

READING EXPERIENCE PREDICTS EYE MOVEMENTS DURING ONLINE AUDITORY
COMPREHENSION

BY

ARIEL NICOLE JAMES

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Master's Committee:

Associate Professor Duane Watson, Chair
Associate Professor Sarah Brown-Schmidt

ABSTRACT

Current theories of language processing suggest that individuals use the probabilities in the language they experience to constrain comprehension as language unfolds, predicting that properties of the linguistic input, such as frequency and predictability, will affect online processing (MacDonald, Pearlmutter, & Seidenberg, 1994; Hale, 2001; MacDonald & Christiansen, 2002; Levy, 2008; Smith & Levy, 2013). It follows that differences between individuals in their idiosyncratic experience with their language should also affect processing, and this has been shown as well (Stanovich & West, 1989; Kuperman & Van Dyke, 2011; Mani & Huettig, 2012; Mishra, Pandey, Singh, & Huettig, 2012). The current set of studies addresses three questions related to individual differences in language experience: (1) whether experience with language that is not specific to the spoken domain nonetheless affect eye movements during auditory comprehension, (2) if so, does experience show an influence even when controlling for other theoretically motivated cognitive factors, and (3) whether language experience shows its influence on more low-level word recognition processes, top-down processes, or both. This paper describes two studies that use individual differences in language experience to predict performance in a visual world eyetracking task following Altmann and Kamide (1999). The design allows an effect of predictability of the target (top-down process) to be assessed separately from an overall facilitation in fixating the target (bottom-up effect). Study 1 finds trending evidence that language experience predicts an overall facilitation in fixating the target, and Study 2 replicates this effect and finds that it remains significant even when controlling for working memory, inhibitory control, phonological ability, and perceptual speed.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: STUDY 1.....	10
CHAPTER 3: STUDY 2.....	20
CHAPTER 4: GENERAL DISCUSSION.....	40
REFERENCES.....	44
APPENDIX A: SENTENCE STIMULI FOR THE EYETRACKING TASK IN STUDY 1.....	52
APPENDIX B: SENTENCE STIMULI FOR THE EYETRACKING TASK IN STUDY 2.....	54

CHAPTER 1: INTRODUCTION

In psycholinguistics, it is widely accepted that the frequency and predictability of a linguistic construction can influence processing. Linguistic elements that are more frequent and more predictable are easier to process than those that are not (MacDonald, Pearlmutter, & Seidenberg, 1994; Hale, 2001; MacDonald & Christiansen, 2002; Smith & Levy, 2013). These theories also predict that individual variability in linguistic exposure will result in processing differences. Individuals with more language experience should be better at processing language. A large body of work suggests that such a relationship exists (Stanovich & West, 1989; Creel, Aslin, & Tanenhaus, 2008; Fine & Jaeger, 2011; Mani & Huettig, 2012; Mishra, Pandey, Singh, & Huettig, 2012; Kamide, 2012; Fine, Jaeger, Farmer, & Qian, 2013). Here, we examine this second prediction more closely and investigate whether greater experience with *written* language predicts facilitated processing in *spoken* language processing. Not only is this a test of experience-based theories of online processing, but if it is the case that reading exposure benefits spoken language comprehension, it would underscore the importance of formal literacy education more broadly. We also examine whether language exposure has effects on language processing that go above and beyond other cognitive mechanisms that might underlie the relation between written and spoken language processing: working memory, inhibitory control, phonological ability, and processing speed. If language experience predicts performance beyond these other cognitive skills, all of which have been argued to underlie individual differences in language processing, it would provide evidence that language exposure plays a central role in online comprehension by literate adults, even in the spoken domain.

Language experience within individuals

Constraint-based processing models serve as the starting point for many current language processing theories (MacDonald, Pearlmutter, & Seidenberg, 1994; Hale, 2001; MacDonald & Christiansen, 2002; Levy, 2008; Smith & Levy, 2013). Within this framework, the processing system considers multiple possible interpretations of the linguistic input in parallel, as various types of information constrain the most likely parse. In the connectionist approach, activation of linguistic units is based on the weights of the connections between them, with the weights being contingent on the frequency of exposure to linguistic patterns (Pearlmutter & MacDonald, 1995).

More recently, surprisal theory has been incorporated into constraint-based models, predicting that processing difficulty is proportional to the total probabilities of alternative parses that have been discarded up to that point (Hale, 2001; Levy, 2008; Smith & Levy, 2013). Therefore, the role of linguistic experience is central; the real-world frequency and the conditional probability of linguistic input should influence activation in the language-processing model.

A great deal of evidence supports the importance of linguistic experience. Comprehenders show sensitivity to the frequencies of words, phrases, and syntactic structures in their language. The overall frequency of words (e.g. according to corpus-based measures such as Kučera & Francis, 1967) predicts reading times such that infrequent words are read more slowly than frequent words (e.g. Broadbent, 1967; Just & Carpenter, 1980; Rayner & Duffy, 1986). Word frequency with respect to the sentential context has also been shown to play an important role. Probable words, as indexed by their cloze probability or n-gram probability are read more quickly than less probable words

(Ehrlich & Rayner, 1981; Staub, 2011; Smith & Levy, 2013).

Individuals are also sensitive to the co-occurrence frequencies of words with syntactic structures. Garnsey, Pearlmutter, Myers & Lotocky (1997) investigated reading of sentences with clauses that are temporarily ambiguous as to whether they will resolve as a direct object (DO; 1) or a sentential complement (SC; 2) For some sentences, the complementizer *that* was included, which disambiguates the structure.

(1) The lawyer acknowledged the judge in the red sweater (DO).

(2) The lawyer acknowledged (that) the judge had been lying (SC).

Some verbs, like *acknowledged* in (1) and (2) are equally likely to take a DO or SC continuation, as measured by sentence completion norms. Verbs such as *believed* are biased to take a SC, while verbs like *warned* are biased to take a DO. Readers are sensitive to these biases, having more difficulty reading the disambiguating region of a SC when the verb is DO-biased, and less difficulty when it is SC-biased, relative to equally biased verbs. Similar to the frequency effects described above, readers' knowledge of the probabilities in their language plays a role during sentence processing.

While this work demonstrates effects of experience within individuals and between items, there are also differences *between* individuals within items that may also be explained by experience.

Language experience between individuals

In addition to predicting that the frequency of linguistic structures affect processing, constraint-based models also predict that individual differences in processing are a result of individual differences in exposure (Pearlmutter & MacDonald, 1995, MacDonald & Christiansen, 2002). Pearlmutter & MacDonald (1995) found that while both highly-skilled

and less-skilled readers (as measured by reading span) are aware of the plausibility of alternative interpretations of sentences offline, only the highly skilled readers are sensitive to this information during online processing. They suggest that this may be due to the skilled readers' greater exposure to, and subsequent familiarity with these constraints, allowing more efficient processing of this information online. MacDonald & Christiansen (2002) suggest that individuals with more language skill have more exposure to all types of sentence structures, including more complex or irregular ones. Importantly, the connectionist approach predicts that exposure will differentially affect irregular input: while processing of regular input benefits from exposure to similar input, processing of irregular input relies on exposure to that particular item (MacDonald & Christiansen, 2002, p. 39).

Manipulated differences in exposure between individuals within a laboratory setting have also given rise to differences in online processing (Creel, Aslin, & Tanenhaus, 2008; Wells, Christiansen, Race, Acheson, & MacDonald, 2009; Kamide, 2012; Fine, Jaeger, Farmer, & Qian, 2013). For instance, Fine and Jaeger (2011) measured participants' reading times for sentences that contain the DO/SC ambiguity explored in Garnsey et al (1997). In the initial self-paced reading task, all critical sentences took the SC continuation. In the exposure phase, participants were split into two groups: in one group, (high reliability) all critical sentences resolved in SC continuations while the second group (low reliability) had half DO continuations and half SC continuation. After three days of exposure, both groups completed a self-paced reading task identical to the first. The researchers found that participants in the low reliability group showed a stronger effect of the disambiguating complementizer *that*, demonstrating that when the verb becomes a less

reliable cue to sentence structure, readers shift to a different cue. This experiment demonstrates that adult readers continuously adapt to their language environment, and that differences in recent exposure lead to different outcomes online in a similar context. However, these results leave open how individual differences in long-term learning prior to coming into the lab influence processing. Correlational studies have been useful in investigating this question, and point to both bottom-up and top-down influences of experience on processing, and these are reviewed briefly below.

Bottom-up influences

General exposure to print materials benefits offline sentence comprehension, even among college students with developed reading skills. Concerned with the role of orthographic processing in explaining individual differences in reading, Stanovich and West (1989) developed a print exposure measure as a way to capture differences in practice with varied orthographic forms. They found that exposure to print, as measured by the Author Recognition Test, predicted orthographic processing skill, as measured by spelling tasks, and that print exposure mediated the relation between orthographic processing and reading comprehension generally. This led the authors to suggest that print exposure affords individuals more practice with uncommon orthographic forms, leading to more efficient reading. This in turn creates a "rich get richer phenomenon" or "Matthew effect", such that individuals with more exposure will become better readers, which will in turn allow them to gain more exposure, and so on (see Stanovich, 1986 for further discussion of Matthew effects in reading).

More recent work using online processing measures has found that individual differences in exposure predict reading times. Kuperman and Van Dyke (2011) monitored

eye movements during sentence-reading from a community sample of 16-24 year-old participants. The researchers were interested in exploring the relation between an individual's online processing and their unique skills and experience, in contrast with the well-documented effects of *word-level* effects, such as length and frequency. The authors argue that "efficient word recognition is not simply about linguistic characteristics of words, but rather about the linguistic characteristics of *particular* words *as learned by particular individuals*"¹ (p. 43), with the assumption that highly experienced or skilled readers will have more precise representations of lexical items at a semantic level or in the phonology-orthography mapping. In support of this claim, the authors found that scores in rapid automated naming and word identification uniquely predicted reading times over and above word-level predictors, such that more skilled readers had faster reading times overall. In fact, these subject-level measures strongly modulated the word-level effects. This work suggests that individual differences in experience affect online reading behavior by way of more efficient lexical representations. Underpinning the explanations of these effects is a link between experience and increased efficiency in low-level processes such as word decoding during reading.

Top-down influences

Alternatively, individual differences in experience might play a role in higher-level top-down processing. Prediction or pre-activation during online comprehension is a robust phenomenon within subjects and is thought to be an important aspect of rapid, efficient language comprehension (for reviews, see Federmeier, 2007 and Kamide, 2008). The importance of prediction is compatible with constraint- and probability-based theories

¹ Emphasis is the authors'.

outlined above (Hale, 2001; Levy, 2008; Smith & Levy, 2013). While the expectation-based theory outlined by Levy (2008) deals primarily with syntactic prediction, it generalizes to other types of prediction, including semantic domains. It makes a straight-forward prediction that the ratio of cloze probabilities of alternative completions should determine processing difficulty. In one demonstration of semantic influences, Federmeier and Kutas (1999) found evidence from event-related potentials (ERPs) that readers predict features of upcoming words. In a context such as (3), readers have difficulty with integration, as indexed by the N400 component, when the word does not match what is expected (*palms*).

(3) They wanted to make the hotel seem more like a tropical resort. So along the driveway, they planted rows of...

Crucially, this is moderated by the degree of featural similarity to the expected word, such that *pin*es is facilitated relative to *tul*ips. This explanation is that the pre-activation of *pal*ms lends activation to *pin*es as well (e.g. being a type of tree). To the extent that a concept shares those activated features, it should be facilitated.

In addition to semantic features, DeLong, Urbach, and Kutas (2005) demonstrated that readers also pre-activate the form of the expected word, such that a mismatched preceding article (e.g. “an” rather than “a”) leads to a disruption before the unexpected lexical item occurs. Both of these demonstrate that during reading, individuals make fairly specific predictions online based on the preceding context.

To examine the role of an individual's experience on online prediction, Mishra, Pandey, Singh, and Huettig (2012) compared low-literate adults to high-literate adults in a visual world paradigm task. On each trial, there were grammatical and semantic cues that the listener could use to predict the sentence-final target. The authors found that only the

high-literates made anticipatory eye movements to the target, suggesting that language experience plays an important role in affording predictive processing. However, there may be important differences between the two groups that gave rise to this apparent experience effect, such as general cognitive skills or experience working with computers. Importantly, these results do not allow us to tease apart the top-down prediction explanation from a bottom-up explanation; i.e. experience leads individuals to be faster to fixate targets overall, regardless of whether the context licenses prediction.

Current study

The current set of visual world studies investigates a prediction of constraint-based and surprisal-based explanations to online comprehension, as outlined in the beginning of this section: just as the corpus-level frequency of items affects processing, an individual's experience with these items should also affect processing. The first study assesses differences in language experience as measured by five different tasks that are primarily based on exposure in reading, and relates this to online auditory comprehension in a visual world task. The second study includes measures of other cognitive skills to rule them out as alternative explanations for the experience effect.

The design addresses both bottom-up and top-down affects of language experience in a spoken language comprehension task. A bottom-up explanation of the role of experience predicts that high-experience individuals will be more efficient at processing the incoming linguistic input by virtue of their more finely honed word-decoding skills in both predictive and non-predictive contexts. In contrast, a strictly top-down explanation predicts that differences due to experience will only be seen in the predictive contexts, such that high-experience individuals will more rapidly and effectively make use of the

constraining contexts to select the intended referent, and performance would be equivalent in neutral contexts. It could also be the case the both bottom-up and top-down processes occur simultaneously.

Overall, the primary goal of this study is to determine whether there are *any* effects of individual language experience on spoken language processing. Most previous work on individual differences in language processing has focused on reading, and to our knowledge, with the exception of Mishra et al. (2012), none have investigated effects of individual differences of literacy on the comprehension of relatively simple spoken sentences. Unlike Mishra et al. (2012), the current study includes trials with and without predictive cues, so that we may observe baseline differences between individuals in this task. For instance, individuals with more experience may look more at the target overall, but the prediction effect might be the same across groups.

Another potentially interesting aspect of the current study design is that we investigate differences within a sample of literate University-affiliated adults. To the extent that there are differences in language experience that can explain processing differences, this is likely an underestimate of the variance in the effect that exposure to language has in the general population.

CHAPTER 2: STUDY 1

The first study investigated the relation between language experience and spoken language processing in a visual word task in which the target object could be predicted from the preceding verb. In a version of the Altmann and Kamide (1999) paradigm, participants viewed scenes such as Figure 1 and heard either a constraining sentence such as (4) or a less constraining sentence (5). At the verb in (4), a listener is able to anticipate that *cake* is a likely object, as it is the only edible object in the scene. In contrast, the verb in (5) does not afford this prediction, as all of the objects in the scene can be moved by the boy.

(4) The boy will *eat* the cake.

(5) The boy will *move* the cake.

If language experience affects top-down processing of sentences, we expect that individuals with more experience will anticipate the target more after processing the predictable verbs, but there will be no difference on trials without predictive verbs. If experience affects only more general bottom-up spoken comprehension, we expect that more experience will facilitate comprehension of the target regardless of its predictability, resulting in a main effect of experience on target fixations. Critically, we were interested in whether there are any effects of linguistic experience on fixations at all.

Method

Participants. Participants were undergraduate students at the University of Illinois in Urbana-Champaign participating for class credit. They were all native speakers of English with normal hearing and normal or corrected to normal vision. 124 subjects participated in the study; 13 were dropped due to missing data (three were missing one of

the language experience measures due to computer failure, ten were missing all eyetracking data due to calibration failure), resulting in 111 participants included in the analyses. One person did not report demographic information. Of the remaining 110 participants, 70 were female and the average age was 19 years and 2 months (range: 18-22 years).

Materials. Measures for the language experience assessment, and stimuli for the visual world comprehension task are described below.

Language experience. Language experience was measured with five different tasks. The primary goal in selecting these five particular tasks was to find a diverse set of measures that do not specifically probe sentence comprehension or prediction.

Author Recognition Test. The Author Recognition Test (ART) was developed as a measure of exposure to print materials (Stanovich & West, 1989). In the current study, we used an updated and slightly lengthened version of the task developed by Acheson, Wells, and MacDonald (2008) that included the names of 65 authors' names and 65 foil names. In their version, all 130 names were randomized and presented to participants on a sheet of paper and participants circled the names that they believed belonged to the authors of books. For the current study, the test was adapted for the computer. Participants in the current study saw names presented one at a time and made a judgment about each name. Names were presented in a random order, and two response buttons appeared at the bottom of the screen reading "Author" and "Don't know". Participants were told that there was a penalty for guessing, so they were encouraged to only respond with "Author" if they were sure, and to otherwise choose "Don't know".

Extended Range Vocabulary Test. The Extended Range Vocabulary Test (ERVT; Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D., 1976) includes 48 words of varying difficulty. Participants chose which among five single words has the most similar meaning to the given word. Participants are told there is a penalty for guessing, so they are encouraged to select a sixth “Not sure” option if they are unfamiliar with the word.

North American Adult Reading Test. The North American Adult Reading Test (NAART) was developed as a way to estimate pre-morbid IQ in brain trauma patients (Blair & Spreen, 1989). Participants receive a list of 61 words with irregular spellings, presented one at a time at increasing difficulty. The participants’ task is to read the word and correctly pronounce it.

Comparative Reading Habits. Comparative Reading Habits (CRH) is a self-report survey in which participants answer five questions comparing their own reading habits to what they perceive to be the norm for their fellow college students (Acheson et al., 2008).

Reading Time Estimate. Reading Time Estimate (RTE) is a self-report survey in which participants estimate how many hours in a typical week they read various types of materials, including fiction, newspapers, and online materials (Acheson et al., 2008).

Eyetracking. The design of the eyetracking task closely followed that of Altmann and Kamide (1999). Sixteen scenes were created using Photoshop and cartoon images from the ClipArt database. Two sentences were recorded for each of these scenes, one with a predictive verb and one with a neutral verb. For instance, for a scene with a boy sitting on the floor surrounded by a toy train, toy car, ball, and a piece of cake, participants heard either *The boy will eat the cake* or *The boy will move the cake* (see Figure 1). Scenes either contained four or five total objects. In scenes with five objects, one object did not make

sense in either sentence context. An additional sixteen filler scenes were created such that the target object described in the sentence was not present in the scene. The 16 critical sentences were taken from Altmann and Kamide (1999) and the corresponding scenes were edited and re-colored. The 16 filler scenes and sentences were created for the study. All sentences were recorded by the same female speaker of Midwestern American English and read at a natural speech rate. A full list of the stimuli are presented in Appendix A.

Procedure. All participants completed the tasks in the same order to minimize variability between subjects that is due to differences in the experimental session (see Swets, Desmet, Hambrick, & Ferreira, 2007 for discussion). Participants completed the ART, the ERVT, the NAART, the CRH, the RTE, and then the eyetracking task. The entire procedure took 35 to 50 minutes.

Language experience. All language experience measures were programmed and displayed using Matlab, on the same computer as the eyetracking task. All of the language experience measures together typically took 15 to 30 minutes for participants to complete.

Eyetracking. Participants were seated at an Eyelink 1000 desk-mounted eyetracker. Their heads were stabilized using a chin rest. Participants were instructed to decide whether the recorded sentence could be a possible description of the scene. They were instructed that they should respond *yes* to *The man will choose the watch* if there was a watch in the corresponding scene, and *no* otherwise. Before calibration, participants completed a practice trial in which they viewed a scene and heard the sentence *The man will light the candle*. After the participants chose a response, they were told that they should have responded *yes* because there was a candle present in the scene, even though there was no visible lighter or match.

Participants then completed a calibration procedure and began the task. Before each trial, the eyetracker was recalibrated by having the participant fixate a centrally presented white dot on a black screen.

Results

Language experience. Table 1 summarizes performance on the five different measures of language experience. Table 2 presents the correlations among the five measures. With the exception of the Reading Time Estimate (RTE), the measures are reliably correlated with one another.

Eyetracking. The eyetracker failed to record 26 trials (0.74%). The results of the eyetracking task replicate the Altmann & Kamide (1999) results. Figure 2 shows the relative proportion of fixations to objects in the scene over time, where the onset of the verb is aligned at 0 milliseconds. Statistical results of the condition effect are given in the following section.

Language experience and prediction. To predict the proportion of target fixations, a multi-level mixed effects model was fit with condition, language experience, and their interaction as fixed effects. Condition was coded using effects coding, where the predictable condition was approximately 0.5 and the control condition was approximately -0.5 (weighted for the unbalanced number of trials per condition). Language experience was entered in the model as the average of z-scores for each of the language experience tasks. The maximal random effects structure was included, following Barr, Levy, Scheepers, and Tilly, 2013, excluding terms only when their inclusion lead to collinearity in the model. Separate models were fit for each of three windows: during the verb, during the noun, and from the noun offset to the end of the trial. Start and end times were offset by 200 ms to

allow for the time taken to plan and execute an eye movement (Allopenna, Magnuson & Tanenhaus, 1998; Huettig & Altmann, 2005). P-values reflect the significance of the likelihood ratio test for comparing nested models.

Verb region. Random slopes for subjects were dropped due to collinearity. Condition had a positive and marginal effect on target fixations ($\beta = 0.0353$, $SE = 0.0178$, $p = 0.061$), with more target fixations in the predictable condition. Language experience main effects were also positive and marginal ($\beta = 0.0263$, $SE = 0.0146$, $p = 0.0728$). The interaction was not significant ($\beta = 0.0082$, $SE = .0230$, $p = 0.719$).

Noun region. Random slopes for subjects were dropped due to collinearity. The predictable condition resulted in more fixations than the control ($\beta = 0.1043$, $SE = 0.0347$, $p < 0.01$) and more language experience led to a marginal increase in fixations ($\beta = 0.0292$, $SE = 0.0166$, $p = 0.0803$), with no interaction ($\beta = -0.0147$, $SE = 0.0241$, $p = 0.5414$). The language experience main effect reached significance when the verb and noun windows are collapsed ($\beta = 0.036$, $SE = 0.0124$, $p < 0.01$), and the condition effect remained significant ($\beta = 0.0777$, $SE = 0.0235$, $p < 0.01$).

Post-noun region. Random slopes for subjects were retained in the model. The condition effect remained significant but in the opposite direction of previous regions ($\beta = -0.1037$, $SE = 0.0313$, $p < 0.01$), such that there were more target fixations in the control condition, driven by the delayed peak in fixations relative to the predictable condition. The language effect no longer approached significance ($\beta = -0.0038$, $SE = 0.0248$, $p = 0.9489$), and there was no evidence of an interaction ($\beta = -0.0464$, $SE = .029$, $p = 0.9489$).

It was unexpected that subjects with more experience would show more fixations to the target across both conditions as early as the verb because in the non-predictive

condition, there should not have been information to identify the target until the XXXX region. However, dummy coding by condition revealed that there was only a marginal effect of language experience when looking at just the predictable condition, with more experience leading to more target fixations ($\beta = 0.0396$, $SE = 0.019$, $p = 0.067$). This effect did not reach significance when looking at the control condition separately ($\beta = 0.0194$, $SE = 0.0191$, $p = 0.3380$), suggesting that the effect of experience may be driven by the predictable condition, consistent with the logic of an experience by condition interaction.

Fixations to other objects. The main effect of language experience across both conditions in the early window suggest the curious result that individuals are able to look more to the target during the verb *even when the verb contains no predictive semantic information*. A possible explanation is that individuals with more experience are simply faster to look away from the agent and toward any of the potential referents in the scene. Consistent with this, there was a significant effect of language experience on fixations away from the agent ($\beta = -0.0604$, $SE = 0.0287$, $p < 0.05$), and a positive effect on fixations to the competitor objects ($\beta = 0.0267$, $SE = 0.0125$, $p < 0.05$) in the verb time window. There was no effect of condition on fixations to either object (agent: $\beta = -0.0025$, $SE = 0.0327$, $p = 0.8948$; competitor: $\beta = 0.0026$, $SE = 0.0157$, $p = 0.8918$). This suggests that more language experience allows listeners to more quickly process the agent, which always precedes the verb, allowing them to more quickly shift away from agent in the scene during the processing of the verb.

Discussion

From these results, there is no evidence of a language by condition interaction. Rather, we find a language main effect over the verb and noun regions. This suggests that

language experience affects bottom-up processes, such as word recognition, allowing for more efficient integration of the sentence and accompanying scene. However, it is possible that our study did not have sufficient power to detect the interaction. Further, it is not clear that our effect is the result of language experience *per se* and not some other factor that covaries with experience, such as higher verbal working memory or perceptual speed. Our second study attempted to address these issues by increasing the number of trials per subject and including measures of other cognitive abilities.

Study 1 Tables and Figures

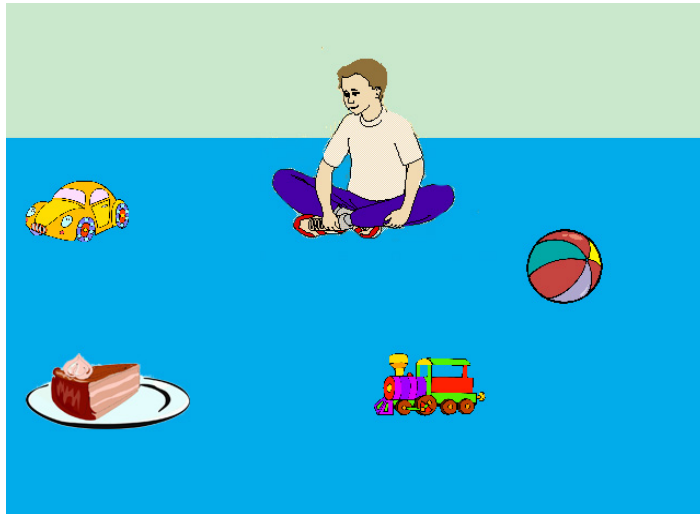


Figure 1. Display for the sentence *The boy will eat the cake* or *The boy will move the cake*. Images were updated from the original Altmann & Kamide (1999) stimuli.

Table 1

Summary of performance on language experience measures

Measure	Possible Range	Observed Range	Mean Score (SD)
ART	Min: -65 Max: 65	Min: 0 Max: 30	12.05 (6.34)
Extended Range ERV	Min: -12 Max: 48	Min: 1.5 Max: 29.25	13.97 (6.19)
NAART	Min: 0 Max: 61	Min: 10 ^a Max: 48	29.06 (7.68)
CRH	Min: 5 Max: 35	Min: 8 Max: 31	21.14 (4.32)
RTE	Min: 0 Max: 63	Min: 6 Max: 41	19.77 (8.34)

Notes: ART = Author Recognition Test, ERVT = Extended Range Vocabulary Test, NAART = North American Adult Reading Test, CRH = Comparative Reading Habits, RTE = Reading Time Estimate. ^aThis subject accidentally skipped one item, so the score is out of 60. Proportion correct are used in the analyses.

Table 2

Correlations among language experience measures

	ART	NAART	ERV	CRH
NAART	0.55***	--		
ERV	0.66***	0.66***	--	
CRH	0.39***	0.34***	0.47***	--
RTE	0.17 ⁺	0.14	0.053	0.32***

Notes: ART = Author Recognition Test, NAART = North American Adult Reading Test, ERVT = Extended Range Vocabulary Test, CRH = Comparative Reading Habits, RTE = Reading Time Estimate. ⁺ $p < 0.1$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$

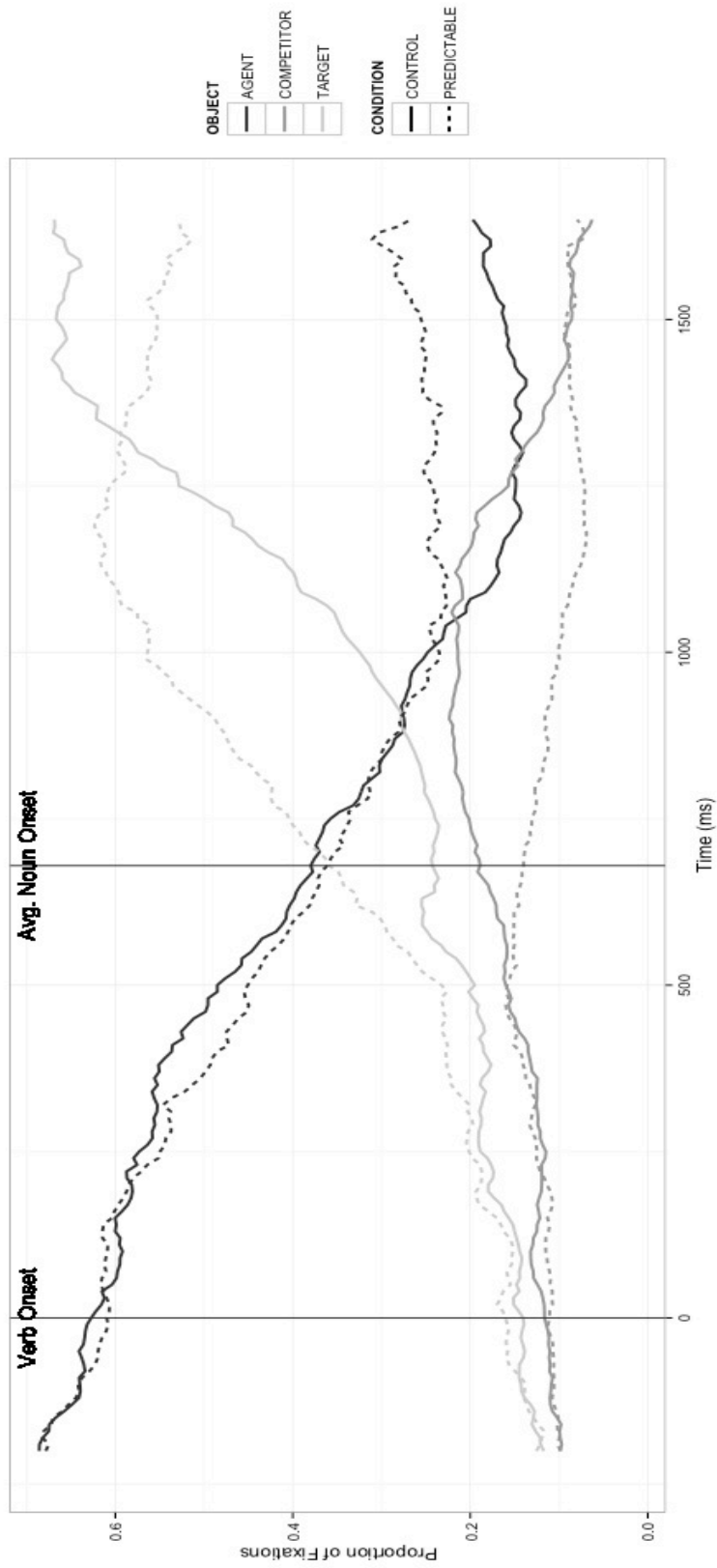


Figure 2. Proportions of fixations to the objects in the visual world display in Study 1. The verb onset is align at 0 milliseconds and the average noun onset is shown for illustration.

CHAPTER 3: STUDY 2

The second study follows the same design as Study 1 but with two important changes: the number of trials per subject² in the eyetracking task was doubled to increase power, and measures of other cognitive abilities were also included. Working memory, inhibitory control, perceptual speed, and phonological processing were included because they address possible mechanisms that underlie the language experience effect found in Study 1, and because they have individually been implicated in previous research on individual differences in sentence processing.

Theories of sentence processing have traditionally focused on cognitive load during reading, and the mechanisms readers use to cope with it (Just & Carpenter, 1980; Frazier, 1987; Gibson, 2000). This earlier work largely focused on the role of verbal working memory, either by way of a single system (Just & Carpenter, 1992) or two separable memory capacities (Waters & Caplan, 1996) that support language comprehension. The special status of working memory was later challenged by MacDonald and Christiansen (2002), who argue that verbal working memory and language comprehension are both measures of linguistic skill. Effects that were previously attributed to differences in working memory are instead explained using language experience. The authors describe “capacity” as a result of the knowledge and architecture of the network itself, rather than an independent component. Wells, Christiansen, Race, Acheson, and MacDonald (2009) tested these claims using both a training study and a computational model, demonstrating that exposure to relevant structures can give rise to individual differences in online processing. Other work has included inhibitory control (Gernsbacher, 1993; Novick,

² The number of trials was doubled rather than subjects because, due to the length of the study session, doubling the number of subjects would have been intractable.

Trueswell, & Thompson-Schill, 2005, 2010), phonological ability (Wagner & Torgeson, 1987; MacDonald & Christiansen, 2002; Acheson & MacDonald, 2011), and perceptual speed (Salthouse & Babcock, 1991) as important factors in language processing.

Our present design allows us to simultaneously address the contributions of these different factors within individuals by including them all in the current study. A strength of this approach is that there are multiple measures of each of these five constructs, as no one measure is process-pure. If language experience *per se* guides eye movement behavior in spoken sentence processing, as is suggested by Study 1, the effect of experience should remain even after these other cognitive factors are accounted for.

Method

Participants. Participants were undergraduate students at the University of Illinois in Urbana-Champaign participating for class credit or for \$8 per hour. They were all native speakers of English with normal hearing and normal or corrected to normal vision. 131 participated in the study. A total of 31 subjects have missing data. Fifteen are missing eyetracking data and are excluded from the analyses; seven failed calibration and eight did not show up for the second session of the study during which the eyetracking task took place. An additional 16 subjects are missing at least one individual differences measures; nine ran out of time during the session and were not able to finish the remaining tasks, five experienced a technical malfunction, and two misunderstood a task. Subjects that had at least one measure for each individual differences domain were included in the analyses; excluding these subjects does not substantively change the results. Of the remaining participants, 75 were female and the average age of the remaining participants was 20

years and 10 months (range: 18-67; excluding the 67-year old participant, the average is 20 years and 5 months, and the maximum age is 35).

Materials. Measurements for the five different cognitive domains, and for the visual world eyetracking task, are described below.

Language experience. Participants completed the same language experience battery as described in Study 1.

Working memory. Working memory has played a prominent role in the investigation of individual differences in sentence processing, although much of this literature deals with complex syntactic structures (e.g. long distance dependencies in Gibson, 2000). Individuals who fixate the target more quickly may do so because they can effectively hold the sentential context in mind and making inferences about it before the sentence has concluded. Working memory was assessed using three complex span tasks, described in more detail below.

Reading span. The reading span task, adapted from Daneman and Carpenter (1980), required participants to read sentences out loud and make a judgment about whether the sentence was true. Sentences were taken from Stine and Hindman (1994). After the judgment was made, a single letter was displayed for the participant to remember, following Unsworth, Heitz, Schrock, and Engle (2005). While other versions of the task require participants to remember the final word of each sentence, a random letter was used so that participants' memory performance is less likely to be confounded with their skill at reading the sentences, or familiarity with the sentence-final words. To further correct for overall differences in reading ability, participants completed a calibration phase at the beginning of the task that excluded the letter-memory component. This determined

how long they would be given to read the sentences during the test phase (Unsworth et al., 2005). Participants were then given two practice trials, each containing two sentence-letter pairs (a span length of two). The test trials then tested span lengths two to six in a random, rather than ascending order. The random presentation of all span lengths was done to gather information on the subject's ability at each level (rather than stopping once they fail a span length, as is often done; see Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005) and to deconfound span length with the increasing likelihood of proactive interference over time (Lustig, May & Hasher, 2001).

Listening span. The listening span task followed the same procedure as above, except that the sentences and letters were presented auditorily and the calibration phase was based on the latency to make the true/false judgment. No sentences during this phase were repeated from the reading span task, although they did also come from Stine and Hindman (1994).

Operation span. The operation span task procedure was similar to the reading and listening span. Rather than comprehend sentences, participants were asked to complete and verify math equations involving two operations. Each equation involved either multiplication or division followed by either addition or subtraction. Once participants solved the problem, they pressed a spacebar to see a probe number, and participants indicated whether or not it was the correct solution to the preceding problem. As in the previous span tasks, participants completed a calibration phase that determined the maximum time they were permitted to spend on the processing portion of the task. After the judgment was complete, participants were presented with a letter to recall later, as in the previous tasks.

Inhibitory control. Inhibitory control is a domain-general ability falling under the general umbrella of executive attention. Inhibitory control is typically used to describe the ability to resist distraction from either internal or external stimuli, although Friedman and Miyake (2004) point out that definitions have been vague and inconsistently applied across literatures. Here, we conceptualize inhibitory control as the ability to override a conflicting response in favor of responding according to task goals.

It should also be noted that the complex span tasks described above can be conceptualized as measures of inhibitory control processes (Chun et al., 2011; Conway et al., 2005). This is true of other tasks, such as the n-back, which is treated as an inhibitory control measure and a working memory capacity measure (Kane & Engle 2002; Conway et al., 2005; Hussey & Novick, 2012). Inhibitory control as treated here is more specifically a measure of conflict resolution, or the ability to override salient cues or prepotent responses in favor of task-relevant information and responses. Conflict resolution in particular has played an important role in recent investigations of individual differences in sentence processing, specifically in ambiguity resolution and garden-path recovery (Gernsbacher, 1993, 1995; Novick et al., 2005, 2010). Even so, based on previous research on the relation between working memory and inhibitory control, we expect the tasks described here to be correlated with our measures of working memory (antisaccade: Kane et al., 2001; Unsworth, Schrock, & Engle, 2004; flanker: Heitz & Engle, 2007; Stroop: Kane & Engle, 2003).

Antisaccade. The basis of the antisaccade task is that it is a prepotent response to make a saccade to a suddenly presented stimulus in the visual field. Following Kane, Bleckley, Conway, and Engle (2001), participants began critical trials by fixating the in the

center of the screen. An anti-predictive cue appeared at one side of the screen after a variable length of time to prevent participants from predicting when this cue would appear. A target letter (either B, R, or P) then appeared on the opposite side of the screen as the cue, preceded by a forward mask (the letter H) and followed by a backward mask (the number 8). Participants then needed to indicate the identity of the target letter. Prior to the 72 critical trials, participants completed a response-mapping phase to learn which keys to press (1, 2, and 3) in response to the target letters, then 52 practice trials that gave a feedback tone only in response to incorrect responses.

Flanker. Participants completed a version of the "flankers" response competition paradigm (Eriksen & Eriksen, 1974; see Eriksen, 2007 for review) in which a visually-presented target item is flanked by either congruent items that facilitate correct responding, or incongruent items that inhibit correct responding. Participants in this task indicated the direction of an arrow that was flanked by four arrows of the same (< < < <) or different (> > < > >) direction. The incongruent items are thought to activate the incorrect response, making it more difficult to select the correct response, as measured by response latency (Eriksen, 2007).

Stroop. Participants completed a self-paced version of the Stroop task (Stroop, 1935) in which they completed a conflict-free phase followed by a conflict phase. In both phases, the task is to name the color presented against a black background on the computer screen as quickly as possible. Participants were trained on the appropriate color names before the task. These were *red, orange, yellow, green, blue, and purple*. In the conflict-free phase, participants named aloud the color of a filled rectangle. In the conflict phase, the stimulus to be judged is a word rather than a box, giving rise to a conflicting response of

simply reading the word. The words were maximally conflicting, as they were task-relevant color terms that never matched the color that the stimulus was presented in (e.g. the word *blue* presented in green, where the correct response is to say *green*). Accuracy is typically high in the self-paced version of the task, so the difference in reaction time between the two phases is used as a measure of interference.

Perceptual Speed. Measures of perceptual speed were included in order to address concerns that individuals with more language experience are faster at the task overall due to a domain-general ability to process perceptual stimuli quickly, which could independently enhance reading skill and the ability to search a visual display for objects of interest. We included two measures of perceptual speed.

Letter comparison. In the letter comparison task, following Salthouse & Babcock (1991), participants were asked to compare two arrays of consonant letters as quickly as possible. Trials were presented in six blocks: two blocks comparing three-letter arrays, two blocks comparing six-letter arrays, and two blocks comparing nine-letter arrays. During each block, participants were given 20 seconds to complete as many comparison trials as possible. On all mismatching trials, only one letter differed between the arrays. Participants completed two practice trials with feedback, each with three-letter arrays, in which one trial contained a match and the other contained a mismatch.

Pattern comparison. The procedure of the pattern comparison task was largely the same as for letter comparison, except that arrays of line segments rather than letters are compared (Salthouse & Babcock, 1991). Blocks of three-, six-, and nine-segment arrays were presented in an order identical to that in the letter comparison task. After completing one match and one mismatch practice trial, participants were asked to perform the critical

trials as quickly as possible.

Phonological ability. Phonological ability is a possible factor that underlies sentence processing skill, as the ability to create phonological representations may aid in the maintenance of the words that have been encountered so far, as is required in reading as well as verbal working memory tasks (MacDonald & Christiansen, 2002; Acheson & MacDonald, 2009, 2011). For this reason, we expect that phonological ability may be related to individual differences in language experience, as well as the working memory span tasks. It is also possible that the clarity of phonological representations may aid in the comprehension of the spoken sentences presented in the eyetracking task, independent of differences in language experience. Phonological ability was assessed using three measures. Two of these, Blending Nonwords and Phoneme Reversal, are taken from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999).

Blending nonwords. On each trial of the blending non-words task, participants heard a list of phonemes or syllables and were asked to combine these elements into one nonword. For instance, if the participants heard /h/, /ε/, and /t/, they would need to produce /het/ as one word. The number of elements ranged from two to eight. Participants were given six practice trials and eighteen critical trials.

Phoneme reversal. In the phoneme reversal task, participants heard a nonsense words and were asked to repeat the word and then pronounce it backwards, creating a real English word. For instance, if the participants heard /stu:b/, they would need to produce the word *boots*. Participants were given four practice trials and eighteen critical trials.

Pseudoword repetition. The pseudoword repetition task, following Gupta (2003),

asks participants to listen to a non-word and immediately repeat it back. Materials, taken from Gupta (2003), were created by combining syllables from English words into novel, phonotactically legal strings, such as *waydish* and *spentonymidderoxing*. After completing six practice trials, participants were given 96 items of either two, four, or seven syllables.

Eyetracking. The design of the eyetracking task is the same as described in Study 1. The number of predictive trials, non-predictive trials, and filler trials were each doubled, such that each participant completed 32 experimental trials and 32 filler trials. The additional materials are listed in Appendix B.

Procedure. As in Study 1, all participants completed all tasks in the same order. The individual differences battery was comprised of the 16 total tasks described above. The entire procedure took place over two sessions, scheduled 24 hours apart in order to minimize fatigue in each session. During the first session, participants completed a self-paced reading task that is not discussed here, then the working memory tasks, the perceptual speed tasks, the inhibitory control tasks, and began the language experience tasks (ERV and ART). The first session typically lasted 90-120 minutes. During the second session, participants completed the prediction eyetracking task as well as another eyetracking task that is not covered here. They then completed the three remaining language experience tasks and the phonological ability tasks. The session was concluded with a participant questionnaire and a debriefing. The second session typically lasted 40-60 minutes.

Individual differences battery. All individual differences tasks were programmed and displayed using Matlab. Participants completed tasks at their own pace, without additional separation between the subcategories of tasks.

Eyetracking. The procedure for the eyetracking task is as described in Study 1.

Results

Individual difference measures. Performance on the individual difference measures across the five domains is summarized in Table 3.

Correlations. Table 4 lists the correlations among the composite scores from each domain. In general, these scores were not correlated with one another, suggesting that they are measuring unique skills. However, there is a reliable correlation between phonological ability and language experience, which is expected given the vast literature linking phonological processing and reading ability, lead by Wagner and Torgeson (1987) who later developed the CTOPP test battery used in the current study. There is also a reliable correlation between perceptual speed and inhibitory control, driven by the anti-saccade task, which is measured here in reaction time (rather than a difference of reaction times, as is the case for Stroop and Flanker). It is interesting to note that we do not see the correlation between working memory and inhibitory control that we anticipated given the prior research discussed earlier. We also predicted a correlation between working memory and phonological ability, and this was marginal in the current study.

Within each domain, task performance is generally reliably correlated, with the exception of inhibitory control. Anti-saccade is not correlated with Stroop ($r = 0.17, p = 0.0997$) and is marginally correlated with Flanker ($r = -0.11, p = 0.2945$). For perceptual speed, Letter Comparison and Pattern Comparison are highly correlated ($r = 0.70, p < 0.001$). For working memory, operation span is correlated with reading span ($0.5, p < 0.001$) and listening span ($0.54, p < 0.001$), and reading and listening span are also correlated ($r = 0.45, p < 0.001$). For phonological ability, pseudoword repetition is

correlated with both blending non-words ($r = 0.28, p < 0.01$) and phoneme reversal ($r = 0.25, p < 0.05$). Blending non-words and phoneme reversal, both from the CTOPP battery (Wagner, Torgesen, & Rashotte, 1999), are correlated ($r = 0.30, p < 0.01$). Correlations among language experience measures are summarized in Table 5.

Eyetracking. A total of 55 trials (0.73%) were not recorded by the eyetracker. The results of the eyetracking task replicate Study 1 and the original Altmann & Kamide (1999) results. Figure 2 shows the relative proportion of fixations to objects in the scene over time, where the onset of the verb is aligned at 0 milliseconds. Statistical results of the condition effect are given in the following section.

Language experience and eyetracking. The analysis strategy here is the same as that of Study 1. The first set of analyses is a replication of the models fit in Study 1, ignoring the other cognitive abilities. The second set of analyses uses a residualized language experience predictor, created by regressing the four other individual difference composite measures onto the language experience composite measure. This new measure represents the portion of language experience that is not already explained by working memory, executive attention, phonological ability, or processing speed. If the results from the first analyses still hold when language experience is residualized, it suggests that language experience plays a role in target fixations *beyond* these other measures. Finally, a third analysis includes all five individual difference measures so that the contributions of the non-language predictors can be assessed.

Raw language experience. As in Study 1, the condition effect remained significant across all three time windows, with the effect again reversing direction in the post-noun window (verb: $\beta = 0.0343, SE = 0.012, p < 0.01$; noun: $\beta = 0.1697, SE = 0.021, p < 0.001$;

post-noun: $\beta = -0.1216$, $SE = 0.0226$, $p < 0.001$). Thus, the predictable condition leads to more target fixations until the latest window, when there are more fixations in the control condition. Individuals with more language experience showed more target fixations overall in the verb window, but not in the two later windows (verb: $\beta = 0.0217$, $SE = 0.008$, $p < 0.01$; noun: $\beta = 0.0055$, $SE = 0.0124$, $p = 0.6562$; post-noun: $\beta = -0.0109$, $SE = 0.0215$, $p = 0.6146$). As in the Study 1, the language experience by condition interaction never reached significance (verb: $\beta = 0.0136$, $SE = 0.015$, $p = 0.3647$; noun: $\beta = 0.0163$, $SE = 0.0158$, $p = 0.3035$; post-noun: $\beta = -0.01$, $SE = 0.0185$, $p = 0.605$).

Residualized language experience. When residualized language experience replaced the raw language experience score in the model, the same pattern of results held. The condition effect was significant across all three time windows, with the effect reversing direction in the post-noun window (verb: $\beta = 0.0358$, $SE = 0.0128$, $p < 0.01$; noun: $\beta = 0.1731$, $SE = 0.0212$, $p < 0.001$; post-noun: $\beta = -0.1217$, $SE = 0.0229$, $p < 0.001$). Again, the language experience main effect was significant in the verb window, but not in the two later windows (verb: $\beta = 0.0239$, $SE = 0.0088$, $p < 0.01$; noun: $\beta = -0.0073$, $SE = 0.0136$, $p = 0.5933$; post-noun: $\beta = -0.0127$, $SE = 0.0238$, $p = 0.5953$). As in Study 1, the language experience by condition interaction never reached significance (verb: $\beta = 0.0003$, $SE = 0.0166$, $p = 0.984$; noun: $\beta = 0.0138$, $SE = 0.0176$, $p = 0.4325$; post-noun: $\beta = -0.0235$, $SE = 0.0205$, $p = 0.2514$).

The results again seem to suggest that in the earliest window, as the verb unfolds, more language experience leads to more target fixations across conditions, although the control condition, by design, should contain no information that predicts the target. Figure 4 presents a scatterplot showing a slight positive relation between residualized language

experience scores and a subject's mean proportion of fixations to the target. As in Study 1, a follow-up analysis was performed to investigate each condition separately using dummy coding. When looking in the predictable condition, there is a marginal effect of language experience ($\beta = 0.0243$, $SE = 0.0127$, $p = 0.0896$). This effect is not significant in the control condition ($\beta = 0.0196$, $SE = 0.0128$, $p = 0.159$). As in Study 1, this suggests that the language experience effect could be driven by the predictable condition, but it is important to note that the difference in statistical significance we see between these two conditions does not imply that there is an interaction (Gelman & Stern, 2006).

Fixations to other objects. Fixations to other objects in the scene again shed light on the language experience effect seen in the verb window. As in Study 1, fixations to the agent drop off more quickly in the predictable condition ($\beta = -0.0254$, $SE = 0.0131$, $p = 0.05284$) and for individuals with more language experience ($\beta = -0.0458$, $SE = 0.0195$, $p < 0.05$) but there is no significant interaction ($\beta = -0.02042$, $SE = 0.0213$, $p = 0.338$). This suggests that both being able to predict the target and having more language experience allow the listener to process and move on from the agent more quickly. To illustrate this effect, an extreme groups comparison is plotted in Figure 5. Agent fixations of subjects from the highest and lowest quartile of language experience composite scores are plotted over time by condition.

These data along with the target fixation data suggest that more language experience allows people to look away from the agent more quickly and explore other objects. However, unlike in Study 1, there is no significant effect of condition ($\beta = 0.0045$, $SE = 0.0131$, $p = 0.7347$) or language experience ($\beta = 0.0024$, $SE = 0.0114$, $p = 0.8432$) on competitor fixations in the verb window and no interaction ($\beta = 0.0149$, $SE = 0.019$, $p =$

0.432), although an experience effect would be predicted if individuals with more experience were exploring the entire scene more.

Model with all individual differences as predictors. Although the effect of language experience is of primary importance for the current study, contributions of the other cognitive factors are of theoretical importance and the current design allows the simultaneous analysis of each. Because some of our measures were correlated, they created collinearity in our model even after centering these predictors. Therefore, the model estimates should be interpreted with caution, but the reported p-values from the model comparisons are robust to this collinearity. Model results are shown in Table 6. It is worth noting that increased inhibitory control and phonological ability lead to more target fixations, consistent with theoretical claims to their role in efficient processing (MacDonald & Christiansen, 2002; Gernsbacher, 1993, 1995; Novick et al., 2005, 2010). However, the important result for the current study is that the early main effect of language experience remains even when these other factors are included.

Discussion

The second study replicated the first, showing that language experience predicted more looks to the target across the two conditions, but only as the verb unfolded and before the noun information came online. We found no evidence for an interaction, suggesting that people with more experience are overall facilitated in interpreting the sentences but are not more likely to predict upcoming linguistic structure than those with less experience. It is possible that there was not sufficient power to detect such an interaction. Importantly, the effect of language experience still held when controlling for the other cognitive measures, suggesting that there is something specific to language

exposure that contributes to online spoken language comprehension beyond these other more general measures that have been found to explain individual differences.

Study 2 Tables and Figures

Table 3				
Summary of performance on language experience measures				
Domain	Measure	Possible Range	Observed Range	Mean (SD)
Working memory	Reading Span	Min: 0	Min: 2.783	6.684 (1.759)
		Max: 10	Max: 10	
	Listening Span	Min: 0	Min: 4.133	8.796 (1.097)
		Max: 10	Max: 10	
	Operation Span	Min: 0	Min: 1.852	10.231 (3.524)
		Max: 15	Max: 14.8333	
Perceptual speed	Letter Comparison	Min: 0	Min: 47	75.54 (14.302)
		Max: undefined ^a	Max: 126	
	Pattern Comparison	Min: 0	Min: 52	86.75 (13.427)
		Max: undefined ^a	Max: 114	
Inhibitory control	Antisaccade log reaction time	n/a ^b	Min: .1660	0.6418 (0.205)
			Max: 1.3882	
	Stroop log interference	n/a ^b	Min: -1.0378	0.2113 (0.181)
			Max: .6019	
	Flanker log interference	n/a ^b	Min: -0.1344	0.1419 (0.064)
			Max: 0.3094	
Language experience	ERVT	Min: 0	Min: 1.25	17.53 (7.387)
		Max: 48	Max: 36.75	
	ART	Min: -65	Min: -9	10.67 (11.807)
		Max: 65	Max: 47	
	CRH	Min: 7	Min: 9	22.08 (5.103)
		Max: 35	Max: 33	
Phonological ability	RTE	Min: 0	Min: 5	20.39 (10.760)
		Max: 63	Max: 63	
	NAART	Min: 0	Min: 17	33.9 (7.592)
		Max: 61	Max: 52	
Phonological ability	Phoneme Reversal	Min: 0	Min: 0.1111	0.6436 (0.169)
		Max: 1 ^c	Max: 1	
	Blending Non-words	Min: 0	Min: 0	0.5742 (0.195)
		Max: 1 ^c	Max: 0.9444	
	Gupta	Min: 0	Min: 0.4808	0.7142 (0.099)
		Max: 1 ^d	Max: 0.9183	
Notes: ^a Participants complete as many items as possible in six 20-second blocks. ^b Due to the logarithmic transformation, the range is not meaningful here. ^c Proportion correct out of 18 items. ^d Proportion correct out of 416 syllables across 96 items.				

Table 4

Correlations among cognitive domains

	Language Experience	Working Memory	Inhibitory Control	Perceptual Speed
Working Memory	0.022	--		
Inhibitory Control	0.051	0.092	--	
Perceptual Speed	0.054	0.063	0.30**	--
Phonological Ability	0.32**	0.18 ⁺	0.067	0.19 ⁺
<i>Notes:</i> ⁺ $p < 0.1$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$				

Table 5

Correlations among language experience measures

	ART	NAART	ERVT	CRH
NAART	0.45***	--		
ERVT	0.47***	0.65***	--	
CRH	0.26*	0.35***	0.38***	--
RTE	0.20*	0.12	0.17	0.45***
<i>Notes:</i> ART = Author Recognition Test, NAART = North American Adult Reading Test, ERVT = Extended Range Vocabulary Test, CRH = Comparative Reading Habits, RTE = Reading Time Estimate. ⁺ $p < 0.1$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$				

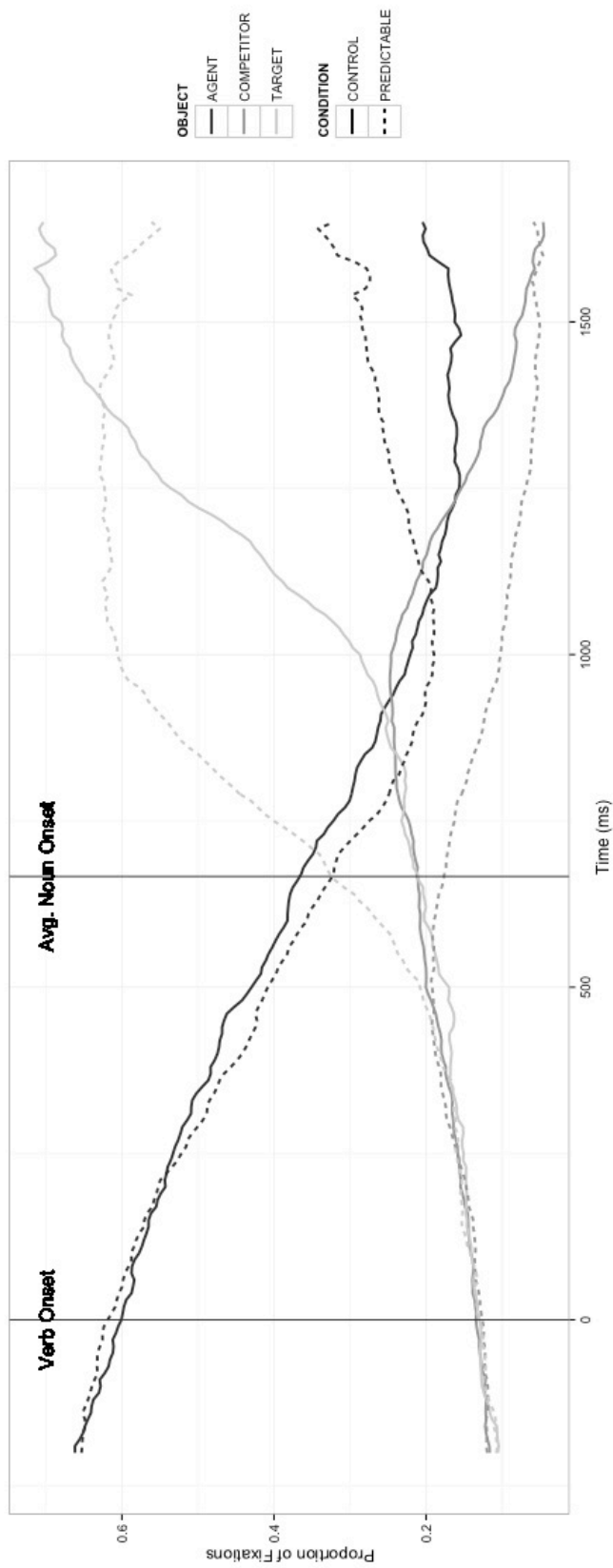


Figure 3. Proportions of fixations to the objects in the visual world display in Study 2. The verb onset is align at 0 milliseconds and the average noun onset is shown for illustration.

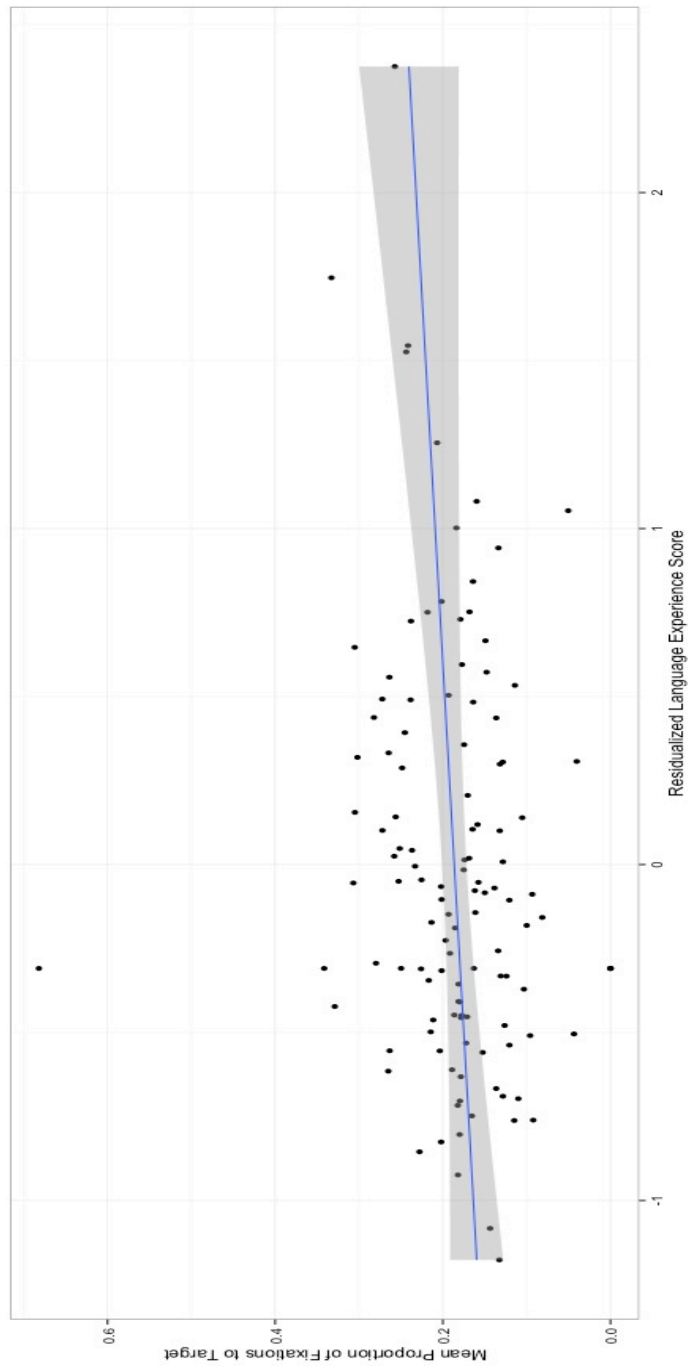


Figure 4. Scatterplot of subjects' residualized language experience scores against their mean proportion of fixations to the target, collapsed across conditions.

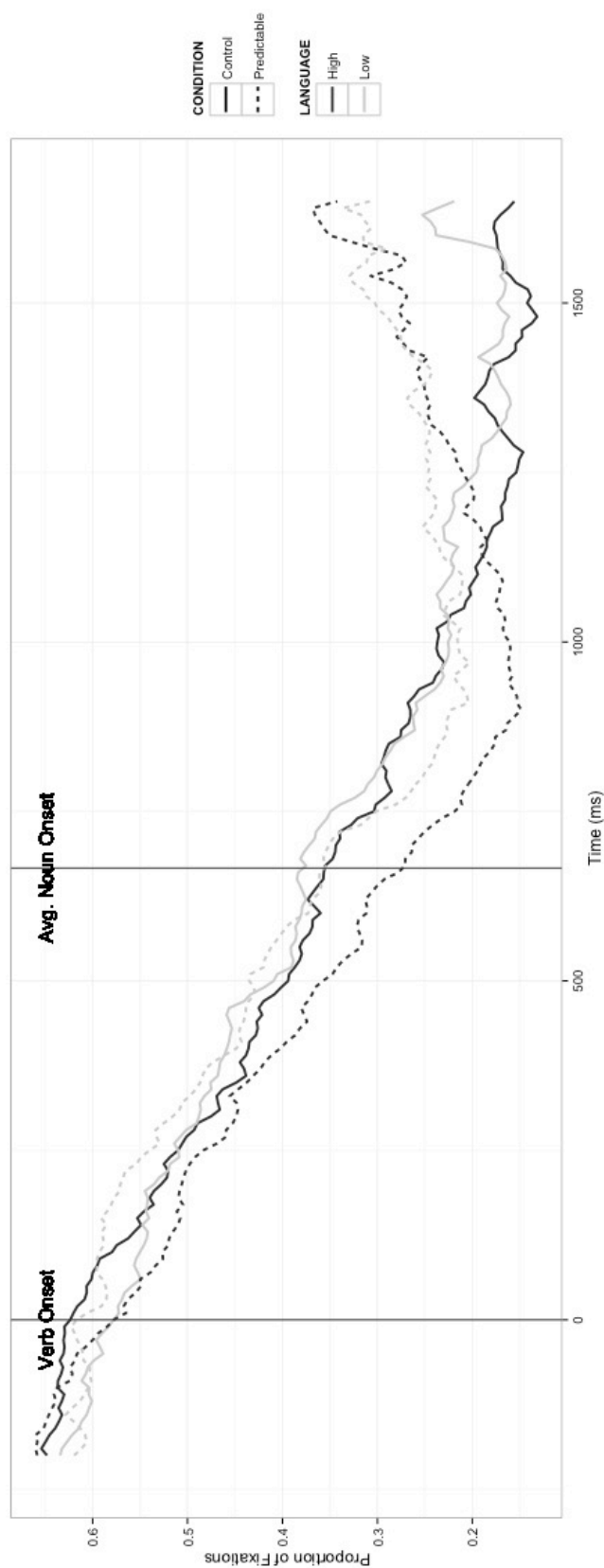


Figure 5. Proportions of fixations to the agent (e.g. boy) in the visual world display in Study 2. The verb onset is align at 0 milliseconds and the average noun onset is shown for illustration.

Table 6 Model results for a multi-level mixed effects model predicting target fixations from condition and all cognitive domains.												
	Verb region				Noun region				Post-noun region			
	β (SE)	t-value	χ^2		β (SE)	t-value	χ^2		β (SE)	t-value	χ^2	
Main Effects	Condition	0.036 (0.013)	2.838	7.1757**	0.173 (0.021)	8.139	35.881***		-0.121 (0.023)	-5.303	20.132***	
	Language	0.024 (0.009)	2.73	7.2008**	-0.007 (0.013)	-0.561	0.3156		-0.013 (0.023)	-0.548	0.2988	
	Memory	0.003 (0.007)	0.472	0.2176	-0.013 (0.010)	-1.27	1.6026		0.010 (0.019)	0.529	0.2793	
	Inhibition ^a	-0.009 (0.01)	-0.926	13.854***	-0.012 (0.015)	-0.792	18.146***		-0.026 (0.027)	-0.959	23.176***	
	Phonological	-0.004 (0.008)	-0.488	23.969***	0.015 (0.012)	1.27	39.042***		0.015 (0.022)	0.705	80.429***	
	Speed	-0.006 (0.006)	-0.985	0.9708	0.022 (0.009)	2.357	5.4208*		0.015 (0.017)	0.888	0.7876	
	Language	0.0005 (0.016)	0.027	0.0007	0.013 (0.017)	0.775	0.6005		-0.023 (0.020)	-1.129	1.2752	
	Memory	0.010 (0.013)	0.781	0.6099	0.006 (0.014)	0.435	0.1891		-0.000 (0.016)	-0.022	0.0005	
	Inhibition	0.020 (0.019)	1.063	1.1304	-0.018 (0.020)	-0.907	0.8222		0.004 (0.024)	0.168	0.0281	
	Phonological	0.015 (0.015)	0.969	0.9379	-0.008 (0.016)	-0.466	0.2169		0.036 (0.019)	1.896	3.5922+	
Interaction with condition	Speed	0.019 (0.012)	1.6	2.5599	0.008 (0.013)	0.613	0.3754		-0.011 (0.015)	-0.722	0.5208	
Notes: ^a Because higher scores in this task reflect more inhibition, a negative effect here suggests that more inhibitory control leads to more target fixations. + $p < 0.1$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$												

CHAPTER 4: GENERAL DISCUSSION

Taken together, the two studies described here suggest that language experience is associated with the efficient recognition of spoken words and their integration into a given visual context. Therefore, there is evidence for an overall benefit of experience that is not limited to contexts where prediction of the upcoming referent is effective. These findings support probability- and frequency-based theories of processing that predict effects of idiosyncratic experience on language processing. Further, this work is unique in demonstrating that the effects of reading experience generalize to natural spoken language while simultaneously controlling for theoretically important cognitive skills, each measured by multiple tasks. All of this points to the ability of language experience to capture unique variability in spoken comprehension, even within a literate population that represents a restricted range of language ability.

Returning to the Mishra and colleagues (2012), it is possible that their results are consistent with our evidence for a more general benefit of experience. Highly literate adults showed facilitation relative to low-literate adults in trials that licensed prediction, in-line with the experience effect shown here. However, their study did not include control trials without predictive information, and it is possible that the literate adults would have been facilitated on these trials as well. Mishra and colleagues suggest that individuals of higher literacy have “fine-tuned” their anticipatory mechanisms through practice with reading and writing. In light of our current evidence, it could be the case that this practice also fine-tunes other mechanisms important for efficient processing that facilitate comprehension on the non-predictable trials.

Although we found no evidence of an experience by condition interaction, it is

unlikely that experience has no bearing on predictive processing. It may be that the measure of prediction that we used was not sensitive enough to detect effects of experience. Our failure to find random slopes on condition for subjects supports this, as this indicates that there is not sufficient variability in participants' predictability effect to justify a random effect to capture it. Furthermore, assuming the interaction has the same or a smaller effect size as the main effect, we necessarily have less power to detect the interaction than the main effect (Smith & Day, 1984). While the dummy coding results suggest that the main effect is driven by the predictable condition, a significant interaction is needed in order to conclude that the experience effects in these two conditions are indeed statistically different (Gelman & Stern, 2006).

It is also unlikely that experience has no bearing on top-down processing more generally. The current study was designed to measure a specific kind of top-down influence on processing: the use of semantic information in a verb to anticipate the most likely object. However, there is possibly other top-down information available in the stimuli we provided that were not controlled in this study, such as co-articulatory information or affordances of the visual scene itself. Future work could attempt to disentangle these effects.

Individuals with more experience may be more likely to try to make sense of the visual scene, exploring the various objects to anticipate what will be referred to. They may be facilitated in processing the sentence as it unfolds due to bottom-up word recognition processes, and, having processed the unfolding sentence more easily, participants might have resources free to search for the upcoming target in the scene as the sentence continues. The current study found support for this in looks away from the agent, which

was always the subject of the accompanying sentence (e.g. “The *boy* will...”). Individuals with more language experience look less at the agent while processing the verb across conditions (e.g. “*eat/move*”).

Another potential explanation of this effect is that individuals with more experience are simply more motivated to complete the task. It is possible that individuals with more language experience are more motivated to complete the eyetracking task, and make more of an effort to find ways to predict the target. Under this hypothesis, the effect may go away if participants were no longer performing an explicit judgment task (although they may still implicitly consider other task goals; see Salverda, Brown, & Tanenhaus, 2010 for discussion). We find this explanation unlikely since highly motivated individuals would likely try harder (and be more likely to succeed) at the other measures, and we only found language experience task performance to correlate with phonological processing tasks, the other group of explicitly language-oriented tasks.

An important aspect of this work is the demonstration of a link between measures of spoken comprehension and measures of language experience that are mostly related to reading, and are not themselves measures of comprehension. While listening and reading comprehension have long been found to be correlated (Abrams, 1966), it is not clear from the current work how listening comprehension and general written language exposure are related to one another. The potential mechanism that motivated this work is that experience with reading benefits auditory comprehension by providing the processing system with information about the probabilities of the language, leading to efficient use of this information to guide comprehension. This assumes that the language processing system applies knowledge gained in the written modality and applies it to the spoken

domain. Of course, this explanation does not exclude the possibility that *auditory* experience and listening comprehension influence *reading* experience. Given that phonological ability both predicted eye movements and correlated with the language experience measures, it is possible that there is a link between reading and listening experience: phonological ability facilitates word recognition during listening, but also promotes efficient reading (Stanovich & West, 1989; MacDonald & Christiansen, 2002; Acheson & MacDonald, 2009, 2011). Efficient readers may then gain more reading experience (see Matthew effects discussed earlier; Stanovich, 1986), which provides a benefit in auditory processing over and above that provided by increased phonological skills.

While questions remain open regarding the mechanism linking more experience to performance in this auditory task, the current work makes two important contributions to the study of individual differences in sentence processing. First, we found that experience with language, largely related to *reading* experience, predicts online performance in the *auditory* domain, which speaks to the potential general benefit of exposure to various linguistic contexts. Second, by measuring a variety of other constructs previously involved in individual differences research, we were able to demonstrate a benefit of exposure that goes beyond more general cognitive mechanisms. These results reinforce the importance of literacy education, suggesting that reading skill may generalize to listening skills.

REFERENCES

- Abrams, A. G. (1966), The relation of listening and reading comprehension to skill in message structuralization. *Journal of Communication*, 16: 116–125.
- Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, 135, 50-68. doi: 10.1037/a0014411
- Acheson, D. J., & MacDonald, M. C. (2011). The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. *Journal of Memory and Language*, 65, 193-207. doi: 10.1016/j.jml.2011.04.006
- Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior research methods*, 40(1), 278–289.
- 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* 1,* 2,* 3,* 4,* 5. *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.
- Barr, D.J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: a revision of the National Adult Reading Test. *The Clinical Neuropsychologist*, 3(2), 129–136.

- Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological review*, 74(1), 1.
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769-786.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 106(2), 633-664.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, 8(8), 1117-1121.
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational Testing Service*. Retrieved from http://www.ets.org/Media/Research/pdf/Manual_for_Kit_of_Factor-Referenced_Cognitive_Tests.pdf
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior*, 20(6), 641-655.
- Eriksen, C. W. (2007). The flankers task and response competition: A useful tool for investigating variety of cognitive problems. *Visual Cognition*, 2:2-3, 101-118.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, 16(1), 143-149.

- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491-505.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469–495.
- Fine, A. B., & Jaeger, T. F. (2011). Language comprehension is sensitive to changes in the reliability of lexical cues. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 925-930). Austin, TX: Cognitive Science Society.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one*, 8(10), e77661.
- Frazier, Lyn (1987), "Sentence processing: A tutorial review", in Coltheart, M., *Attention and Performance XII: The Psychology of Reading*, Lawrence Erlbaum Associates.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science*, 4, 294-298.
- Gernsbacher, M. A. (1995). The Structure-Building Framework: What it is, what it might also be, and why. In B. K. Britton, & A. C. Graesser, (Eds.), *Models of text understanding* (pp. 289-311). Hillsdale, NJ: Erlbaum.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain:*

- Papers from the first mind articulation project symposium* (pp. 94-126). Cambridge, MA: The MIT Press.
- Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology*, 56A, 1213-1236. doi: 10.1080/02724980343000071
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL* (Vol. 2, pp. 159-166).
- Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, 136(2), 217.
- Huetting, F., & Altmann, G. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96(1), B23-B32.
- Hussey, E. K., & Novick, J. M. (2012). The benefits of executive control training and the implications for language processing. *Frontiers in psychology*, 3.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87, 329-354.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, 2(4), 647-670.
- Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124(1), 66-71.

- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130, 169-183. doi: 10.1037//0096-3445.130.2.169
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic bulletin & review*, 9(4), 637-671.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of experimental psychology: General*, 132(1), 47.
- Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, 68, 42-73. doi: 10.1016/j.jml.2011.03.002
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177. doi: 10.1016/j.cognition.2007.05.006
- Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, 130(2), 199.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comments on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109, 35-54. doi:10.1037//033-295X.109.1.35
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, 101(4), 676.

- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843.
- Mishra, R., Singh, N., Pandey, A. & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: evidence from Indian low and high literates. *Journal of Eye Movement Research*, 5(1):3, 1-10.
- 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, 906-924. doi: 10.1111/j.1749-818x.2010.00244.
- Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34, 521-542.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3), 191-201.
- Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental psychology*, 27(5), 763.
- Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta psychologica*, 137(2), 172-180.

- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 402-433.
- Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic bulletin & review*, 18(2), 371-376.
- Stine, E. A. L., & Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging and Cognition*, 1, 2-16.
doi:10.1080/09289919408251446
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: a psychometric approach. *Journal of Experimental Psychology: General*, 136(1), 64.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498-505.
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1302.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological bulletin*, 101(2), 192.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing: CTOPP*. Pro-Ed.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, 103, 761-772.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250-271. doi: 10.1016/j.cogpsych.2008.08.002

APPENDIX A: SENTENCE STIMULI FOR THE EYETRACKING TASK IN STUDY 1

Critical trials use the same sentences and distractor objects as in Altmann and Kamide (1999) unless otherwise noted, with the former version in brackets. Nonsensical items are indicated with italics.

1. The boy will eat/move the cake. (toy car, ball, toy train).
- *2. The woman will drink/try the wine. (cheese, lipstick, perfume [chair], *plant*).
3. The policeman will arrest/search the man. (car, dustbin, houses, *cat*).
4. The woman will bathe/touch the baby. (plant, rocking-horse, stool).
5. The boy will bounce/throw the ball. (paper plane, shuttle-cock, acorns, *bicycle*).
6. The hiker will climb/photograph the mountain. (mountain lion, moon, cactus).
- *7. The housewife will fry/grab [wash] the shrimp. (knife, jug, weighing scales).
8. The doctor will inject/check the child. (TV monitor, microscope, books, *toy bear*).
9. The woman will play/dust the piano. (table, television, telephone).
10. The woman will read/open the book. (door, bag, jar, *cup*).
- *11. The man will repair/wipe the washing machine. (mirror, goggles [waste bin], wrench [dog]).
12. The baby will ring/kick the bell. (drum, bricks, duck).
13. The man will sail/watch the boat. (birds, car, sun).
14. The man will smoke/collect the cigarettes. (binder, glasses, briefcase, *clock*).
15. The boy will walk/feed the dog. (bird, pig, hen, *ball*).
16. The businessman will wear/forget the hat. (wallet, folder, magnifying glass, *chair*).

Sentences and distractor objects for the filler trials were invented for the current study. Following Altmann and Kamide (1999), they were designed such that either none, one, two, or three of the items in the scene made sense given the preceding context. Nonsensical items are indicated with italics. As in the critical scenes, there were either four or five objects in the scene and an agent.

1. The baby will lift the rattle (*rocking chair, grandfather clock, bureau, trunk*)
2. The woman will rip the page (*ice cream, egg, phone, bowling ball*)
3. The woman will sew the dress (*stool, clock, laptop, tomato, hat box*)
4. The man will shut the window (*puppy, vacuum, grapes, pumpkin, oven mitt*)
5. The man will shake the maraca (*salt, toaster, cheese grater, coffee maker*)
6. The maid will fold the slacks (*shirt, cup, stapler, vase*)
7. The man will stir the coffee (*soup, banana, money, tissue box, lamp*)
8. The child will taste the candy (*cookie, yo-yo, teddy, bench, parrot*)
9. The girl will ride the unicycle (*horse, scooter, shrub, football*)
10. The teenager will zip the hoodie (*boot, jacket, sandwich, newspaper*)
11. The silversmith will polish the spoon (*teapot, watch, child, couch, pillow*)
12. The man will shatter the vase (*jar, bottle, hotdog, purse, onion*)
13. The farmer will hear the siren (*rooster, telephone, radio, bread*)
14. The girl will pet the puppy (*kitten, bunny, hamster, pie*)
15. The musician will tune the saxophone (*violin, guitar, banjo, cabbage, bag*)
16. The man will sip the martini (*tea, milkshake, beer, candle, basket*)

APPENDIX B: SENTENCE STIMULI FOR THE EYETRACKING TASK IN STUDY 2

An additional 32 stimuli, half fillers, were created according to the same guidelines as Study 1, with none of the target objects or verbs repeated.

Critical items:

1. The worker will chop/carry the wood (nails, ladder, box).
2. The student will sharpen/sort the pencils (paper clips, markers, push pins, *juice box*).
3. The vendor will slice/sell the meat (pasta, pop, corn).
4. The boy will peel/enjoy the orange (chocolate, crackers, taco, *lunchbox*).
5. The athlete will dribble/take the basketball (towel, water bottle, gym bag, *basketball hoop*).
6. The girl will fly/choose the kite (baby carriage, clown doll, top).
7. The woman will water/keep the flowers (watch, mascara, teapot).
8. The thief will drive/steal the van (jewels, stereo, necklace, *spider*).
9. The cleaner will flush/scrub the toilet (sink, bathtub, tiles, *mouthwash*).
10. The chef will roast/prepare the chicken (salad, rice, sundae, *rolling pin*).
11. The cowboy will tie/hold the rope (horseshoe, saddle, gun, *puddle*).
12. The woman will spend/use the money (nail polish, pen, lamp).
13. The child will blow/see the horn (baseball bat, toad, chalk).
14. The dog will chase/notice the squirrel (steak, fire hydrant, food dish).
15. The boy will stack/arrange the cards (marbles, balloons, lemons, *mitten*).
16. The man will fill/clean the pot (plate, spatula, tongs).

Filler items:

1. The boy will crush the can (*anvil, safe, dumbbell, iron, pipe*).
2. The woman will close the curtains (open cereal box, *light bulb, tooth brush, fork, lighter*).
3. The man will finish the chili (cupcake, baked potato, popcorn, *cane*).
4. The dog will bury the bone (slippers, chew toy, *flower, flag, dog house*).
5. The girl will lace the skates (*hula hoop, slinky, Frisbee, globe, leaf*).
6. The man will hug the grandpa (grandma, boy, young woman, *Christmas tree, fireplace*).
7. The boy will smell the pizza (bacon, *lock, toy soldier, ruler*).
8. The bully will tease the girl (*catcher's mitt, sling shot, rock, helmet*).
9. The waitress will serve the fries (pancakes, onion rings, apron, *note pad*).
10. The hunter will shoot the rabbit (deer, duck, bear, *snowman*).
11. The CEO will sign the checks (forms, *potted plant, statue, framed picture*).
12. The gardener will pick the tomatoes (raspberries, peppers, roses, *caterpillar, hose*).
13. The cat will scratch the boy (chair, scratching post, *peanut, pin*).
14. The girl will bend the straw (*peanut butter, yogurt, pear, rolling pin*).
15. The butler will iron the jeans (skirt, blazer, *belt, shoes*).
16. The fireman will rescue the cat (toddler, *scissors, umbrella, sock, battery*).