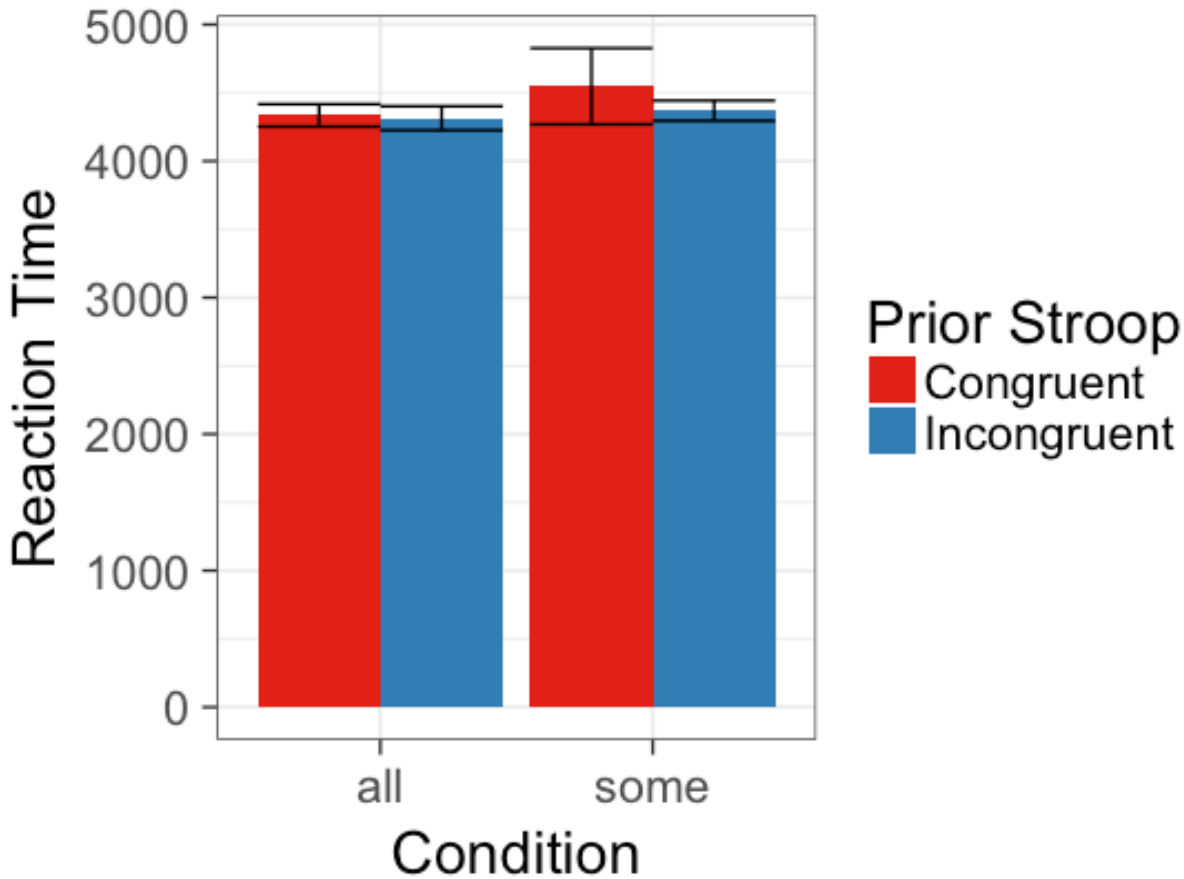


Figure 17. Reaction time on the word-learning task by prior Stroop type, Experiment 3. Error bars represent standard errors.

	Estimate	Std. Error	t-value
Intercept	0.5770	0.0266	25.46
Scalar term	-0.0898	0.01042	-8.61
Prior Stroop	0.0108	0.0279	0.39
Scalar term x Prior Stroop	0.0296	0.0147	2.015

Table 8. Results of linear mixed effect model of reaction time on the word-learning task, Experiment 3.

lmer(rt ~ 1 + scalar * priorstroop + (1 + scalar * priorstroop |Subject) + (1 +

Condition * priorstroop |Item), data = stroopsent, REML = FALSE)

The next analysis looked at eye movements during the Early and Late quantifier regions of the word-learning task. As in Experiments 1 and 2, the primary dependent measure examined Target preference, calculated by taking the ratio of looks to the Target over looks to the Target and Distractor. Target preference was analyzed in a mixed multinomial logistic model with scalar term and prior Stroop included as fixed effects. The models were created with the “maximal” random-effects structure for both subjects and items. However, when the models failed to converge within 50,000 iterations, the models were estimated without random intercepts. Figure 18 shows the proportion of looks to the Target from the onset of the quantifier to the end of the sentence, separated by prior Stroop trial type. Analyses of the Early quantifier region revealed a significant main effect of scalar term ($F(1) = 104.11$, $p < 0.01$), no significant main effect of prior Stroop type ($p = 0.34$), but a significant interaction between scalar term and prior Stroop type ($F(1) = 4.06$, $p < 0.05$). Planned comparisons revealed a significant difference in Target looks during “some” versus “all” word-learning trials preceded by congruent Stroop trials ($t(39) = -2.81$, $p < 0.01$). However, there was no significant difference in Target looks during “some” versus” all trials preceded by incongruent Stroop trials ($p = 0.39$). There was also no significant difference within the “all” trials when preceded by congruent versus incongruent Stroop trials ($p = 0.80$). Critically, within “some” trials, there was a significant difference in Target looks when the prior Stroop trial was congruent versus incongruent ($t(39) = -2.68$, $p < 0.05$) such that Target looks were higher when word learning was preceded by incongruent rather than congruent trials. These results suggest that engaging cognitive control prior to the word learning task facilitated scalar inference processing by promoting consideration of the pragmatic interpretation over the semantic meaning of “some”. By the Late quantifier region, Target looks were similar across conditions and trials preceded by different Stroop trials. Analyses of the Late quantifier region revealed no significant main effect of condition ($p = 0.18$). There was a marginally significant effect of prior Stroop type ($F(1) = 4.12$, $p = 0.05$). There was no significant interaction between condition and prior Stroop type ($p = 0.14$).

To determine when the facilitating effect of cognitive control engagement occurred, a second analysis examined when Target preference exceeded chance for each of the four pair sequences. The results show that Target looks in the “all” trials rose above chance during the 200-300ms time window when preceded by congruent Stroop trials ($t(39) = 2.28$, $p < 0.05$) and incongruent Stroop trials ($t(39) = 2.75$, $p < 0.01$). Target looks in the “some” trials preceded by congruent Stroop trials exceeded chance in the 800-900ms time window. In contrast, Target fixations in “some” trials preceded by incongruent Stroop trials exceeded chance in the 500-600ms time window. This finding suggests that cognitive control engagement accelerates scalar inferencing by 300ms.

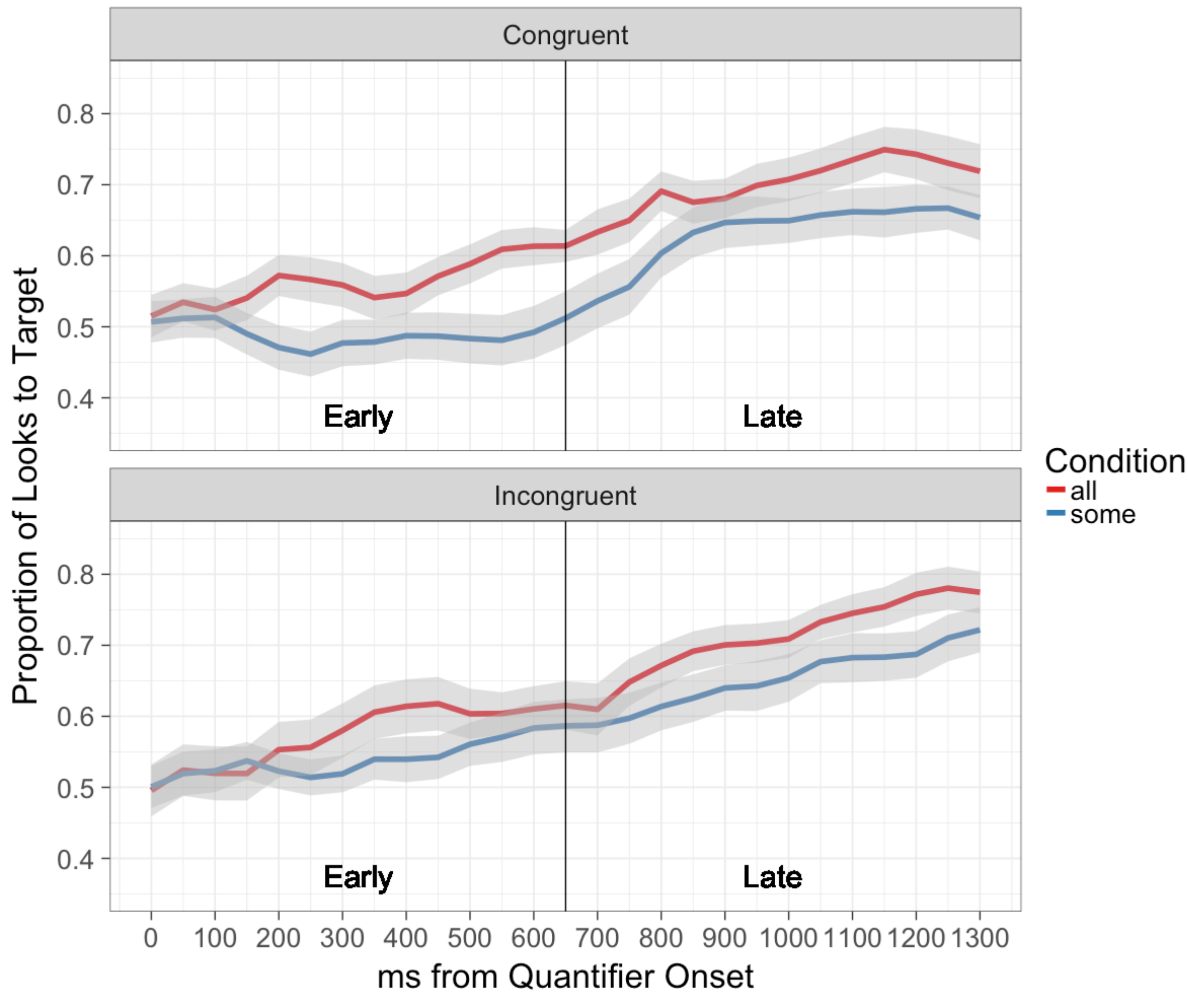


Figure 18. In Experiment 3, Target preference fixations during the Early and Late quantifier regions within the instruction. Error bars represent standard errors.

	Estimate	Std. Error	t value
Intercept	0.8318	0.0376	22.15
Condition	-0.0072	0.0402	-0.18
Prior Stroop	-0.1306	0.0460	-2.84
Condition x prior Stroop interaction	0.1269	0.0555	2.29

Table 9. Results of linear mixed effects model in the Early quantifier region in Experiment 3.
lmer(targetprop ~ 1 + condition * priorstroop + (0 + condition * priorstroop |subject) + (0 + condition * priorstroop |trial), data = early, REML=FALSE)

	Estimate	Std. Error	t value
Intercept	0.8335	0.0379	22.00
Condition	-0.0050	0.0402	-0.13
Prior Stroop	-0.1206	0.0580	-2.08
Condition x prior Stroop interaction	0.1059	0.0703	1.50

Table 10. Results of linear mixed effects model in the Late quantifier region in Experiment 3.
lmer(targetprop ~ 0 + condition * priorstroop + (0 + condition * priorstroop |subject) + (1 + condition * priorstroop |trial), data = late, REML=FALSE)

4.5 Discussion

Chapter 4 looked for evidence that the semantic and pragmatic meanings of "some" are in competition prior to the inference being adopted. To test this, listeners completed the word-learning task featured in Experiments 1 and 2, preceded by a Stroop trial in which cognitive control was either engaged or not engaged. Because cognitive control is recruited when there is competition between representations, evidence that cognitive control is involved in scalar inferencing prior to the inference would suggest that both semantic and pragmatic meanings of "some" are activated during this time. The results of Experiment 3 showed a significant difference in the pattern of eye movements in trials preceded by incongruent and congruent Stroop trials, suggesting that cognitive control is recruited during scalar inferencing. Specifically, a preference for the Target character emerged 300ms earlier in "some" trials when cognitive control was engaged versus when it was not engaged. This suggests that the pragmatic inference was speeded by the engagement of cognitive control.

Importantly, a Target preference emerged in “all” trials, regardless of prior Stroop type, in the 200ms time window. This provides a baseline for the earliest Target preference can be observed when only the semantic meaning of a word must be retrieved. Therefore, there was still a 300ms delay before the pragmatic interpretation was adopted in “some” trials preceded by incongruent Stroop trials. This finding is consistent with an account of scalar inferencing where the semantic meaning of “some” is computed prior to the generation of the pragmatic inference (Huang & Snedeker, 2009, 2011). Thus, it may be the case that while only the semantic meaning of “some” is accessed initially, the pragmatic meaning of “some” is available as early as 300ms later. The results suggest that cognitive control is engaged to facilitate the adoption of the pragmatic inference, possibly by inhibiting the semantic meaning of “some” once the pragmatic meaning is available.

Chapter 5

5.1 Summary of empirical findings

Incremental processing is a key feature of language comprehension. This dissertation investigated the extent to which incrementality extends to the encoding of sentence interpretations in memory. In Chapter 2, I described a word-learning and recall experiment where participants had to calculate a scalar inference in order to learn the meaning of a novel word. Later, they were asked to recall the novel word's meaning. The eye-tracking data showed that participants were considering the semantic meaning of "some" prior to making the pragmatic inference. The recall data showed that participants were less likely to remember the meaning of a novel word if it had been learned via pragmatic inference. This suggests that the semantic meaning of "some" was encoded in memory prior to the inference being made, and that it interfered with recall of the novel word.

In Chapter 3, I described a follow-up experiment that replicated the experiment in Chapter 2 but without the inferencing aspect. Stimuli were presented in downward-entailing contexts to block the pragmatic inference from being drawn. The results showed that when the same amount of attention was paid to the Distractor object but there was no scalar inference, recall of the Target object was similar across conditions. This suggests that visual attention to the Distractor is not enough to interfere with recall. Thus, the lower proportion of matches in the "some" condition of Experiment 1 were not due to greater visual attention to the Distractor object during the word-learning task in the “some” condition compared to the “all/two/three” conditions.

Chapter 4 looked for evidence of conflict between the semantic and pragmatic meanings of “some” prior to the inference being adopted. Listeners completed a conflict adaptation task composed of interleaved Stroop and word-learning trials. Because cognitive control is recruited when there is competition between representations, incongruent Stroop trials that engage cognitive control may facilitate processing of following word-learning trials if there is conflict. The results showed an increase in Target fixations during word-learning trials featuring “some” that were preceded by incongruent Stroop trials, suggesting that the engagement of cognitive control led to stronger consideration of the pragmatic interpretation of “some”. Specifically, a preference for the Target, indicating the adoption of the pragmatic interpretation, emerged 300ms earlier in “some” trials when cognitive control was engaged. This suggests that cognitive control engagement actually speeded the pragmatic inference. However, Target preference in incongruent-“some” trials was still delayed relative to “all” trials preceded by either Stroop trial type, suggesting that there is additional processing following access of the semantic meaning of “some” before the pragmatic inference is adopted.

In the remainder of this discussion, I will examine how the project’s findings pertain to models of sentence processing. Next, I look at the implications for theories of the relationship between semantic and pragmatic processing. Then I discuss remaining questions and future directions of this line of research.

5.2 Implications for models of sentence processing

The findings of Experiments 1-3 were predicted by the model of interpretation building outlined in Chapter 1. This model had two characteristics that were either different than or not clearly defined by previous models: (1) Incremental memory encoding means multiple interpretations of a single sentence can be committed to memory; (2) interpretations based on semantic/pragmatic analysis (in addition to syntactic analysis) persist in memory. First, in contrast to traditional accounts that assume that comprehension results in a single interpretation carried forward in memory, the model described an incremental memory system that could encode multiple interpretations of a sentence in memory. In two-stage and constraint-based accounts, reanalysis completely overrides initial misinterpretations, leading to a single interpretation as the output of comprehension. However, in the incremental memory model, initial interpretations are encoded before competition between lower-level representations is resolved, and prior to the selection of a final interpretation. When more than one interpretation is considered, as in the case of scalar inferences, both are encoded. Experiment 1 found evidence of interference from the semantic meaning of "some" in the memory of an utterance where a scalar inference was drawn. Experiment 2 found no evidence of interference when participants recalled an interpretation where a scalar inference was not drawn. This suggests that multiple interpretations can persist in memory.

This finding is consistent with prior work showing evidence that interpretations based on syntactic analyses are maintained even after reanalysis, and extends it to the semantic-pragmatic level of linguistic processing. This is important because traditional language processing models tend to focus on syntactic analysis as the primary means of interpretation building. In two-stage models, semantic and pragmatic processing only begin after syntactic analysis is complete and a single syntactic representation is chosen. In constraint-based models, sentence meanings are built on syntactic frames, and semantic and pragmatic information can inform structure building, but they cannot override syntactic analysis. However, syntax does not seem to have such a privileged role in interpretation building because interpretation can sometimes be inconsistent with syntactic forms. For example, sentences that follow non-canonical word order like passives are often misinterpreted, especially when they express implausible meanings (e.g., The dog was bitten by the man) (Ferreira, 2003). Also, the language processing system seems to have a mechanism for arriving at sensible interpretations based on disfluencies that make the input ungrammatical (Bailey, 2004; Ferreira et al., 2004). These observations suggest that interpretations were assigned in these cases before syntactic analysis was completed.

Rather than maintain multiple interpretations in a temporary memory store (such as working memory), the evidence suggests that they are committed to longer-term memory, since these memory traces can be accessed to answer comprehension questions, paraphrase sentence meanings, and select depictions of the sentences well after a sentence has been completed (Christianson et al., 2001; Christianson et al., 2006; Ferreira et al., 2001; Patson et al., 2009; Christianson, 2008; Slattery et al., 2013; Malyutina & den Ouden, 2016, Experiment 1). Thus, within the memory encoding account, final interpretations are considered "encoded" when they are committed to a long-term/non-temporary memory store. Prior research suggests that the longer the distance between the generation of an initial interpretation and a disambiguating cue, the more difficult it is to adopt the revised interpretation (Ferreira & Henderson, 1991). This suggests that the longer an interpretation is considered, the more likely it is to get entrenched in memory. This is supported by the results of Experiment 1 linking proportions of Distractor looks to recall performance. If a new interpretation is generated, it will not replace the initial interpretation because it has already been shifted from working memory into a longer-term store. This would result in multiple interpretations of the sentence being carried forward. Future research could investigate factors that may cue encoding, such as the time spent considering a single interpretation, or key pieces of linguistic material like main verbs (Parker & Phillips, 2014).

As noted in the Introduction, recent theories of memory in language comprehension draw a distinction between *limited capacity buffers* that retrieve representations from *declarative memory*, which can include short- and long-term memory (e.g., Lewis & Vasishth, 2005). Encountering a word triggers the retrieval of an associated bundle of features into the focus of attention (e.g., the lexical representation "horse" has features of being a singular noun). Within a small set of buffers, the new word is integrated with previously retrieved feature bundles. Out of the focus of attention, the final interpretation would be represented in declarative memory by a collection of related feature bundles that has an associated level of activation based on prior usage and decay over time. Later, these feature bundles can be retrieved, but there is a chance that a related item will interfere with this retrieval at the point of recall.

The proposed model of processing is may be consistent with this activation-based framework of memory recall. For example, encountering the word “raced” may trigger the retrieval of both a transitive structure and a reduced relative clause structure from declarative memory and into the focus of attention. Under the focus of attention, both structures would be combined with previously retrieved representations (e.g., “horse”) to form higher-level representations. Out of the focus of attention, some features activated by the competing representations would overlap (e.g., “horse”) but an association would extend from these features to both syntactic structures. A feature of the interpretation could be retrieved from declarative memory later on, with the chance that the wrong syntactic structure is reactivated.

However, this memory model of language processing is inconsistent with the results of Experiments 1 and 3 showing a delay in adoption of the pragmatic inference, even when cognitive control is engaged. In the activation-based framework, bundles of features have an assigned activation level based on prior use. Because the pragmatic interpretation of “some” is preferred in upward entailing contexts (Noveck & Posada, 2003; Bott & Noveck, 2004; Geurts & Pouscoulous, 2009; Van Tiel, Van Miltenburg, Zevakhina, & Geurts, 2016), this interpretation (represented by a bundle of features) should have a higher baseline activation level. This would lead to faster retrieval of this meaning of "some" compared to the semantic meaning of "some". However, in Experiment 3 there was a 500ms delay in Target preference in "some" trials compared to "all" trials when no cognitive control was engaged and a 300ms delay when cognitive control was engaged.

5.3 Implications for models of scalar inference processing

As discussed in Chapter 2, a source of debate within experimental pragmatics is whether scalar inferences are generated automatically or after a period of semantic analysis. The present results are incompatible with a Default model of scalar inference processing. Replicating prior research using a visual-world paradigm (Huang & Snedeker, 2009, 2011, in press), eye-tracking results from Experiment 1 suggest that listeners considered the semantic meaning of "some" prior to making the scalar inference. Adding to this data, the recall task revealed evidence that the semantic meaning was encoded in memory. For this to be the case, that meaning had to be accessed before the pragmatic inference was adopted. This could not be the case if the interference was generated by default.

The results are also incompatible with the hypothesis that both the semantic and pragmatic meanings of "some" are activated initially. The constraint-based account of scalar inferences predicts that the more probabilistic support from cues in the sentence means it is more likely that the inference will be made. If there is less support, listeners will take longer to arrive at the inference (Degen & Tanenhaus, 2015). However, the evidence presented here suggests that there was a delay after the onset of "some" and prior to the inference being adopted. Under the constraint-based view, this delay could be due to (1) late arriving information or (2) time taken to compare the pragmatic meaning to other alternatives. However, there was no new information introduced during this time, linguistic or contextual. Moreover, even when cognitive control was engaged, there was a 300ms delay in Target preference when an inference was made ("some" trials) compared to when it was not ("all" trials).

Taken together, these findings are more in line with an account of scalar inference processing where the semantic meaning of "some" is computed prior to the generation of the pragmatic inference (Huang & Snedeker, 2009, 2011). Huang and Snedeker (in press) suggest that there are two routes to inference processing depending on the predictability of set descriptions. Recall that one explanation for the delay in inference generation observed by Huang and Snedeker (2009, 2011) was that more felicitous set descriptors were available (number terms) (Degen & Tanenhaus, 2015). Huang and Snedeker (in press) varied the proportion of scalar versus number labels in their visual world paradigm and found that the timing of the scalar inference was influenced by how the set quantities were referenced throughout the study. When participants heard both scalar and number terms, "some" was interpreted slower than "all." However, when the sets were only described using scalar terms, "some" and “all” were interpreted just as quickly. Moreover, “some” and “two” were judged to be sensible descriptions of subsets when there was a rich discourse (an introduction story) preceding the instructions. Contrary to Huang and Snedeker (2009, 2011), the present experiments did not include a preceding story. According to Huang and Snedeker (in press), listeners rely on bottom-up activation of the lexical entry for “some” when set descriptions are unpredictable. Based on their findings, set descriptions in the current experiments were relatively unpredictable because not only were scalar and number terms used to describe sets, but there was no story adding to the discourse context. Thus, the delay associated with processing “some” may be due to this set label unpredictability. In contrast, Huang and Snedeker predict that when set descriptions are predictable, listeners can encode referents using scalar terms even before they hear the instructions. Thus, in prior work showing rapid pragmatic inferencing (Grodner et al., 2010; Breheny et al., 2013; Degen & Tanenhaus, 2015), pre-encoded descriptions were quickly mapped onto the incoming input, leading to rapid pragmatic interpretations. This suggests that, given more predictable set descriptions, the present set of experiments may have observed faster inferencing. It would be interesting to replicate these experiments under those conditions to see whether the semantic meaning of “some” is still accessed and committed to memory.

Appendices

If needed.

Glossary

If needed.

Bibliography

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of memory and language, 38(4), 419-439.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition, 73(3), 247-264.

Attardo, S., Attardo, D. H., Baltes, P., & Petray, M. J. (1994). The linear organization of jokes: analysis of two thousand texts. Humor-International Journal of Humor Research, 7(1), 27-54.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. Trends in cognitive sciences, 5(3), 119-126.

Awh, E., Dhaliwal, H., Christensen, S., & Matsukura, M. (2001). Evidence for two components of object-based selection. Psychological Science, 12(4), 329-334.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. Journal of Experimental Psychology, 113(4), 518.

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. Neuron, 47(6), 907-918.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255-278.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Grothendieck, G. (2014). Package 'lme4'. R foundation for statistical computing, Vienna, 12.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. Memory & cognition, 8(6), 493-512.

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5), 1450.

Bezuidenhout, A., & Cutting, J. C. (2002). Literal meaning, minimal propositions, and pragmatic processing. Journal of Pragmatics, 34(4), 433-456.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. Journal of memory and language,51(3), 437-457.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. Psychological review,108(3), 624.

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. Language and Cognitive Processes, 28(4), 443-467.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. Cognition, 100(3), 434-463.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. Journal of experimental psychology: Learning, memory, and cognition, 30(3), 687.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. Structures and beyond, 3, 39-103.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. Behavioral and brain sciences, 31(5), 489-509.

Christianson, K. (2008). Sensitivity to syntactic changes in garden path sentences. Journal of psycholinguistic research, 37(6), 391.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. Cognitive psychology,42(4), 368-407.

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and Older Adults' "Good-Enough" Interpretations of Garden-Path Sentences. Discourse Processes, 42(2), 205–238.

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In Linguistic structure in language processing (pp. 273-317). Springer, Dordrecht.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of verbal learning and verbal behavior, 11(6), 671-684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. Journal of experimental Psychology: general,104(3), 268.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. Journal of Experimental Psychology: General, 125(2), 159.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(2), 498.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. Experimental psychology, 54(2), 128-133.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. Cognitive science, 39(4), 667-710.

Duncan, J. (1984). Selective attention and the organization of visual information. Journal of Experimental Psychology: General, 113(4), 501.

Dynel, M. (2009). Humorous garden-paths: A pragmatic-cognitive study. Cambridge Scholars Publishing.

Egner, T., Ely, S., & Grinband, J. (2010). Going, going, gone: characterizing the time-course of congruency sequence effects. Frontiers in psychology, 1, 154.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. Journal of Memory and Language, 30(6), 725-745.

Ferreira, F., & Patson, N. D. (2007). The 'good enough'approach to language comprehension. Language and Linguistics Compass, 1(1-2), 71-83.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. Current directions in psychological science,11(1), 11-15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. Journal of psycholinguistic research, 30(1), 3-20.

Fisk, A. D., & Schneider, W. (1984). Memory as a function of attention, level of processing, and automatization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(2), 181.

Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. Cognition, 6(4), 291-325.

Freitas, A. L., Bahar, M., Yang, S., & Banai, R. (2007). Contextual adjustments in cognitive control across tasks. Psychological Science, 18(12), 1040-1043.

Gadzar, G. (1979). Logical functors. Gerald Gazdar, Pragmatics: Implicature, Presupposition and Logical Form.

Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures?!?. Semantics and pragmatics, 2, 4-1.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation.Proceedings of the National Academy of Sciences, 110(20), 8051-8056.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. Cognition, 73(2), 135-176.

Gleitman, L. (1990). The structural sources of verb meanings. Language acquisition, 1(1), 3-55.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. Language Learning and Development, 1(1), 23-64.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. Journal of memory and language, 28(5), 501-518.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. Journal of Experimental Psychology: General, 121(4), 480.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). “Some,” and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. Cognition, 116(1), 42-55.

Hale, J. (2001, June). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies (pp. 1-8). Association for Computational Linguistics.

Horn, L. (1989). A Natural History of Negation. University of Chicago Press, Chicago.

Horn, L., (1972). On the Semantic Properties of Logical Operators in English. UCLA dissertation, distributed by Indiana University Linguistics Club.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. Psychological science, 0956797615625223.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. Cognitive psychology, 58(3), 376-415.

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. Language and Cognitive Processes, 26(8), 1161-1172.

Huang, Y. T., & Snedeker, J. (in press). Some inferences still take time: Prosody, predictability, and the speed of scalar inferences. Cognitive Psychology.

Huang, Y.T., & Arnold, A. (2016). Word learning in linguistic context: Processing and memory effects. Cognition.

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. Journal of Cognitive Neuroscience, 21(12), 2434-2444.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. Journal of Memory and language, 49(1), 133-156.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. Journal of psycholinguistic research, 32(1), 37-55.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. Science, 303(5660), 1023-1026.

Levinson, S. C. (2000). Presumptive meanings: The theory of generalized conversational implicature. Cambridge, MA: MIT press.

Levy, R. (2008, October). A noisy-channel model of rational human sentence comprehension under uncertain input. In Proceedings of the conference on empirical methods in natural language processing (pp. 234-243). Association for Computational Linguistics.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. Proceedings of the National Academy of Sciences, 106(50), 21086-21090.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. Psychological review, 101(4), 676.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. Cognitive Science, 31(1), 133-156.

Malyutina, S., & den Ouden, D. B. (2016). What is it that lingers? Garden-path (mis) interpretations in younger and older adults. The Quarterly Journal of Experimental Psychology, 69(5), 880-906.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. Journal of experimental psychology: General, 130(4), 621.

Mulligan, N. W. (1998). Perceptual interference at encoding enhances recall for high- but not low-imageability words. Psychonomic bulletin & review, 5(3), 464-469.

Naigles, L. (1990). Children use syntax to learn verb meanings. Journal of child language, 17(2), 357-374.

Notebaert, W., Gevers, W., Verbruggen, F., & Liefooghe, B. (2006). Top-down and bottom-up sequential modulations of congruency effects. Psychonomic bulletin & review, 13(1), 112-117.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. Brain and language, 85(2), 203-210.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. Cognitive, Affective, & Behavioral Neuroscience, 5(3), 263-281.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. Language and Linguistics Compass, 4(10), 906-924.

Panizza, D., Chierchia, G., & Clifton, C. (2009). On the role of entailment patterns and scalar implicatures in the processing of numerals. Journal of memory and language, 61(4), 503-518.

Panizza, D., Huang, Y. T., Chierchia, G., & Snedeker, J. (2009, September). Relevance of polarity for the online interpretation of scalar terms. In Semantics and Linguistic Theory (Vol. 19, pp. 360-378).

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. Cognition, 157, 321-339.

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: evidence from a paraphrasing task. Journal of Experimental Psychology. Learning, Memory, and Cognition, 35(1), 280–285.

Rips, L. J. (1975). Quantification and semantic memory. Cognitive Psychology, 7(3), 307-340.

Ritchie, G. (1999). Developing the incongruity-resolution theory.

Ritchie, G. (2002, April). The structure of forced reinterpretation jokes. In Proceedings of the The April Fools Day Workshop on Computational Humour (TWLT20) (pp. 47-56).

Roland, D., Elman, J. L., & Ferreira, V. S. (2006). Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. Cognition, 98(3), 245-272.

Sauerland, U. (2004). Scalar implicatures in complex sentences. Linguistics and philosophy, 27(3), 367-391.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. Journal of Memory and Language, 69(2), 104-120.

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. Weaving a lexicon, 257294.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. Cognitive psychology, 45(4), 447-481.

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions.Journal of experimental psychology, 18(6), 643.

Swinney, D. A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. Journal of verbal learning and verbal behavior, 18(6), 645-659.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. Journal of Memory and Language, 50(4), 355-370.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268(5217), 1632-1634.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. Proceedings of the National Academy of Sciences, 94(26), 14792-14797.

Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. Proceedings of the National Academy of Sciences, 95(26), 15855-15860.

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.

Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict adaptation effect: It's not just priming. Cognitive, Affective, & Behavioral Neuroscience, 5(4), 467-472.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3), 443.

Van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. Journal of logic, language and information, 13(4), 491-519.

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. Journal of Semantics, 33(1), 137-175.

Wilson, D., & Sperber, D. (2002). Relevance theory. In Handbook of pragmatics. Blackwell.

Wühr, P., & Ansorge, U. (2005). Exploring trial-by-trial modulations of the Simon effect. The Quarterly Journal of Experimental Psychology Section A, 58(4), 705-731.

Ye, Z., & Zhou, X. (2009). Executive control in language processing. Neuroscience & Biobehavioral Reviews, 33(8), 1168-1177.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(1), 1.