LANGUAGE-VISION INTERACTIONS IN THE VISUAL WORLD PARADIGM: EXAMINING THE EFFECTS OF DISPLAY AND WORKING MEMORY

BY

SAJIN STANISLAV

BBA, SUNY ALFRED, 2010
MS, BINGHAMTON UNIVERSITY, 2013

DISSERTATION

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Cognitive and Brain Sciences
in the Graduate School of
Binghamton University
State University of New York
2016

Accepted in partial fulfillment of the degree requirements for
The degree of Doctor of Philosophy in Cognitive and Brain Sciences
In the Graduate School of
Binghamton University
State University of New York
2016

May 2, 2016

Dr. Cynthia Connine, Chair
Department of Cognitive and Brain Sciences, Binghamton University

Dr. Albrecht Inhoff, Member
Department of Cognitive and Brain Sciences, Binghamton University

Dr. Sarah Laszlo, Member
Department of Cognitive and Brain Sciences, Binghamton University

Dr. Lijun Yin, Outside Examiner
Department of Computer Science, Binghamton University

Abstract

Visual World Paradigm (VWP) is commonly used to study a variety of psycholinguistic topics. Nonetheless, as pointed out in several reviews (Huettig, Mishra, & Olivers, 2012; Huettig, Olivers, & Hartsuiker, 2011a; Salverda, Brown, & Tanenhaus, 2011), using the VWP to study language mediated eye-movements should properly account for both linguistic and non-linguistic cognitive factors. In this dissertation, I review studies from the visual search and visual world literature indicating that linguistic representations can be accessed based on visual information and that working memory plays an important role in guiding visual attention. An additional theme that is examined is whether visual search in the VWP is driven by just-in-time processing or from a sparse visual working memory buffer. Following the review of the literature, five experiments examine (1) the pattern of eye movements to printed word displays during the preview period (Experiment 1) , (2) if competitor effects can be guided by linguistic representations stored in working memory rather than online processing of spoken utterances (Experiment 2), (3) if cognitive load and non-linguistic representations stored in working memory affects language mediated eye-gaze (Experiment 3 & 4), and (4) the influence of display size (Experiment 5). The results of these experiments point that linguistic and non-linguistic representations stored in working memory play an important role in guiding eye-movements in the VWP, suggesting that the paradigm can be adapted for a variety of other cognitive tasks (e.g., semantic priming, memory recognition). Moreover, the results of these experiments will be used to argue for a more complex understanding of what drives eye-movements in the VWP. In particular, I will argue that task parameters could affect the mapping of language to visual referents through either a memory-less visual search or a memory-dependent search.

Dedication

This document is dedicated to my loving wife and parents.

I love you mostestest.

TABLE OF CONTENTS

List of Tables

List of Figures

Introduction

An impressive characteristic of human cognition is that humans are able to process and integrate information across multiple modalities seemingly without effort. For instance, when someone asks to “pick up the pen on the desk,” we can effortlessly orient our eyes towards the pen on the desk within milliseconds of hearing its label. Even more impressive is that we can construct affordances based on the context in which the utterance is presented: If the table has only one pen and some coffee spill, our eye-movements will orient towards the pen the moment we hear the verb “pick” instead of orienting towards the coffee spill. The assumption that visual search is mediated by activation of linguistic representations stand at the foundation of using VWP to study language processing.

Over the past two decades, the VWP paradigm (Cooper, 1974; Tanenhaus et al., 1995), has received significant prominence in psycholinguistics. The popularity of this paradigm largely stems from the fact that it is believed to reveal the time-course of spoken language processing and does not require participants to engage in meta-linguistic decision making (e.g., lexical decision task). In the VWP, participants are presented with speech input as they explore a visual display of semi-realistic scenes, arrays of objects, or printed words. The paradigm has been used to study syntactic ambiguity processing (Ferreira, Foucart, & Engelhardt, 2013), speech perception (Allopenna, Magnuson, & Tanenhaus, 1998; McQueen & Viebahn, 2007; Sajin & Connine, 2014; Tobin & Cho, 2010), semantic, thematic, and visual shape competition effects (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005, 2007; Huettig & Hartsuiker, 2008; Kalénine, Mirman, Middleton, & Buxbaum, 2012; Rommers, Meyer, & Huettig, 2013; Yee, Huffstetler, & Thompson-Schill, 2011; Yee & Sedivy, 2006), prediction in sentence processing (Altmann & Kamide, 1999, 2007, 2009), and individual differences in language processing (Mirman & Graziano, 2012), among other topics in psycholinguistics (see Huettig, Rommers, & Meyer, 2011; Salverda, Brown, & Tanenhaus, 2011; Michael K. Tanenhaus & Trueswell, 2006 for reviews).

Although the VWP has been used predominantly to study speech processing by examining how visual attention is guided by the unfolding of the speech signal, a limitation of the VWP is that the mapping of the speech input interacts with the features of the visual display (set size, using objects vs. words as referents, preview duration). In other words, the properties of the display enable participants to engage in visual processing that could, in turn, affect linguistic processing. For instance, Sajin and Connine (2014), when examining how semantic information affects spoken word recognition, have observed that listeners use semantic information when presented with displays that contain phonological competitors, but don’t rely on semantic information for displays that present the target without any phonological competitors, indicating that speech processing is contingent on the type of information present in the display (but see Dahan, Magnuson, & Tanenhaus, 2001 who found an effect of frequency in displays with no competitors). This issue makes it particularly challenging to study the effects of competition for displays that contain phonological competitors, since certain parameters of the task (e.g., using objects instead of words) lead to different levels of representation (e.g., semantic, lexical, visual) being activated, thus limiting our understanding of how listeners eye-movements map to the visual display while processing a spoken utterance (Huettig & McQueen, 2007).

Themes

Through the experiments presented in this dissertation, I plan to convince the reader that the VWP provides a much broader venue through which one can explore how language interacts with visual processing and working memory. The strengths of the paradigm do not lie only in understanding how speech is mapped onto visual referents, but in understanding how vision, attention, memory, language, and task goals interact, thus providing very rich prospects for studying cognition in general rather than only language processing.

One theme that I will discuss in this dissertation is the distinction between just-in-time (or memory-less) processing and processing that is driven by binding of representations in working memory. One assumption behind binding representations in working memory is that when we look at a display with objects or scenes, we immediately make hypotheses about how things are going to interact or how objects are related to each other lexically, visually, semantically, or syntactically (Ferreira et al., 2013; Huettig, Olivers, et al., 2011a; Knoeferle & Crocker, 2006, 2007a, 2007b). For instance, a model proposed by Huettig, Olivers, and Hartsuiker (2011) suggests that the information in the display allows the observer to access from long-term memory semantic and visual-from representation of objects and tie them to specific spatio-temporal indices in working memory. Similarly, hearing the spoken input allows the listener to access linguistic and non-linguistic representations, which are then mapped on the pre-activated representations in working memory. The details of this account will be discussed later, nonetheless I will point out that the account heavily relies on working memory as the nexus where representations are bound together. In other words, the mapping of the spoken input and pre-activated representations from visual input is done in working memory. On the other hand, a slew of research in visual search suggests that visual working memory is generally fairly poor, with a buffer of 3-4 items for tasks that involve active search for very simple objects, such as colored geometrical shapes or line drawings (see Horowitz & Wolfe, 1998; S J Luck & Vogel, 1997; Steven J. Luck & Vogel, 2013; McCarley, Wang, Kramer, Irwin, & Peterson, 2003). Moreover, work in hand-eye coordination tasks (Ballard, Hayhoe, & Pelz, 1995; Hayhoe & Ballard, 2005; Horowitz & Wolfe, 1998; Sprague, Ballard, & Robinson, 2007) indicates that it often more effective, computationally, to not create working memory representations when the costs of making saccadic eye-movements is very low and one can simply engage in just-in-time processing. The analogy here would be that one does not need to engage in the arduous task of remembering the birthdays of each of their friends, when Facebook provides a fairly simple access to that information without any effort. Therefore, eye-movements are not necessarily driven by the mapping of the spoken input to internal working memory representations. Instead, when the display is constantly present in front of the participant, the visual world acts as an external memory buffer, and the mapping is done serially, as participants move their eyes from one object to another. For instance, work done by Wolfe, Klempen, and Dahlen (2000) indicates that participants prefer to not memorize contents of the display when performing visual search. Wolfe, Klempen, and Dahlen (2000) presented to participants a display filled with a 4 to 8 geometric shapes of different colors (e.g., red circle or white square). The display remained constantly on the screen from trial to trial, thus allowing participants to become familiarized with the contents of the display. The only thing that changed from trial to trial was the specific target that participants were required to find (e.g. on one trial the target could be a red circle and on the next it could be a white square). It was expected that over the course of the experiment, participants would memorize the objects that they were searching and thus show an improvement in search times. After all, a memory-driven search allows one to quickly orient their eyes to the location of the memorized object instead of engaging in a serial, object-by-object scan. Remarkably, the results indicate that even after 350 trials, participants search times were as slow as at the beginning of the experiment[1], indicating that they engage in a slower, serial, memory-less search.

Serial search should lead to competition effects that are not closely time-locked to the spoken language onset because there is no pre-processing advantage in which the object features and their locations are memorized. Indeed, work by Andersson, Ferreira, and Henderson (2011) that indicates that with more complex scenes, with several dozens of objects in them, the time-locking of language to eye-movements gets delayed by about 2s. With complex scenes, participants simply don't have the capacity to store every object and their location in working memory, so the search times are less efficient. In other words, search in a complex display is

[1] To be more precise, they found no improvement in search times in four experiments.

necessarily serial. Moreover, if participants are presented with the auditory labels for multiple objects within a sentence, a recency and primacy effect is observed, where fixations are allocated only to objects that are heard first or last (Andersson, Ferreira, & Henderson, 2011). This indicates that not every label (i.e., word) we hear triggers an automatic movement to an associated object. Intuitively, the lack of time-locking and selective allocation of visual attention has good ecological appeal. After all, it would be fairly difficult to live in a world where language-driven eye movements are entirely automatic, since such automaticity would presume that listeners get distracted by every single label they hear in a spoken utterance. Instead, it is more likely that listeners are more selective in regard to how they orient their eye-movements when hearing spoken language, and that this selectivity is influenced by the complexity of the scene and goals of the task.

The distinction between just-in-time mapping and working memory mapping is important, since it speaks to the time-course of how speech is mapped to visual referents. For example, if a participant is given a preview of four objects in a display, from which they can extract visual, semantic, and phonological information that is bound to specific locations in visual working memory, then they could use the information in working memory to suppress saccades to objects that mismatch spoken input that is presented at a later time. If initially participants are given a display with the words BUCKLE, BUCKET, DIESEL, and KICKER, which they spatially store in working memory, then at the presentation of the spoken target BUCKLE, they should be able to immediately suppress saccades to words mismatching the target, thus having fewer fixations to the words DIESEL and KICKER right from the onset of the spoken utterance. In other words, listeners would never look at the locations of distractors because they have eliminated those distractors as potential targets in their working memory. On the other hand, if the search proceeds more serially and does not depend on working memory representations, participants would have to look at distractors before eliminating them as potential targets. This will lead to delays in how quickly the competitor and the distractors diverge from each other. The divergence periods between competitor and distractors in Experiments 1,2, and 5 will allow us to answer if listeners are using working memory to bind linguistic representations to their locations when guiding visual attention or if listeners likely engage in just-in-time processing.

The second theme of this dissertation focuses on the competition processes during lexical access that are observed in the VWP, with a large emphasis on testing the linking hypothesis (Salverda et al., 2011; M K Tanenhaus, Magnuson, Dahan, & Chambers, 2000). First, later in the dissertation, I will provide some historical coverage for the linking hypothesis account. I chose to present developments related to the linking hypothesis chronologically because this account has undergone some major changes over the last 15 years. An understanding of how this hypothesis has changed over the years will enable one to better assess which parts of the linking hypothesis remain viable and which parts require further exploration. The data from experiments 2,3,4 will be used to argue that competition effects can be observed during typical visual search and that phonological competition can be disrupted by non-linguistic representations stored in working memory (thus putting into question just how automatic are the competition effects in the VWP). These experiments have been inspired by work on visual search (Görges, Oppermann, Jescheniak, & Schriefers, 2013; Meyer & Damian, 2007; Soto, Heinke, Humphreys, & Blanco, 2005; Soto & Humphreys, 2007; Woodman & Luck, 2007b) and by a thoughtful review written by Tanenhaus and Trueswell (2005), in which they acknowledge that the discrete nature of saccadic movements does not allow one to examine competition effects during a single trial. Only by aggregating the fixation proportions across multiple trials, we are able to observe a sense of continuity. Thus, describing the VWP as a continuous measure of processing is somewhat a simplification, since the measurements that we use are essentially discrete. The visual search paradigm also deals in discrete measurements. Fundamentally, eye movements from visual search tasks and VWP are the same, and my suspicion was that aggregating and plotting fixations across time in the visual search paradigm will lead to similar perception of continuous processing. For instance, one could think of fixation proportions between a target and competitor in the VWP not as a measure of ongoing competition, but instead as a measure of how long it takes the listener to disengage the word they are currently fixated on. If the initial fixation falls on the correct target, then there is no reason for the subject to disengage and look at another word. Instead, they simply have to wait for a bit more spoken input to confirm the target. If the eyes fall initially on the distractor, they can immediately disengage from it and search for a matching target. If the eyes fall on a competitor, the amount of time needed to disengage the eyes and start a new search will be a bit longer. Fixations that have this pattern of engagement and disengagement will lead to similar fixation curves that are observed in a typical VWP. The same pattern of eye-movements nonetheless happens when one carries a visual search task, where the participant has to identify the target based on two conjunction features, while the display contains a target, a competitor sharing one of those two features and several unrelated distractors. Under the second theme, I will also examine if phonological competition can occur in the absence of spoken input, and only from the activation of the labels presented during the preview period. I will present evidence from the literature indicating that a preview period of as long as 1s is enough for a participant to access the phonological label from an object.

The data in this dissertation does not suggest that working memory account for information binding is inaccurate or that the linking hypothesis is incorrect. Rather the experiments presented here demonstrate some challenges that these accounts face and brings to light the fact that eye-movements in the VWP are not just a result of simple mapping of speech to objects/words. Some of these challenges have been acknowledged by the proponents of these accounts (Huettig, Olivers, et al., 2011a; Ramesh K. Mishra, Olivers, & Huettig, 2013; Salverda et al., 2011).

Phonological or orthographic competition effects?

VWP studies using printed word displays often refer to phonological competition when in reality competitor words in the display have both shared onset orthography and onset phonology (Brouwer, Mitterer, & Huettig, 2012; Gauvin, Hartsuiker, & Huettig, 2013; Huettig & McQueen, 2007, 2011; McQueen & Viebahn, 2007; Sajin & Connine, 2014). Taking into account that phonological competitor effects are observed with pictures, it is generally implicitly assumed that competitor looks towards a word like *beaker* when hearing *beetle* are due to ambiguity in phonological onset overlap rather because they share the first two letters. Part of the reason why experimenters use words that have both shared onset phonology and orthography is because for the vast majority of the words in English onset orthography and phonology tend to go hand in hand. Nonetheless, evidence from Salverda and Tanenhaus (2010) indicates that language-vision mapping in VWP might occur at orthographic rather than phonological level. In Experiment 1, using printed word displays, Salverda and Tanehnhaus (2010) manipulated the amount of phonological overlap while maintaining the amount of orthographic overlap. For instance, they presented the target *bead* in the presence of a high overlap phonological competitor *bean* or low overlap phonological competitor *bear*. The result of this experiment indicates that the amount of phonological overlap did not affect the strength of the competition process. In Experiment 2, however, where the phonological overlap was controlled by using homophones and orthographic overlap differed (e.g., target *bead* presented either with high orthographic overlap competitor *bear* or low orthographic competitor overlap *bare*), they found that competition was stronger for words with high overlap orthography. Subsequent studies, where subjects had to learn an artificial language in which letter-sound mappings are mismatched to English (e.g., participants learned Colbertian alphabet, where, for example, grapheme *v* stood for sound /t/[2]), found that competition is driven by orthographic rather than phonological overlap (Bartolotti, Daniel, & Marian, 2013). Salverda and Tanenhaus (2010) suggest that orthographic information is accessed from the unfolding of the spoken input and is used in mapping spoken referents to printed words. Although this explanation is entirely plausible, a simpler explanation exists: that orthographic information is directly extracted from the visual display, and eye-movements reflect both phonological

[2] An example of this occurring is when an English speaker learns that grapheme v found at onset of words in German is pronounced as /f/ rather than /v/.

processing of the spoken input as well as visual orthographic processing (see more on how orthographic similarity influence competition effects in VWP, Gregg, 2014). These studies indicate that competition effects reported in the literature might not be due to online, automatic phonological competition, but rather due to orthographic similarity of objects present in the display or a combination of orthographic and phonological competition processes. Dissociating the effects of phonology from orthography in VWP is fairly difficult, since one has to deal with a much smaller sample of words composed of homophones or have subjects learn an artificial language with grapheme-phoneme mismatches conflicting to English. Nonetheless, it is important to acknowledge that in display with printed words, phonology onset and orthography onset is confounded. In the items used for this dissertation, competitor words in printed word displays share at least two letters and two phonemes at the onset, similarly to what is commonly done in the vast majority of VWP literature that uses printed words except for Slaverda and Tanenhaus (2010) and Bartolotti and colleagues (2013).

Access to linguistic information from preview of the visual display

In the VWP, participants are generally given a preview of the display before the auditory target is presented. The preview is usually shorter for displays with printed words than for displays with pictures. The duration of the preview can have important implications with regard to what type of information participants are able to extract before they hear the spoken utterance. There are at least four reasons why the preview is important, and all of the four reasons specify that during the preview listeners actively extract and process the information in the visual display. These four reasons are not entirely independent of each other. The reasons are 1) preview allows implicit naming to occur, 2) preview affects the type of representations participants can extract from the visual display, 3) preview duration can affect pre-activation and storage of representation in working memory that will dictate the speed of visual search, and 4) preview affects pre-target processing, leading to abnormalities observed when comparing the baseline of fixation curves.

The first reason specifies that participants might be implicitly naming the words/pictures and thus the phonological competitor effects might be due to this implicit naming process rather than because of online phonological competition driven by the ambiguity in the spoken word. In other words, the visual display, by allowing participants to access linguistic information about an object through implicit naming, could be used to guide visual attention to referents that share the same spoken onset before the spoken target is even heard. The evidence that participants are able to access the names of objects is very robust in the domain of visual processing. Zelinsky and Murphy (2000) tracked participants' eye-movements as they looked at four pictures in the display. In some cases, the pictures depicted objects with one-syllable names (e.g., ball) while other objects had multiple-syllable names (e.g., elephant). Participants fixated longer towards objects with multi-syllable names than one-syllable names, despite the fact that they were not required to name any of the objects and simply preview them[3].

More compelling evidence that linguistic information can be accessed by naming visual referents has have been previously found in a task called the visual search paradigm (see for discussion between VWP and visual search in Huettig, Olivers, & Hartsuiker, 2011). In visual search paradigm, participants are given a target template and then are asked to look for it in a display of objects, which in some cases contain the target and in some cases don't have the target. Importantly, in the visual search paradigm, participants are given the target template before the display appears. Meyer, Belke, Telling and Humphreys (2007) found that participants took longer to decide if a target was present/absent if the display contained the picture of a homophone related to the target (see Figure 1 for the trial structure). When the search display has a competitor

[3] The effect of syllable number has been heavily understudied/controlled in the VWP with pictures. Out of all the studies cited in this dissertation that used VWP with pictures none have properly controlled syllable length within a display (13 papers don't report norming on syllable number whatsoever and 2 papers report the means across target and competitor rather than norm for syllable number within display). For studies with printed word displays, 4 papers controlled for syllable length within display and 4 others haven't done so. Zelinsky and Murphy (2000) report a 133ms difference in gaze duration between 1-syllable and 3-syllable words. The large size of this effect indicates that controlling for syllable length in VWP items should probably merit more attention than it is currently given.

buoy together with the target *boy*, participants took longer to indicate the presence/absence of the target. This indicates that phonological information becomes rapidly available when allocating visual attention despite the absence of any spoken input. Other studies found that this interference is present not only for homophone related pairs of visual objects, but also for rhyme-related and phonological onset related pairs (Görges et al., 2013; Meyer & Damian, 2007). What is perhaps more remarkable is that both visual and phonological information of depicted objects is activated in parallel when more than one object is presented in the visual display (Meyer, Ouellet, & Häcker, 2008). However, it is important to note that these competition effects from visual display seem to be contingent on having a familiarization task where participants are given the names of the objects beforehand (Görges et al., 2013). Presumably, the visual representation of an object can have many labels, and if participants access a label that is not a homophone for the target template, no phonological competition will be observed.

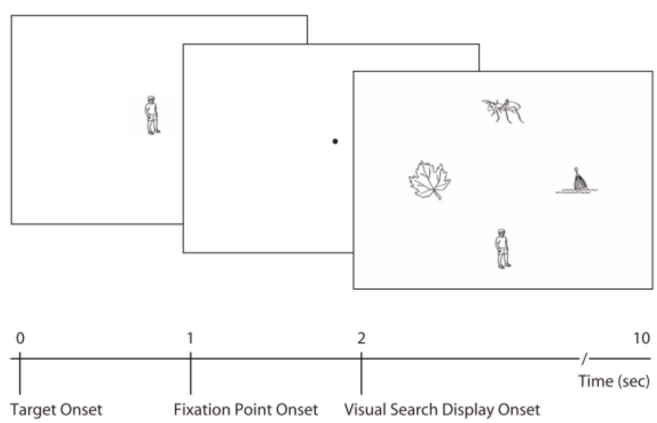


Figure 1: Example of a trial structure for target present displays from Meyer, Belke, Telling and Humphreys (2007).The target template boy is presented, followed by a blank display, followed by a search set that includes the target boy homophone buoy and two unrelated distractors ant and leaf.

The VWP literature has been slow in accepting the effects of preview or display (i.e., the influence of visual information), even though early papers have expressed concern regarding visual similarity of objects in the display and how they affect word recognition (see end of discussion in Allopenna, Magnuson & Tanenhaus, 1998). Anderson, Chiu, Huette, and Spivey, (2011), have eloquently expressed that more attention should be given to how the information from the visual subsystem and language subsystem is integrated:

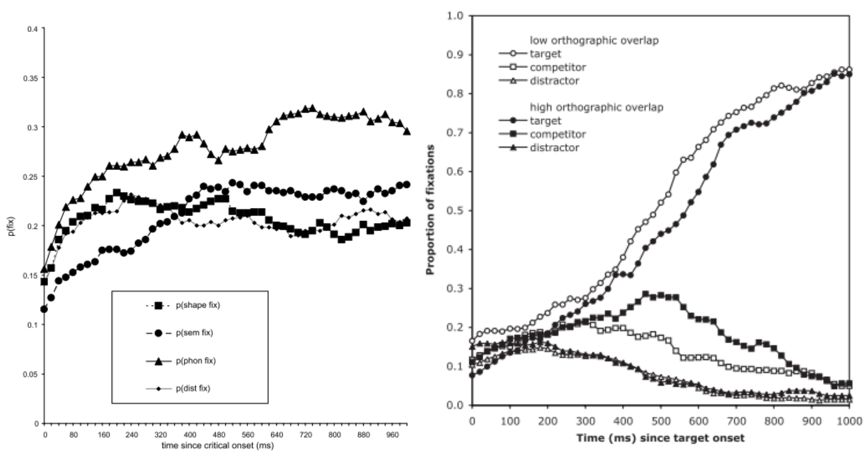
> "Many psycholinguists readily accept the idea that language comprehension can dramatically influence visual cognition, but when they are faced with claims that vision can influence language processing, they bristle. Inversely, many vision scientists accept with aplomb the idea that visual perception can profoundly influence language comprehension, but when you suggest to them that language could profoundly influence vision, they suddenly become skeptics." (p. 182).

The second reason suggests that the preview of the visual display allows for some types of information to be extracted but not others. Dahan and Tanenhaus (2005) found that the preview duration did not affect competition based on semantic shape similarity. In their experiments, subjects were given 300ms or 1000ms preview of a display with two objects that had similar shape features and two distractor objects (e.g., picture of rope, snake, umbrella, and sofa) before the target was presented (e.g., "click on snake"). They found competitor looks to the *rope* competitor in both 300ms and 1000ms preview conditions, and suggested that this argues against subjects' implicitly naming the objects in the display, and that pre-exposure to the display leads to perceptual processing rather than lexical processing. Nonetheless, Huettig and McQueen (2007) found that short preview time (200ms) affects the access to phonological information, but not to semantic information. In Experiment 2, they report that limiting the exposure to the display of objects before the auditory target is presented leads to the disappearance of phonological competitor effects, while competition due to semantic and visual similarity still remains (Huettig & McQueen, 2007). In other words, activation of the names of presented objects can be lessened if the display has brief exposure or is presented concurrently with the auditory target, reflecting that participants need time to engage in implicit naming or in identifying the object. Nonetheless, if the display contains printed words rather than pictures, brief exposure (e.g., 200ms rather than 1000ms) does not eliminate phonological competitor effects. This is likely because it is easier to access the phonological representations of printed words rather than of objects, mainly because for printed words, phonological knowledge can be accessed directly from the orthographic input.

Probably the best exemplification why the visual display matters when examining language mediated eye-movement is by comparing the results of VWP studies using objects versus those using printed words. Huettig and McQueen (2007) found that in object displays, eye-movements seem to be guided by a “tug of war” between phonological representations, semantic representations, and visual representations. With printed word displays, fixations are generally driven by a mapping of semantic (Huettig & McQueen, 2011; Sajin & Connine, 2014), phonological, or orthographic representations (McQueen & Viebahn, 2007; Salverda & Tanenhaus, 2010), with visual representations showing either no or very delayed mapping. In a later study (Huettig & McQueen, 2011) that used printed word displays, even when the spoken referents were embedded into imagery loaded sentences that would enhance the activation of visual information, no shape competitor effects were found during the time-course of the spoken utterance (the shape-competitor effect appeared only about 2.5s later after target onset). This evidence indicates that the time-course of the activation for shape information is based on the type of stimuli that are present in the display. When presented with printed word displays, listeners don’t immediately activate mental imagery from the spoken referent that, in turn, activates shape information. On the other hand, visual shape information becomes more accessible when objects are used. In this dissertation, I do not aim to compare how preview affects the activation of different types of representations, but I do examine how preview affects processing of the spoken target.

Another reason why examining the preview is important is because the preview allows the listener to pre-activate and store information about the objects and their locations in working memory. This information could, in turn, speed up processing because participants don’t have to engage in looking at areas in the display that they know do not contain the object/word matching the onset of the target they hear. In the interest of full disclosure, it important to point out that studying how preview affects the storage and binding of working memory representations has not been proposed in the original formulation of this dissertation, so experiments examining how preview affects binding in working memory need to be replicated using more direct manipulations. Nonetheless, the results of several experimental conditions will be used to argue that with sufficient enough preview, one observes a shift from memory-less visual processing to a more efficient memory-driven processing.

Perhaps another reason why it is important to examine what happens during the preview period in the VWP, when participants have to engage solely in visual processing, is to also get some closure about baseline differences that are often observed between the onset of fixation curves (see examples in Figure 2). Baseline differences indicate that even before the spoken target is presented, participants show preferential looks to certain words in the display. These differences can sometimes make it a challenge to compare fixation curves over time, since the intercepts of these curves have different starting positions. In all experiments in this dissertation, the context in which the preview was presented was neutral and all the words in the display were carefully matched on several lexical dimensions. Therefore, it is expected that no baseline differences should occur.



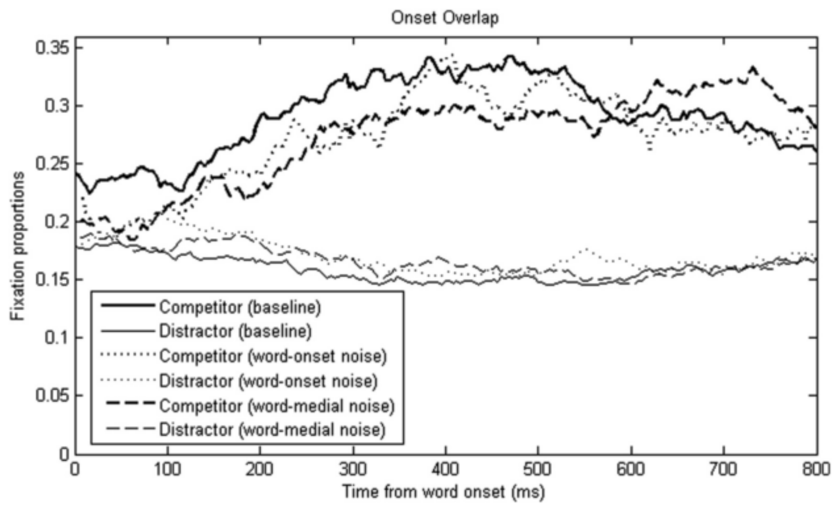


Figure 2: Examples of baseline differences. From top to right; Huettig & McQueen, Figure 5 (2007); Salverda & Tanenhaus, Figure 3 (2010); McQueen & Huettig, Figure 1 (2012);

The Linking Hypothesis

The formalization of the linking hypothesis started with the work by Allopenna, Magnuson, & Tanenhaus (1998), who evaluated the time-course of activation for lexical onset or rhyme competitors. In their experiments, participants were instructed to "pick up the beaker", while being presented with the objects *beaker* (target), *beetle* (onset competitor), *speaker* (rhyme competitor), and *carriage* (distractor). They found support for continuous mapping models (e.g., TRACE; McClelland & Elman, 1986), by observing that activation for both cohort and rhyme competitors, indicating that eye-movements are closely linked to the activation of the name of the object from the spoken input. The linking hypothesis in Allopenna, Magnuson, and Tanenhaus (1998) was initially phrased as a computational definition, where activations of lexical candidates in TRACE simulations were transformed to response probabilities using Luce choice rule. The authors point out that they "made the simplifying assumption that only the activations of those words in the visual display would be evaluated using the [Luce] choice rule." P. 424). In other words, the initial formulation of the linking hypothesis makes a somewhat limiting assumption, since the strength of a lexical candidate (i.e. higher fixation probability) is assessed based on the limited set of lexical candidates that are available in the visual display. This limitation was later fixed by showing that properties such as frequency can affect resting-activation levels, connection strengths, and the post activation bias in selecting the target (Dahan et al., 2001). Dahan and colleagues (2001) also provided behavioral data indicating frequency effects while not using any competitors in the display. Moreover, other studies have found that competition arises even when the target is not present, using this as support for activation not being dependent on circumscribed visual context (Huettig & Altmann, 2007; Huettig & Hartsuiker, 2008; Huettig & McQueen, 2007). It is important to note that lexical activation in the linking hypothesis is driven by the spoken input, and that no account is given as to how visual processing affects word recognition. In a later review, Tanenhaus et al. (2000), make the claim that, "the assumption…is that the activation of the name of a picture determines the probability that a subject will shift attention to that picture and thus make a saccadic eye movement to fixate it." (p. 567) and explicitly argue that the lexical representation (i.e., the name) is not accessed through implicit naming of the visual objects but from the unfolding of the spoken input (see pp. 878-879, Tanenhaus & Trueswell 2006). In other words, the linking hypothesis makes a very directional view about what drives eye fixations: the continuous activation of the lexical representation is accessed from the spoken input and is mapped onto visual referents. In a later version of the linking-hypothesis, which tried to account for visual as well as phonological competition effects, Dahan & Tanenhaus (2005) also argued against implicit naming and even against participants being able to extract abstract shape information from the processing of the visual display. When participants were presented with displays that had a picture of a rope and snake, but which looked visually dissimilar (the rope was coiled and the snake was uncoiled), and heard the target *snake*, looks to the competitor *rope* were driven by visual information activated from the lexical activation of spoken input. In other words, by hearing *snake*, participants are able to access its visual properties, and use that information to map fixations to the competitor *rope*. This set of results is inconsistent to the literature on feature semantics that found that shared feature representations can be extracted from objects, and guide semantic similarity judgements and priming effects (see

McRae, de Sa, & Seidenberg, 1997). In other words, by activating the visual referent snake, one can also activate its visual features that are not displayed (e.g., that it usually coiled), and observe competitor looks to the object rope (and vice-versa).

Later formulations of the linking hypothesis acknowledge the view that eye-movements in the VWP reflect interactions between linguistic and pre-activated visual representations, making it difficult to isolate the effects of single variables (Salverda et al., 2011) and that more work needs to be done in order to understand how language and vision interact. In particular, the most recent revision of the linking hypothesis acknowledges that the VWP is a very complex task and that task parameters and goals play an important role in how visual attention is guided. For instance, the goal based linking hypothesis puts more emphasis on task affordances(Chambers, Tanenhaus, & Magnuson, 2004) and visual-subroutines observed in hand-eye coordination tasks (Ballard et al., 1995; Hayhoe & Ballard, 2005).

In this dissertation, I will test the linking hypothesis by examining if participants are able to implicitly name the words during the preview period and notice the phonological/orthographic similarity between them (see Experiment 1). I also will test if lexical competition effects can be found in the absence of any spoken input, but rather appear as an artifact of how fixations are aggregated across trials (Experiment 2).

Mental representations and working memory

Several proposals have been made that suggest that working memory plays an important role in binding different representations (e.g., semantic, phonological, visual) in paradigms exploring language vision interactions (Anderson et al., 2011; Ferreira et al., 2013; Huettig, Olivers, et al., 2011a; Knoeferle & Crocker, 2006, 2007a, 2007b). Huettig et al. (2011) has proposed a model where long-term memory representations are bound together in working memory when processing the visual and auditory input. In this model, the objects/words in the display are perceptually encoded into a type of visuospatial sketchpad, which allows them to be bound to specific locations. The visual representations in working memory will trigger the retrieval of long-term memory representations (e.g., such as semantic properties of the objects and their names), which will also be bound to the specific object locations. The presentation of the spoken utterance will also trigger activation of phonological and semantic long-term memory codes which will be matched to the long-term memory representations active due to visual input (see for general framework in Figure 3). The model is still unclear on how different mental representations are matched to each other and what exactly is happening in the nexus of working memory in which multiple representations are activated. Nonetheless there is evidence from the blank screen paradigm, in which participants hear the spoken input once the visual display is presented and then removed (Altmann, 2004), that suggests that eye movements are guided by mental representations stored in working memory. In Altmann (2004), upon hearing the target, participants showed saccades towards the blank regions of the screen in which the objects were previously seen, which indicates that their location is stored in memory. Earlier, Spivey and Geng (2001) also found that eye movements were mediated by a mental model of a story that participants heard. In their study, participants faced a blank screen and were unaware that their eye-movements are recorded. They were told to simply listen to several stories. In some stories the scene descriptions had a directional bias (up, down, left, right), such as a person rappelling down from the top of a canyon and falling 10 feet (downward bias). Spivey and Geng (2001) found that listeners moved their eyes with the directional bias of the story, indicating that eye movements can be guided by mental representations.

Altmann (2004; Altmann & Kamide, 2009) suggests that when we examine a visual display, we don’t store all the details about the objects located in the display. Instead, we have an episodic memory trace that indexes the objects’ locations. Once the spoken target is presented, eye-movements are used to orient towards the location of that trace in order to extract more important details about the object. Note that this theory is very similar to theories of visual search espoused by research on scene perception, where it is assumed that visual representation of complex scenes is believed to be impoverished and fleeting (Becker & Pashler, 2002; Rensink, 2000, 2002)

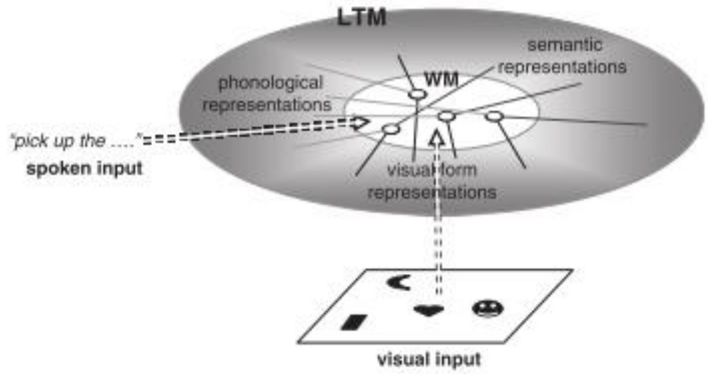


Figure 3: Huettig's working memory model. In this model visual representations are bound to specific locations, which with time will activate semantic and phonological information that will also be bound to that location. The more active representation in working memory will determine the guidance of visual attention. Form Huettig et al, (2011a).

Evidence that participants are trying to interpret the speech input in the context of the visual display by creating and manipulating mental representations has been provided by studies that examined anticipatory fixations for sentence processing (Altmann & Kamide, 1999, 2007, 2009). Altmann and Kamide (2007) found that the verb tense in an utterance can help participants interpret the referent in a visual display before subjects get to hear the name of the referent (see typical trial in Figure 4). In a later study, Altmann and Kamide (2009) found that anticipatory fixations can be guided by dynamically modifying the location of objects in the mental representation. In Experiment 1 (Altmann & Kamide, 2009), subjects were presented with displays such as depicted in Figure 5. Participants got to hear one of these two sentential conditions composed of two sentences:

> "(1) The woman will put the glass onto the table. Then, she will pick up the bottle, and pour the wine carefully into the glass.

(2) The woman is too lazy to put the glass onto the table. Instead, she will pick up the bottle, and pour the wine carefully into the glass." (p. 59)

When participants processed the word *glass* in the second sentence in the first sentential condition (1), they showed more fixations to the table relative to participants who were exposed to the second sentential condition (2). This is because in the first condition subjects had to mentally relocate the glass to the table while in the second condition the glass remained on the floor. This finding indicates that participants don't simply map language to objects located in the visual scene. Instead, eye-movements seem to be modulated by how language interacts with mental representations of the visual scenes.

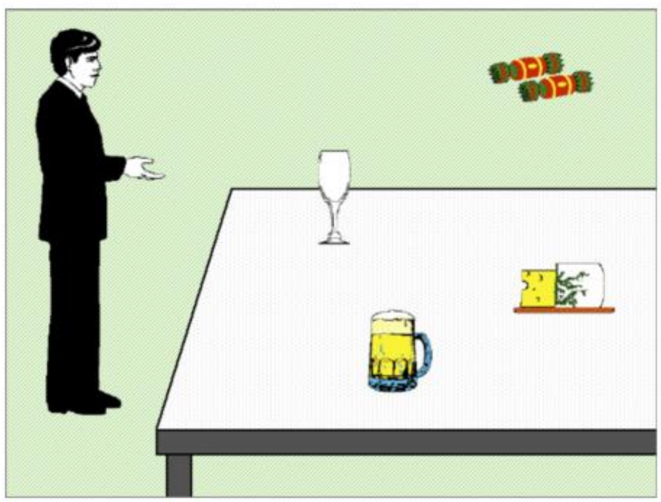


Figure 4: Visual Display from Altmann & Kamide (2007). Participants hear either "The man will drink the beer" or "The man has drunk the beer". The display shows an empty glass of wine, a full mug of beer, cheese, and Christmas crackers. Participants fixate on the beer when processing drink and show more fixations towards the empty glass when hearing drunk.

Figure 5: Example scene from Altmann & Kamide (2009)

Earlier, Altmann and Kamide (1999) found that eye-movement behavior that subjects exhibit depends on the task they are given. In Experiment 1, Altmann and Kamide (1999) presented to subjects semi-realistic scenes having a boy, a cake, and other distractor objects. During the presentation of the scene, subjects also heard sentences such as “the boy will eat the cake” or “the boy will move the cake” and they were asked to verify if sentences describe something that could occur in the scene. They found that subjects show anticipatory saccades towards cake upon hearing the onset of the verb *eat* rather than the verb *move.* This evidence was used to argue that participants use language information to restrict the domain of visual referents to which they should attend. Interestingly, in Experiment 2, Altmann and Kamide (1999) found that if participants are told to not pay any attention to the spoken sentences, then anticipatory saccades are not time-locked to the onset of the verb but rather occur slightly later. This pattern of results indicates that goal directed behavior influences the linguistically-mediated saccadic eye movements. When participants are instructed to verify if spoken utterances are related to visual referents, visual attention is closely aligned to the unfolding of the spoken input. On the other hand, if participants are asked to look-and-listen, there is a delay in the deployment of visual attention. One possibility for why there is a delay in look-and-listen version of the VWP, is that participants are not required to develop and store in working memory expectations about what possible utterances they might hear. Recently, Ferreira, Foucart, and Engelhardt (2013) found that participants are able to generate expectations of possible instructions for garden path sentences based on the preview of the display, indicating that visual attention is guided by an interaction of expectations built up in the working memory, visual display, and language. If the visual complexity of the display is increased, participants don't generally rely on expectations because the number of possible utterances that could be heard is too large for them to predict and hold in working memory.

Visual attention guided by working memory has also been observed in studies examining visual search (Olivers, 2009; Soto et al., 2005; Soto & Humphreys, 2007). Downing (2000) found that the contents of working memory guide attention even when participants are not explicitly asked to perform a visual search task and Soto and Humphreys (2007) found that visual attention can be guided by items stored in working memory only when subjects are told to remember those objects or name them, indicating that even if objects are attended to, they still have to be actively stored in mental memory in order to interfere with visual search. In a very similar setup as that used in the VWP, Spivey and colleagues (2001; see also Chiu & Spivey, 2014) found that language can constrain the visual search, so that the continuous processing of the spoken utterance limits the effects due to display set size (see Figure 6 for procedural setup). In Experiments 1&2, in one condition participants had to search for the target in a search display after the verbal instructions were presented, while in the second condition, the search display was presented concurrently with the auditory onset of the first feature of the target. Results of the first condition fit the results from the previous literature on visual set size: as the set size increased from 5 to 20 objects, the visual search for the target slowed linearly. In the second condition, the results showed shallower visual search slopes as the set size increased. These results indicate that incremental processing of spoken input constrains the visual search. Upon the presentation of the onset of the first feature (e.g., red), subjects orient their attention towards bars that are in red, thus substantially decreasing the set size before the second feature of the target (e.g., vertical) is presented. The alternative explanation is that the feature red constrains not only the set size (i.e., subjects reduce the set size to red objects only), but also induces a pop-out effect because it allows them to identify the odd vertical bar among horizontal red bars. Therefore, it would not even be necessary to hear the second feature because the information about the first feature plus information to look for the odd shape would be enough to drive attention towards the target. Unfortunately, the authors don't provide time-course data that would allow one to decide which explanation is most appropriate. Regardless of which explanation fits the data better, both accounts indicate that goal-directed language constrains processing of relevant or irrelevant aspects of the visual display.

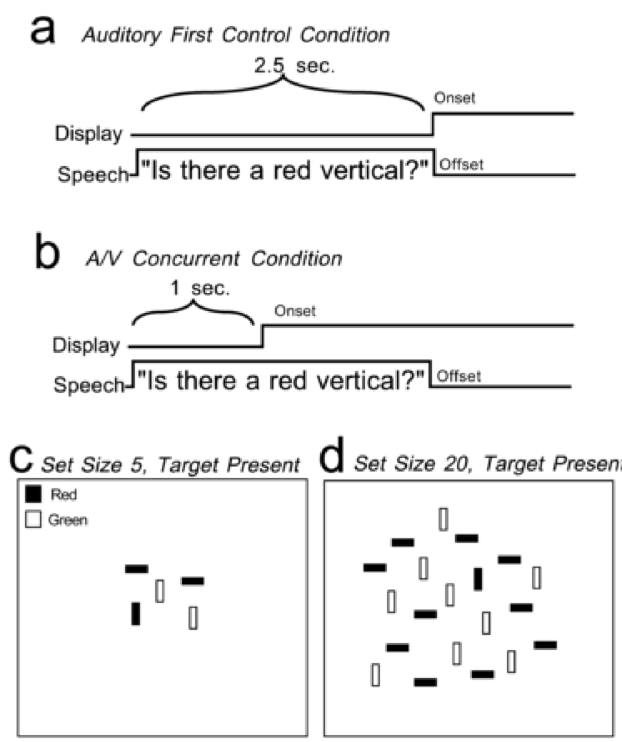


Figure 6: Example trial from Spivey, Tyler, Eberhard, and Tanenhaus (2001).In condition (a), participants are presented with the display with varying set size (such as c and d after the auditory instructions are presented. In condition (b), the display appears concurrently with the onset of the first feature of the object that needs to be searched. The vertical black bar is in red, while the remaining vertical bars are in green and horizontal bars are in red.

Concluding Remarks

As outlined in a review by Huettig (2011), the VWP does not reflect exclusively language processing or visual processing, but rather speaks to the language-vision interaction as listeners try to process both visual and spoken input. Similar views have been expressed by Salverda, Brown, and Tanenhaus (2011), who refined the original formulation of the linking hypothesis in order to account for task related effects, such as participant's goals and affordances in trying to make sense of the spoken information and visual input. The experiments proposed below aim to explore several factors that are likely to be important when processing spoken words in the VWP. In particular, Experiment 1 will measure what guides participant's eye-movements during the preview period, and if competitor effects can be observed before the presentation of the spoken target. Experiments 2 will investigate if eye-movements can be guided by an internal linguistic representation rather than due to an automatic mapping of the spoken utterance to printed words. Experiment 3 will examine if the addition of cognitive load will disrupt the storage of the linguistic representation in working memory and thus eliminate the phonological competitor effect. Experiment 4 will test if in the presence of an additional task that requires listeners to focus on perceptual attributes of the words in the display, the effect of phonological competition can be enhanced or diminished. Experiment 5 will involve a simple, yet untested, manipulation that will investigate how the complexity of the visual display affects phonological competition effects. In all experiments, the printed words version of the VWP will be used (McQueen & Viebahn, 2007), because printed words are less ambiguous in their labeling than pictures (i.e., a picture can have many labels) and because of the benefit of being able to use a larger set of stimuli (e.g., words referring to abstract concepts), which makes it easier to design controlled sets of stimuli for each display.

In addition to the above, I will draw comparisons between the divergence points across Experiments 1,2, and 5. As mentioned previously, serial, memory-less search is likely to be delayed relative to search that is driven by representations stored in working memory. With longer previews, one should expect stronger working memory representations, since participants are given more time to pre-activate the linguistic information from visual word-forms. In other words, longer previews should lead to faster search times. On a similar note, if the display is more complex, one should expect weaker working memory representations, serial visual search, and delayed divergence between competitor and distractor fixations.

Data Analysis

Fixation proportions were used as a measure of processing and were calculated for 25ms time bins based on looks to the whole display rather than to looks only to the four/six interest areas (thus the sum of proportions to the interest areas will not always sum up to 1).

Growth Curve Analysis

In all experiments reported in this thesis, fixation proportions were analyzed using three different types of models. First, fixations were examined using Growth Curve Analysis (GCA; Mirman, Dixon, & Magnuson, 2008; Mirman, 2014), a modeling technique that is fairly prevalent in the eye-tracking literature. GCA allows one to examine the change in fixation proportions between conditions by modeling the interactions between main effects and orthogonal (i.e., uncorrelated) polynomials. For instance, if targets show higher fixation proportion over time when compared to competitors, the effect will be visible in the in the linear slope coefficient between the linear polynomial and the condition examined (Target vs. Competitor). A variety of models were fitted for each experimental comparison. The starting point for model building was the base model, which had no polynomials and included intercepts-only random effects in its structure. Following the base model, subsequent models increased in model complexity by adding the effects of polynomials in the fixed effect and the random effect structure, up until the point where models were indicating convergence issues. Polynomials were added in the following order: linear (T1), quadratic (T2), cubic (T3) and quartic (T4). For each experiment, a table will be provided detailing the model building process and the amount of explained variation using a method by Xu (2003). In addition, a graphical representation of each model's fit to the data is provided. Unless otherwise indicated, the error bars in subsequent graphs represent 95% confidence intervals. Although multiple models will be tested, the discussion and presentation of results will focus only on the models that provide the best fit to the data, assuming that the models successfully converge.

Cluster Based Permutation Analyses (i.e., divergence analysis)

Although GCA provides with an understanding as to how the shape of the two fixation curves are different from each other, it does not provide a way to measure where two fixation curves diverge. In order to measure the divergence between two fixation curves, cluster based permutation analyses were performed (Maris & Oostenveld, 2007). The approach to performing this analysis was as follows:

1. t-tests were performed for each 25ms time bin on the original dataset.
2. Adjacent time-bins with t-values larger than the threshold for a within-subject t-test (usually a value between 2.05-2.09 based on having 21-25 subjects in an experiment) were grouped together to form the t-sum statistic. These adjacent time-bins formed time-clusters where the difference between one fixation curve relative to another was significant.
3. The data was shuffled randomly. For the randomly shuffled data, the largest t-sum statistic was calculated.
4. Step 3 was repeated 10000 times
5. The distribution of the-sum statistic from step 3 is calculated, and the p-value for the t-sum for the clusters found in step 2 is calculated relative to this distribution.

Generally, these analyses will be accompanied by two plots. In one plot, the t-statistic for the time-clusters will be provided during the course of a trial. This plot will showcase where during a trial is the difference significant in fixation proportions between two conditions. The second plot will show the distribution of the t-sums from the 10000 permutation tests. This plot essentially showcases how likely the difference between fixation curves in the original dataset occurred by chance. The divergence analysis was performed using eyetrackingR (Dink & Ferguson, 2015).

Random Forests (see Supplementary Section)

GCA provides a good modeling approach to examining how the shape of fixation curves changes during the course of a trial. Nonetheless, it is not very robust to issues related to overfitting. Moreover, the implementation GCA with many polynomials in linear mixed effects models (Baayen, Davidson, & Bates, 2008) sometimes leads to convergence issues. Moreover, the S.E. calculation of the effects in GCA seems to be susceptible to how the time-bins are partitioned. Shorter duration time-bins (10ms vs 50ms) lead to a larger number of data-points and a smaller S.E. in the calculation of the main effects, creating the impression of highly statistically significant main effects despite having very small coefficients. Because of these issues, the third modeling approach will focus on random forests (Biau, 2010; Sandri & Zuccolotto, 2006), which involve an ensemble of regression trees that are built on random subsets of the data. All random forest models were run using 200 trees and were validated using 5-fold cross-validation, with 80/20 split between train and test sets. The results will report the model with the best average cross-validation performance and the graphical representation of the model fit to the data. Unfortunately, random forests do not have probability values associated with each predictor variable, but rather provide a Gini-like purity index, which indicates the importance of a variable in reducing node impurity (i.e., MSE) across different decision tree splits (see Hastie, Tibshirani, & Friedman, 2009; also see for how it was implemented http://topepo.github.io/caret/varimp.html).

The purpose of reporting random forests models is to identify any potential issues related to overfitting in GCA. For each comparison, the output for a RF model will be provided. One random forest will include all the interactions between the four polynomials and condition and the subject ID variable. This model will be designated as random forest complex. The second model will be a more parsimonious model, including the interaction only between the linear term and condition, together with the subject ID. This model will be designated as random forest simple. All the output from random forests analyses has been included in a supplementary section. The reader is welcome to consult this section if there are concerns regarding overfitting in G.

All analyses were performed in R (Version 3.2.2; R Core Team, 2015) using the following packages: lme4 (Version 1.1-8; Bates, Maechler, Bolker, & Walker, 2015), lmerTest (Version 2.0-29; Kuznetsova, Brockhoff, & Christensen, 2015), dplyr (Version 0.4.2; Wickham & Francois, 2015), tidyr (Version 0.2; Wickham, 2014), and caret (Version 6.0-47; Khun, 2015). Plots were created using ggplot2 (Wickham, 2009) and gridExtra (Version 2.0; Auguie, 2015).

Experiment 1

Experiment 1 aims to replicate the phonological competition effect in the VWP and examine if visual attention during the preview period (i.e., before the auditory target is presented) can be guided by orthographic similarity of words in the display. Currently, there is no empirical evidence that reports if competitor effects are found when participants are given a preview of a display with words that overlap in phonological/orthographic onset without any auditory input. This experiment will test several assumptions proposed by the linking hypothesis. According to the linking hypothesis (Salverda et al., 2011; Tanenhaus, Magnuson, Dahan, & Chambers, 2000), the main assumption is that fixations are guided by visual attentional shift while subjects interpret the ongoing linguistic input. In the VWP, the goal of the participant is to understand the spoken input, and eye-movements are believed to reveal how listeners activate linguistic representations present in the speech input as they process an utterance. For instance, upon hearing an ambiguous onset /sk/, participants try to resolve the ambiguity by making provisional commitments towards objects or words that start with phoneme /sk/ (e.g., scan or scam). As speech input unfolds, the ambiguity is being resolved and subjects fixate on words/objects in the display that fit with linguistic representations that remain active. The linking hypothesis assumes that fixations generally arise because of the representational overlap between the language input and the visual input. Interestingly, it also assumes that visual scene representation is independent of the spoken input. In other words, it would not predict any competitor effects arising from the visual display alone, but it would predict competitor effects arising from the need to resolve ambiguity (whether it is lexical, semantic, or syntactic) present in the speech input. In this experiment, participants will be provided with a preview of 1.9s before the presentation of the spoken target input. According to the linking hypothesis, in the absence of the target provided through the spoken instructions, eye-movements during the preview period should be guided to all four words in the display with equal probability. On the other hand, during the presentation of the spoken target, visual attention should be allocated towards words in the display that have a closer match to the target. When the visual target is presented together with a phonological competitor, eye-movements will be guided based on the representational overlap between phonological input and visual input. Since the phonological input initially is ambiguous, subjects will fixate on both the target and competitor. Nonetheless, as the spoken production unfolds, more fixations will be oriented towards the target word rather than the competitor.

Methods

Materials

Three types of trials were used. In experimental trials, participants were presented with a display with two phonological/orthographic onset competitors and two distractors that were phonologically and orthographically unrelated to the two competitors and to each other. The onset overlap for the two competitors was defined as sharing at least two letters and two phonemes. In experimental trials, the auditory target was one of the two competitors. For these trials, stimuli were normed so that all four words in a display did not differ among each other in terms of the number of syllables, letters, phonemes, age of acquisition (AoA; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) , concreteness (Brysbaert, Warriner, & Kuperman, 2013), emotional valence (Warriner, Kuperman, & Brysbaert, 2013), phonological and orthographic Levenstein Distance (Yarkoni, Balota, & Yap, 2008), subtitle frequency (Brysbaert & New, 2009), and number of word senses (Durda & Buchanan, 2006). All pair-wise comparisons have p>.2. There were a total of 36 experimental trials. The list of items can be found in Appendix A and the norming statistics (i.e., means and standard deviations) are in Table 1.

Table 1: Information on stimuli norming

Mean (SD) statistics for words (Target, Competitor, and Distractors) in experimental trials. Distractor 3 and Distractor 4 columns apply for 6-word display in Experiment 5.

	4-word displays				**Additional Distractors for 6-word displays in Exp.5**	
	Target	Competitor	Dist. 1	Dis. 2	Dis. 3	Dis.4
Syllables	1.64 (.49)	1.67 (.48)	1.67(.48)	1.64(.49)	1.67(.48)	1.67(.48)
Letters	5.44(.65)	5.44(.65)	5.44(.65)	5.44(.65)	5.44(.65)	5.44(.65)
Phonemes	4.61(.69)	4.58(.65)	4.64(.8)	4.64(.72)	4.5(.65)	4.64(.64)
Phonological Levenstein D.	1.65(.45)	1.67(.37)	1.63(.4)	1.72(.45)	1.71(.37)	1.71(.45)
Orthographic Levenstein D.	1.87(.35)	1.82(.31)	1.87(.35)	1.88(.32)	1.9(.33)	1.9(.35)
Subtitle Frequency	2.61(0.61)	2.63(0.73)	2.67(.50)	2.69(.69)	2.61(.46)	2.62(.69)
Concreteness	4.36(.79)	4.27(.77)	4.32(.81)	4.2(.72)	4.23(.84)	4.28(.86)
Senses	4.81(1.72)	4.61(2.07)	5.11(2.01)	4.83(2.78)	5.06(2.03)	5.36(2.89)
Age of Acquisition	6.68(1.92)	6.89(2.47)	7.13(2.04)	6.69(1.81)	7.18(2.41)	7.03(2.32)
Valence	5.58(.98)	5.59(.87)	5.85(1.08)	5.66(1.14)	5.57(1.14)	5.61(0.91)

The second type of trials, called filler overlap trials, are similar in structure to the experimental trials, except that no norming has been done for these items, other than making sure that the words in the display have the same number of letters among each other. Participants are again presented with two phonological/orthographic onset competitors and two distractors that are phonologically and orthographically unrelated to the two competitors. In filler overlap trials, the target will be a word in the display that does not have phonological/orthographic overlap with any another word in the display [e.g., "click on bald" with a display having the words *bald* (Target), *scam* (Competitor), *scan* (Competitor), *cube* (Distractor)]. These trials have been introduced in order to reduce the possibility of subjects developing expectations that whenever they see two words with similar onset in the display, the target will always be one of the two words (see also Salverda & Tanenhaus, 2010 who used similar type of fillers). There are a total of 36 such trials.

In the third type of trials, filler no overlap, all four words in a display share no orthographic or phonological onset, but they have been normed on the number of letters. These trials are used as fillers. There are a total of 36 such trials.

In all trials, subjects will hear the prompt "click on" followed by a target word. For instance, participants might hear "click on buckle", while being presented with the display with the words *buckle* (Target), *bucket* (Competitor), *diesel* (Distractor), and *kicker* (Distractor). The production of the prompt "click on" was recorded separately and then spliced to each target word. In other words, each target word was presented with the same click on production. This was done mainly to eliminate any potential effects of co-articulation. Between the offset of "click on" and the onset of the target word, a silent 200ms pause was inserted. The duration of the "click on" production was 682ms.

The prompt "click on" as well as all the targets was recorded by a native speaker of American English and digitized at 48kHz (16 bit resolution). The targets were recorded in the absence of any sentential context. Root mean square normalization was done so as to make sure that the average intensity is similar across productions. The list of targets and displays can be found in Appendix A.

Participants

A total of 21 Binghamton undergraduate students took part in Experiment 1 in order to fulfil a psychology course requirement. All participants were native speakers of English and had normal or corrected-to-normal vision and no hearing impairments. Three participants were bilingual, speaking mainly English at home, with the remaining being monolingual.

Procedure (see Figure 43a at the end of the document for trial structure)

Gaze position and duration for the right eye was recorded with a head-mounted eye-tracker at 500Hz sampling rate using the EyeLink 2 system. Participants were seated approximatively 150cm from a 16 x 12 inch monitor with a resolution of 1024x768. The calibration of eye-movements was done using 9-point calibration procedure at the beginning of experimental trials and practice trials. At the start of each trial, participants were presented with a fixation cross in the center of the screen for one second. Participants were asked to orient their eyes towards the cross when it appears. After the presentation of the fixation cross, a display with four printed words was presented. The four words were in Times New Roman, font size 24, black color, and were positioned in the quadrants surrounding the fixation cross. The location of the four words was randomly assigned to each quadrant. Eye movements were recorded for interest areas (225x225 pixels; about 3.5 x 3.5 inches) centered around the four words. Participants were instructed to look at the words immediately when they appear on the screen. One second after the presentation of the word display, the audio input is presented through the headphones instructing subjects which target to click (e.g., “Click on lamp”). Participants were instructed to use the mouse to select the target word. The trial ended when subjects clicked on a word. In addition to recording to eye-movements to the four interest areas, click times were recorded. Throughout the experiment, drift correction was done at the start of each trial. Participants were given six practice trials, after which they proceeded to the remaining 108 experimental trials.

Results

Fixations for 0.2% of the trials for which participants selected a word other than the target were removed. Average mouse click time was 1238ms, as measured from the onset of the spoken target. The top panels of Figure 7 plot the fixations proportions from the onset of the trial until 6000ms. The bottom panels plot a zoomed-in portion between the two vertical lines for the top panels, which corresponds to the time from the onset of the spoken target until 4200ms into the trial. Figure 7 depicts fixations for all three types of trials: experimental, filler no overlap, and filler overlap. The bars represent 95% confidence intervals. Bilingual participants performed similarly to monolingual participants, so the data used in figures and analyses includes all subjects. Note that in filler trials there is no competitor word per se, but rather the competitor is a distractor.

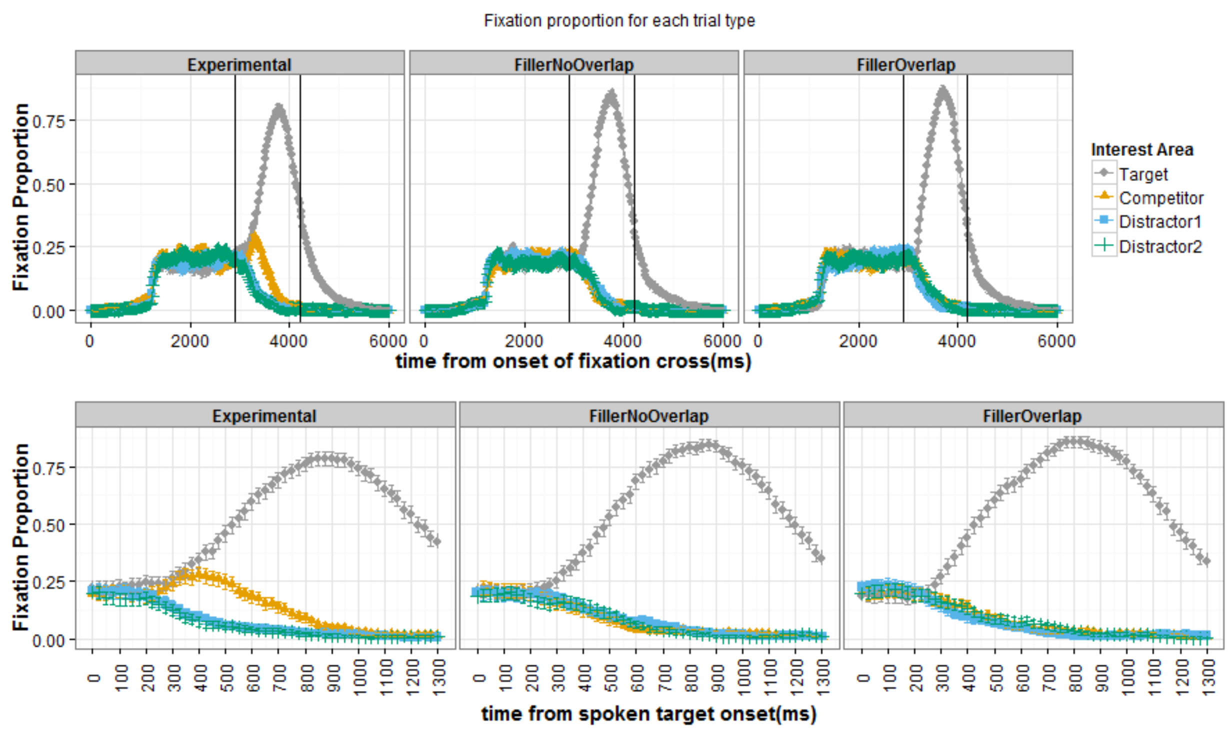


Figure 7: Fixation proportions for Experiment 1 from the start of the trial (top three panels) until 6000ms. The bottom panels represent the fixation proportions from the onset of the target word until 1300ms after the onset of the target.

The top panels in Figure 7 indicate that before the onset of the spoken target is presented, participants fixate on the four words with equal preference. Moreover, the figure also indicates that participants do not look preferentially at words that are similar in their orthography/phonology before the onset of the spoken target. This can be observed even more clearly by zooming in on fixations 200ms after the display onset till 3000ms into the trial (Figure 8).

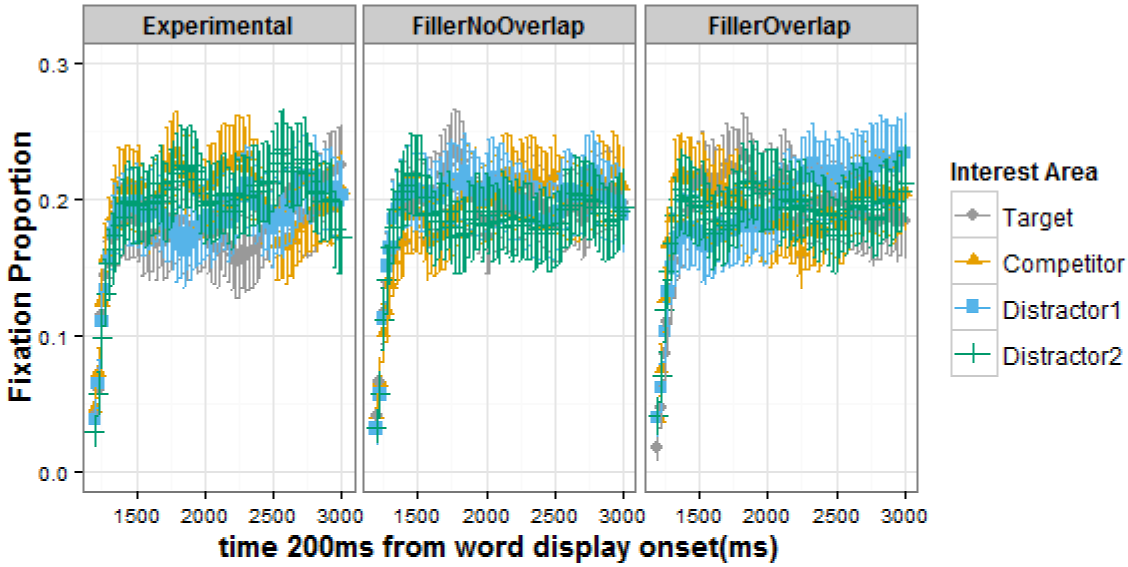


Figure 8: Fixation proportions for Experiment 1 200ms from the onset of the word display until 3000ms into the trial.

Because there is such an overwhelming overlap in CI between conditions during the preview period, the data for preview will not be analyzed. Instead, the data will be analyzed for fixations presented in Experimental trials for the left bottom panel in Figure 8. Fixations will be first compared between target and competitor and then between competitor and average of the two distractors.

Target vs. Competitor

Fixations between the target and competitors are compared using GCA and permutation cluster analysis. Table 2 depicts the structure of all the models that have been tested, the percentage variance explained by each model, and the Aikake Information Criterion (AIC). Figure 9 presents a graphical representation of the models' fit to the data.

Table 2: Models tested for Target vs. Competitor comparison in Experiment 1

Model Type and Name	**Model Structure**	**R^2**	**AIC**
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.606	-997.7
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.912	-3091.9

Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.955	-3925.5
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.982	-5066.9
Model 5 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2|Subject:Condition)	.976	-4802.8
Model 6 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2|Subject)+ (1+T1+T2+T3|Subject:Condition)	.982	-5075.0

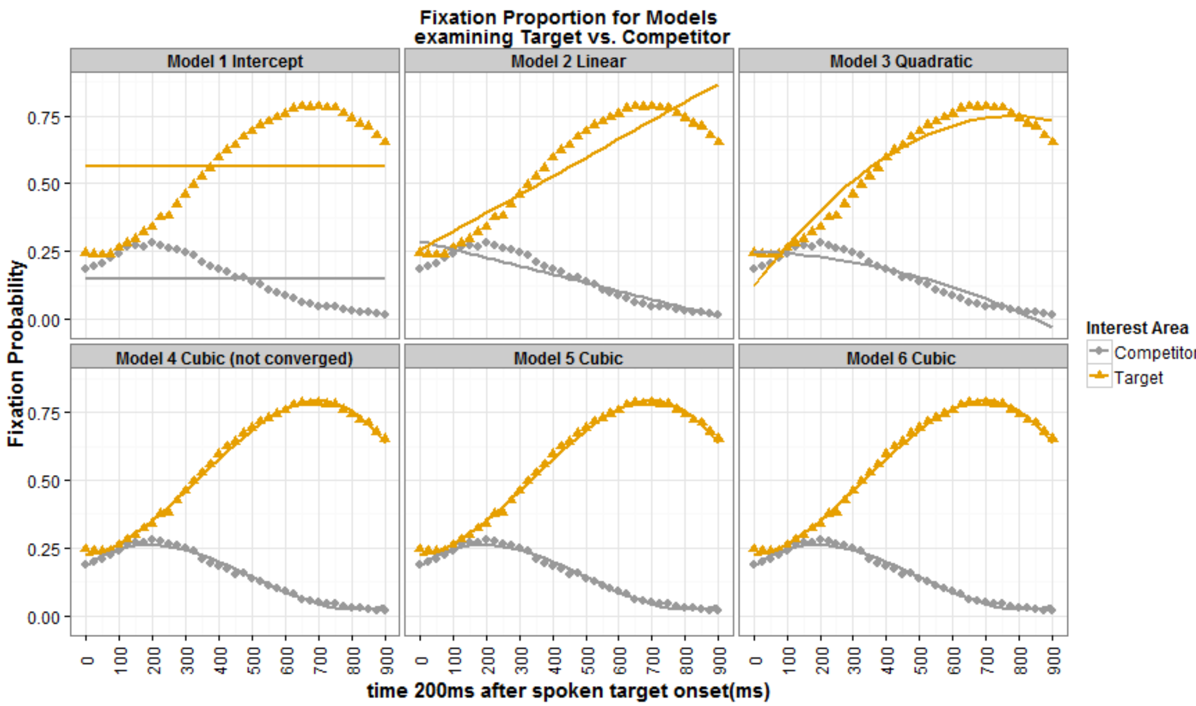


Figure 9: Predictions of the fitted models to the data in Experiment 1 (Target vs. Competitor). The lines represent the predicted values for the models, while the shapes are the mean data for each time point.

For GCA models, Model 6 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 6 can be found in Table 3.

Table 3: : Model 6 output for Target vs. Competitor in Experiment 1

Predictor	β	S.E.	t	p
(Intercept)	.149	.011	13.49	<.0001
Condition (Target vs. Competitor)	.413	.015	26.47	<.0001
T1	-.503	.067	-7.426	<.0001
T2	-.118	.040	-2.490	.0055
T3	.177	.031	5.683	<.0001
T1:Condition	1.607	.095	16.87	<.0001
T2:Condition	-.277	.053	-5.192	<.0001
T3:Condition	-.455	.044	-10.33	<.0001

An examination of the unstandardized coefficients in Table 3, indicates that the effect is mainly driven by an overall difference between conditions, the interaction between condition and linear polynomial, followed by an interaction between cubic polynomial and condition, followed lastly by the interaction between the quadratic polynomial and condition.

The bootstrapped cluster-based permutation analysis indicates that the competitor and the target diverge at the 225ms mark. The cluster of summed t-tests (Sum=593; see cluster 2 in Figure 10) is far outside of a 10000 sample distribution of permutation tests (Mean=-.038; 95% CI [-104,102], p=0).

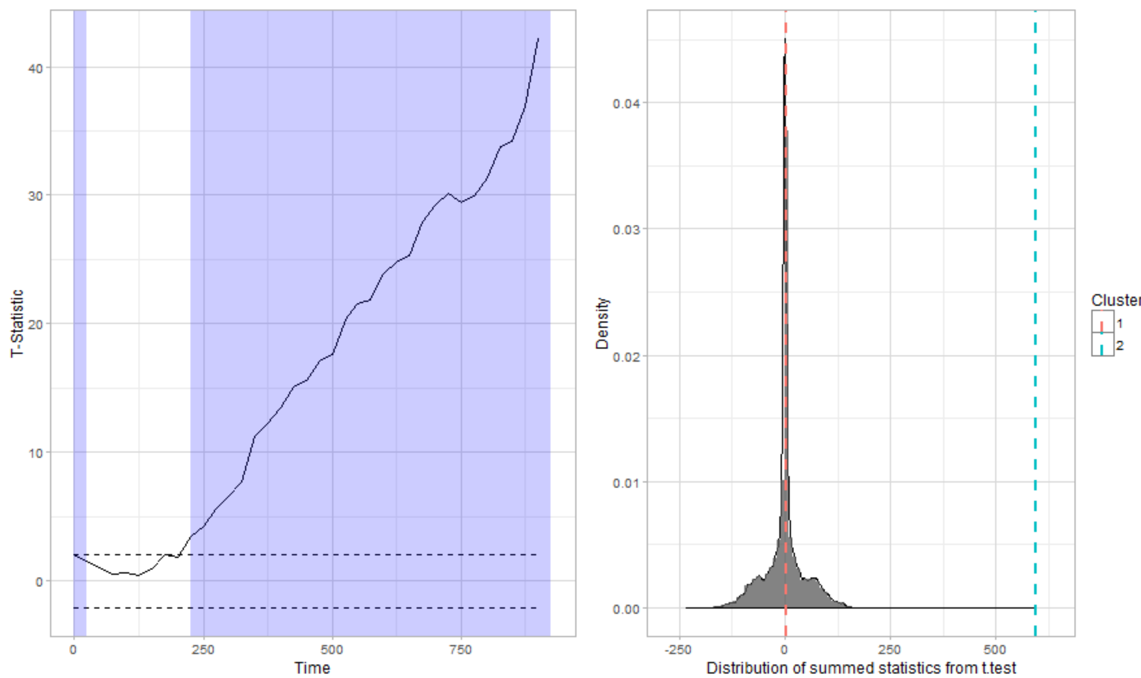


Figure 10: Permutation-cluster outcome for Experiment 1 (Target vs. Competitor). The left panel represent the t-statistic for every 25ms time-bin (time is measured 200ms from spoken target onset). The areas highlighted in blue represent the time ranges where there was a significant difference between the target and the competitor curve. Two time-ranges are significant: 0-25ms and 225-925ms, which correspond to cluster 1 and 2 respectively. The two horizontal, dotted lines in the left plot represent the t-test threshold based on the number of subjects. The right panel indicates the distribution of t-sums from 10000 simulations. This distribution indicates that only the second cluster (i.e., second time-range represented by the blue dotted line) is reliable. The first time-cluster (for period 0-25) is likely to have happened by chance.

Competitor vs. Distractors

Fixations between competitor and the average of the two distractors are compared using the same approach as above. First, several models were fitted to the data (see Table 4 and Figure 11) and the output for the models with the best fit (which also converged is presented).

Table 4: Models tested in Experiment 1, comparing the fixations between Competitor and the average of the two distractors.

Model Type and Name	Model Structure	R^2	Aikake Information Criterion
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.283	-3199.6
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.780	-4856.7
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.859	-5374.6
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.929	-6229.0
Model 5 (quartic)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+ (1+T1+T2+T3+T4|Subject)+ (1+T1+T2+T3+T4|Subject:Condition)	.946	-6495.5
Model 6 (quartic)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+ (1+T1+T2+T3+T4|Subject)+ (1+T1+T2+T3|Subject:Condition)	.938	-6337.4
Model 7 (quartic)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3+T4|Subject:Condition)	.946	-6505.4

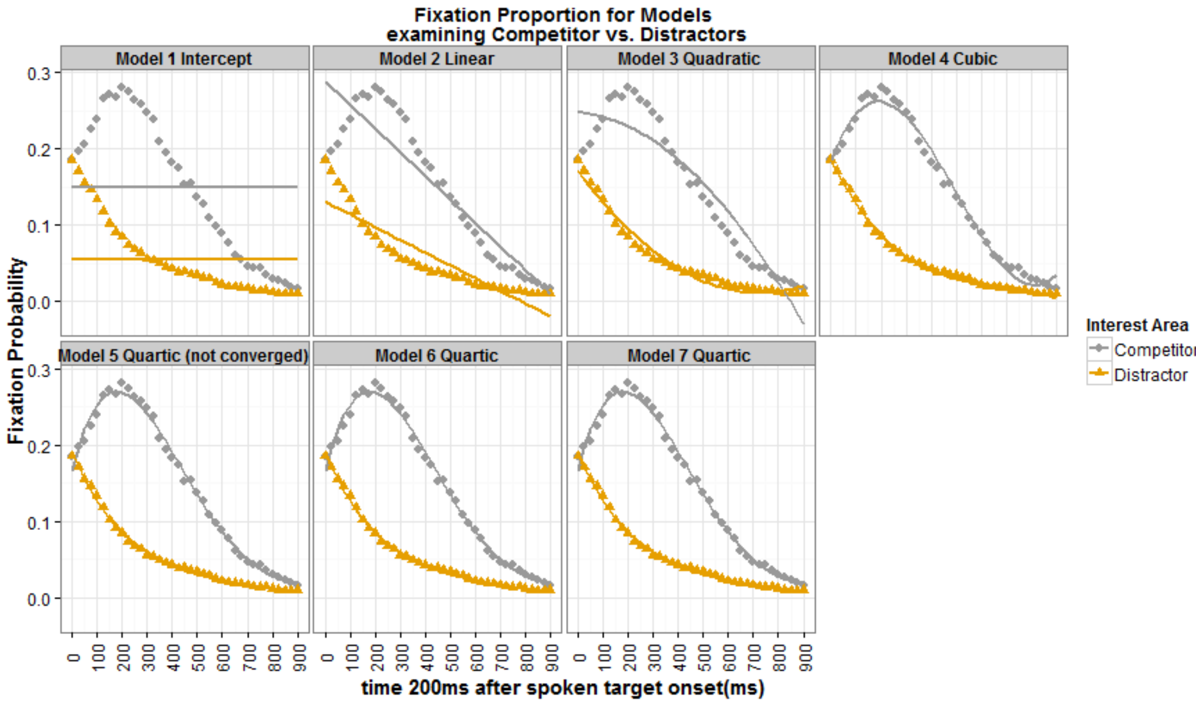


Figure 11: Predictions of the fitted models to the data in Experiment 1 (Competitor vs. Distractors). The lines represent the predicted values for models, while the shapes are the mean data for each time point.

For GCA models, Model 7 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 7 can be found in Table 5.

Table 5: Model 7 output for Competitor vs. Distractors in Experiment 1.

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.149	.006	25.94	<.0001
Condition (Target vs. Competitor)	-.093	.006	-14.42	<.0001
T1	-.501	.029	-17.05	<.0001
T2	-.118	.027	-4.31	.0001
T3	.177	.021	8.46	<.0001
T4	-.043	.016	-2.67	.0109
T1:Condition	.231	.041	5.57	<.0001
T2:Condition	.235	.038	6.14	<.0001
T3:Condition	-.219	.029	-7.51	<.0001
T4:Condition	.054	.022	2.38	.0221

The GCA indicates significant interactions between all four polynomials and condition, though the unstandardized coefficient for T4:Condition is much smaller than the coefficients for the other three interaction terms. The GCA also indicates that there is a significant main effect of Condition, with competitor interest area having more fixations than distractor interest areas.

The permutation cluster analysis indicates that the competitor diverges from the distractor at the 100ms mark (see Figure 12). This divergence is highly unlikely to happen by chance, as demonstrated by the large t-sum statistic (172) relative to the distribution of 10000 permutation tests (M=.32, 95% CI [-51, 52], p=0).

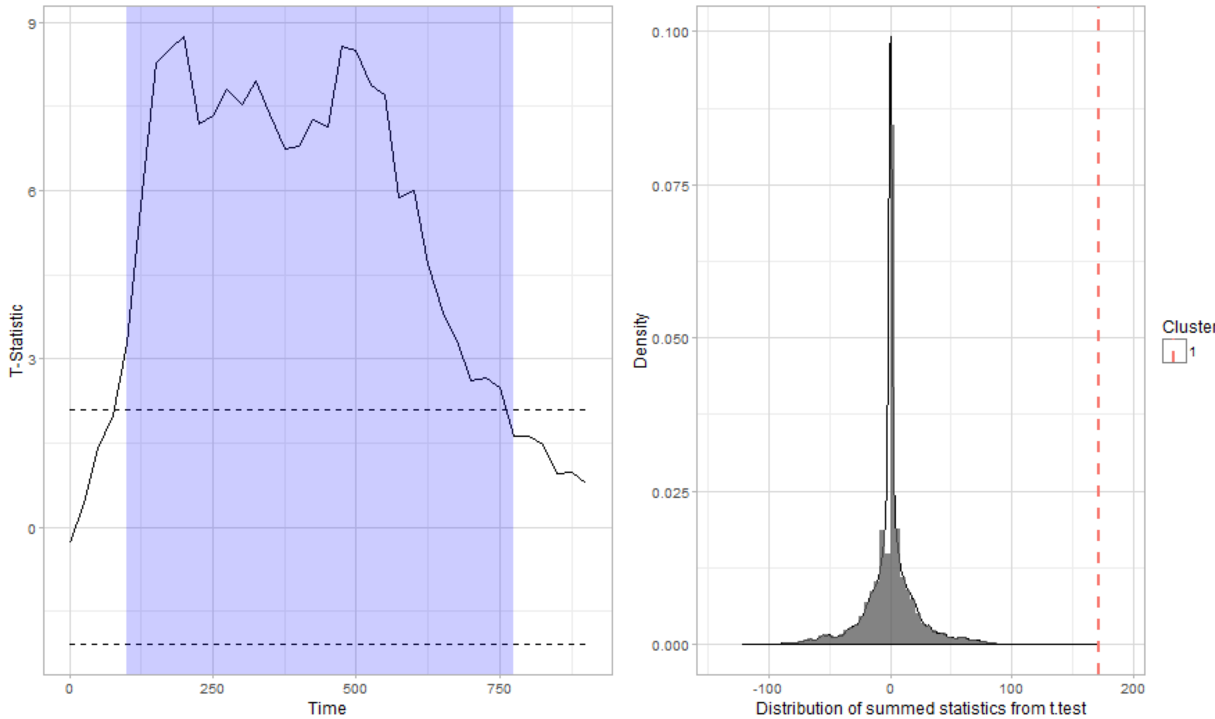


Figure 12: Permutation cluster outcome for Experiment 1 (Competitor vs. Distractors). The left panel indicates that the competitor and the distractors start to diverge at about 100ms until the 775ms mark. The right panel indicates that the time-cluster where the divergence occurs is far outside the distribution of 10000 permutation tests.

Discussion

The analyses reported in Experiment 1 replicated the results of many other experiments that examined phonological competition effects (c.f. Experiment 3 & 4 in Huettig & McQueen, 2007). As the spoken target unfolds, participant's attention is drawn to the target and the phonological competitor. Overtime, as more disambiguating information is received, only the target draws fixations from the listener, while the distractors and competitor receive fewer and fewer fixations. The novel contribution of this experiment is brought from noting that nothing is happening before the spoken target is received. In other words, during the approximately 1.9

second preview (1s[display with no sound]+682[click on]+200[pause]) before the presentation of the spoken target, all four words have similar fixation proportions. This indicates that even if participants access linguistic information from the words in the display, such as noting that the orthography of some words is similar or accessing the phonology of the words while reading them, they do not use this information until they hear the auditory input.

Experiment 1 indicates that simply by having participants look at visual referents is not enough to make them draw associations based on orthographic or phonological overlap. Although there might be concerns about participants activating phonological representations though implicit naming from the visual input, the results of this experiment indicate that visual attention is guided primarily by the ambiguous signal received through the spoken input and not through implicit naming. This experiment indicates that the presence of visual information is not enough to make participants make quick associations between word-forms. One possible reason why participants did not notice that two words in the display were lexically related is that the preview was too short. Typical fixation duration during silent reading or visual search is in the range of 200-250ms (Sereno & Rayner, 2003), which allows enough time for participants to read the words in the display but not enough time to draw any associations between those words. As will be shown in Experiment 5, this explanation does not hold much weight, since even with a longer preview duration, no competitor effects were observed during preview. Moreover, if one were to use longer preview (longer than say 15s), then even if competitor-like effects are observed, we would not be able to conclude that the competition arises during rapid processing of word-forms or simply because of meta-linguistic judgement. This is because with exceedingly longer previews, participants are essentially put in a position where they have nothing more to do but find associations.

The general conclusion that should be drawn from Experiment 1 is that it is the speech signal that initiates the mapping process to the visual-word forms, which overall should be considered a favorable outcome for the linking hypothesis.

Experiment 2

The purpose of this experiment is to test if linguistic representations stored in working memory guide competition effects. A critical distinction of this experiment from the VWP is that instructions indicating which target to click on are presented before the presentation of the display. Although there is no incremental delivery of the target, the expectation is that one will observe similar competition effects that are seen in Experiment 1. Note that this prediction is not supported by the linking hypothesis, since the linking hypothesis posits that competition effects arise because of ongoing lexical processing.

Methods

Materials

Same recordings used in Experiment 1 were used in Experiment 2.

Participants

A total of 21 Binghamton undergraduate students took part in Experiment 2 in order to fulfil a psychology course requirement. All participants were native speakers of English and had normal or corrected-to-normal vision and no hearing impairments. Four participants were bilingual, speaking mainly English at home, with the remaining being monolingual.

Procedure (see Figure 43b at the end of the document for trial structure)

Eye-tracking recording parameters and setup is the same as in Experiment 1.

At the start of each trial, participants were presented with a fixation cross in the center of the screen. Participants were asked to orient their eyes towards the cross when it appears. One second after the presentation of the fixation cross, the audio input was presented through the headphones instructing subjects which target to click (e.g., *Click on lamp*). 1000ms after the offset of the spoken target, the display with the four words appeared on the screen. Participants were instructed to look at the words immediately when they appear on the screen and to use the mouse to select the target word. The trial ended when a word was selected.

Results

Fixations for 0.3% of the trials for which participants selected a word other than the target were removed. Average mouse click time was 1375ms from the onset of the word display. The top panels of Figure 14 plot the fixations proportions from the onset of the trial until 6000ms. The bottom panels plot a zoomed-in portion between the two vertical lines, which corresponds to the time 200ms from the onset of the word display until 4800ms. Bilingual participants performed similarly to monolingual participants, so the data used in figures and analyses includes all subjects.

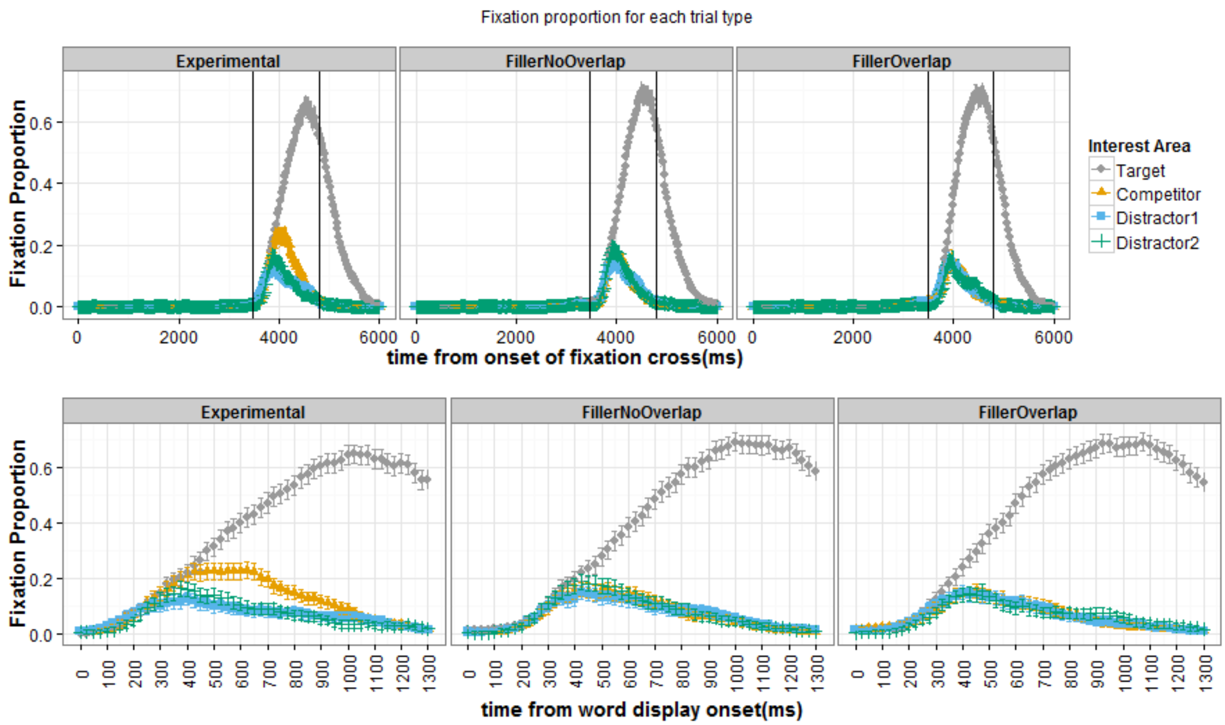


Figure 13: Fixation proportions for Experiment 2 from the start of the trial (top three panels) until 6000ms. The bottom panels represent the fixation proportions from the onset of the word display until 4800ms into the trial.

The following analyses compare the target vs. competitor and the competitor vs. the average of the two distractors in Experimental trials from the onset of the word display.

Target vs. Competitor

Fixations between the target and competitors are compared using GCA and permutation cluster analysis. Table 6 shows the structure of all the GCA models that have been tested, the percentage variance explained by each model, and the Aikake Information Criterion (AIC).

Figure 14 presents a graphical representation of the models' fit to the data.

Table 6: Models tested for Target vs. Competitor comparison in Experiment 2.

Model Type and Name	Model Structure	R^2	AIC
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.361	-1063.6
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.838	-3922.7
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.936	-5774.7
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.965	-6918.3
Model 5 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2|Subject:Condition)	.958	-6603.9
Model 6 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2|Subject)+ (1+T1+T2+T3|Subject:Condition)	.965	-6923.4

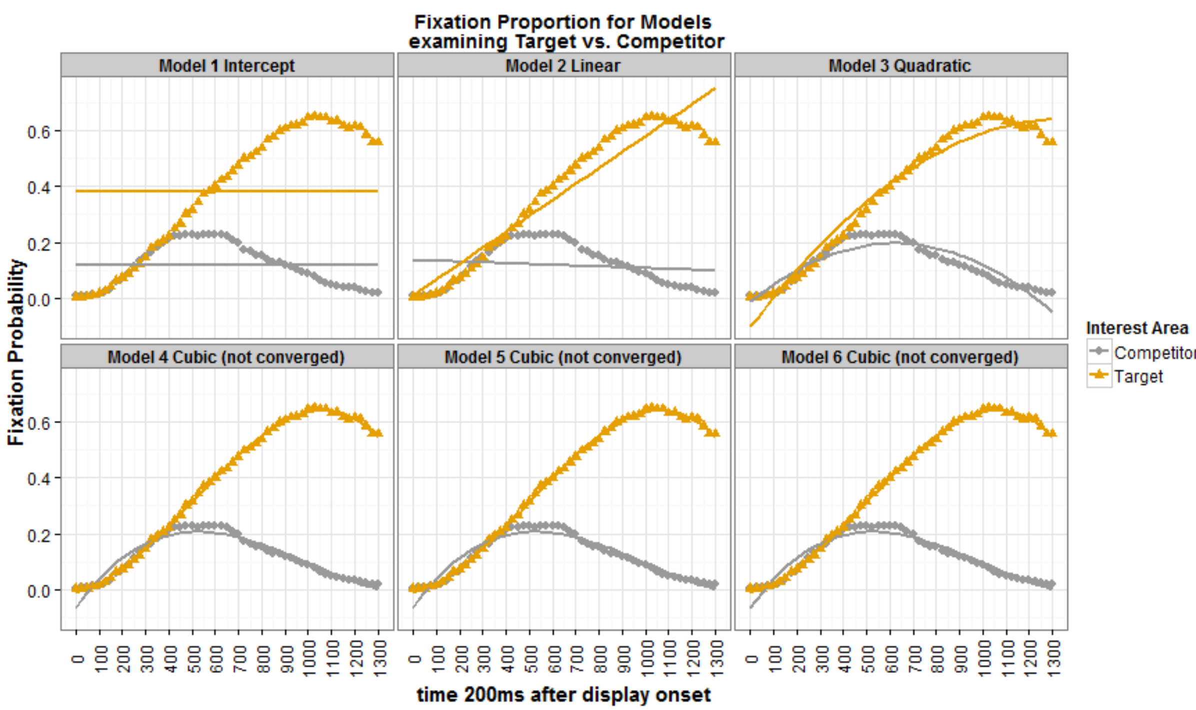


Figure 14: Predictions of the fitted models to the data in Experiment 2 (Target vs. Competitor). The lines represent the predicted values for the models, while the shapes are the mean data for each time point

For GCA models, although Model 6 has the best fit to the data when it comes to explained variance and AIC, the models with the cubic component failed to converge, so Model 3, with quadratic component is reported below in Table 7.

Table 7: Model 3 output for Target vs. Competitor in Experiment 2

Predictor	β	S.E.	t	p
(Intercept)	.118	.011	10.53	<.0001
Condition (Target vs. Competitor)	.263	.015	16.59	<.0001
T1	-.079	.074	-1.07	.291
T2	-.509	.062	-8.18	<.0001
T1:Condition	1.664	.105	15.78	<.0001
T2:Condition	.124	.086	1.45	.155

GCA is generally in line with the output presented for Target vs. Competitor in Experiment 1. A strong interaction between the linear polynomial and condition is observed, with targets receiving more fixations overtime and competitors having fewer fixations as the trial progresses.

The bootstrapped cluster-based permutation analysis indicates that the competitor and the target diverge at the 475ms mark (see Figure 15). The cluster of summed t-tests (Sum=538; see cluster 1) is far outside of a 10000 sample distribution of permutation tests (Mean=-.0; 95% CI [-111,109], p=0).

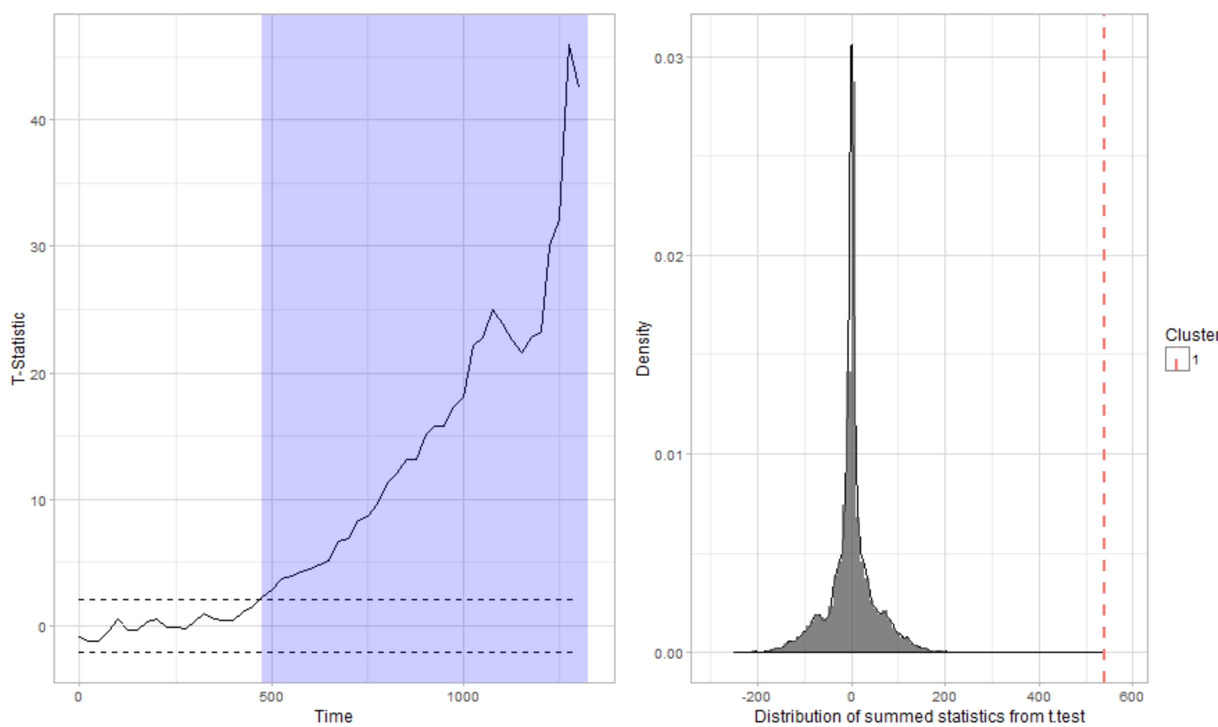


Figure 15: Permutation-cluster outcome for Experiment 2 (Target vs. Competitor). The left panel represent the t-statistic for every 25ms time-bin. The time-cluster from 475ms to 1325ms is

significant. The right panel indicates the distribution of t-sums from 10000 simulations. This distribution indicates that the time-cluster has a very small probability of occurring by chance.

Competitor vs. Distractors

Fixations between competitor and the average of the two distractors are compared. First, several models were fitted to the data (see Table 8 and Figure 16).

Table 8: Models tested in Experiment 2, comparing the fixations between Competitor and the average of the two distractors

Model Type and Name	Model Structure	R^2	Aikake Information Criterion
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.164	-5131.9
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.224	-5207.8
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.702	-7145.6
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.811	-7993.3
Model 5 (quartic)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+ (1+T1+T2+T3+T4|Subject)+ (1+T1+T2+T3+T4|Subject:Condition)	.854	-8450.4

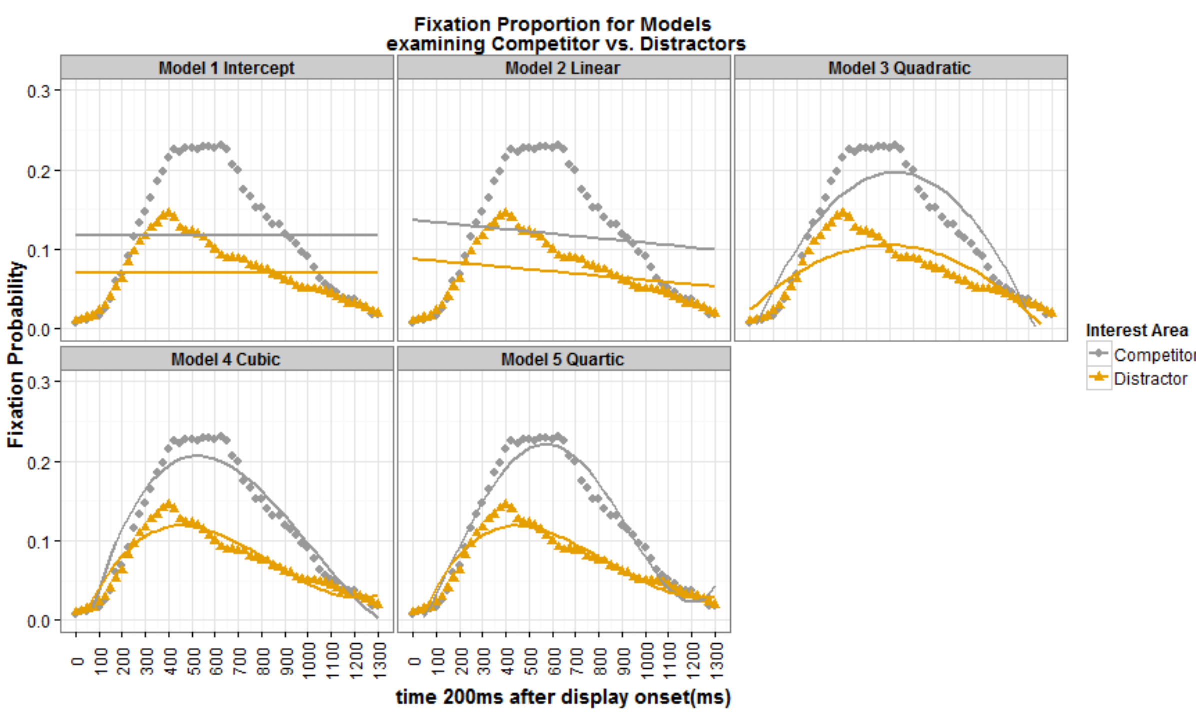


Figure 16: Predictions of the fitted models to the data in Experiment 2(Competitor vs. Distractors). The lines represent the predicted values for models, while the shapes are the mean data for each time point.

For GCA models, Model 5 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 5 can be found in Table 9.

Table 9: Model 5 output for Competitor vs. Average Distractor in Experiment 2

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.118	.005	22.21	<.0001
Condition (Competitor vs. Distractor)	-.047	.005	-9.95	<.0001
T1	-.079	.029	-2.76	.0089
T2	-.509	.025	-19.79	<.0001
T3	.157	.029	5.32	<.0001
T4	.122	.019	6.18	<.0001
T1:Condition	.005	.036	0.15	.8793
T2:Condition	.289	.032	9.06	<.0001
T3:Condition	-.025	.037	-0.67	.5086
T4:Condition	-.133	.027	-4.78	<.0001

GCA shows a strong interaction between quadratic polynomial and condition, which can be observed from the fact that the fixation curve for competitor has a higher peak than the curve for average distractor. Although Model 5 points to an interaction between quartic polynomial and condition, it is important to note that that interaction seems to be driven by responses at the endpoint competitor fixation curve, likely overfitting a few responses (note the change at the end between Model 4 and Model 5 in Figure 16).

The bootstrapped cluster-based permutation analysis indicates that the competitor and the distractor diverge at the 300ms mark (see Figure 17). The cluster of summed t-tests (Sum=124; see cluster 1) is far outside of a 10000 sample distribution of permutation tests (Mean=-.15; 95% CI [-37,35], p=0).

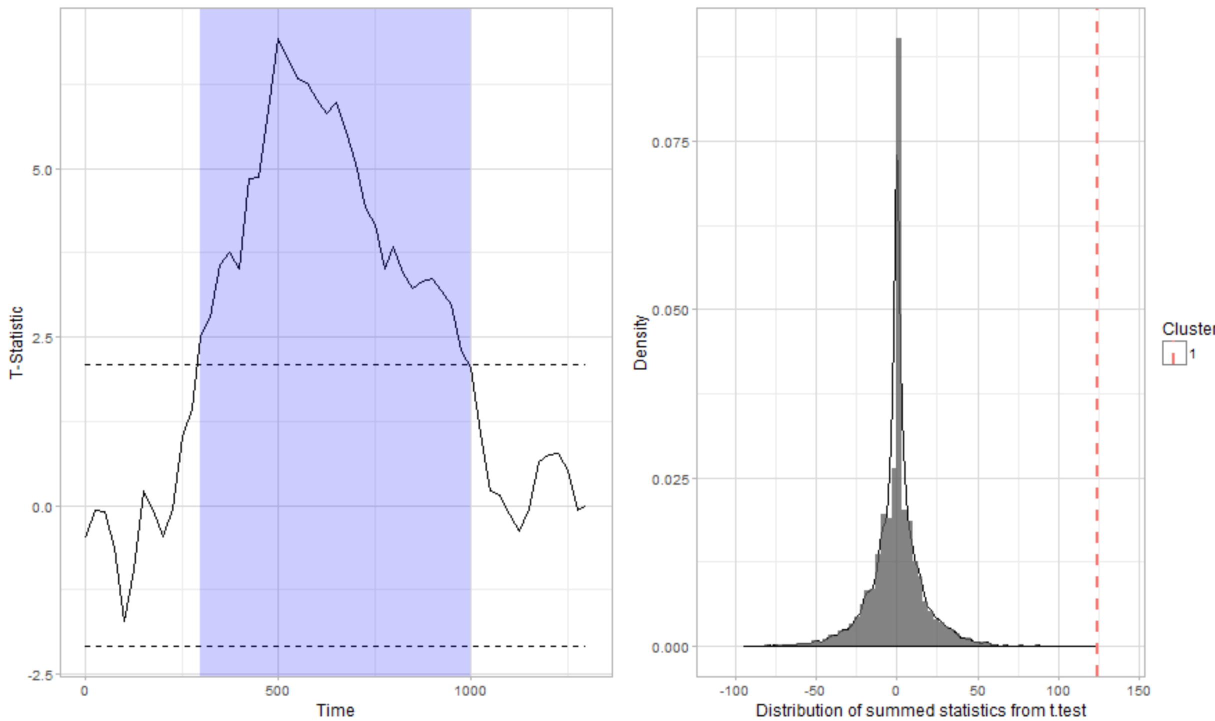


Figure 17: Permutation-cluster outcome for Experiment 2 (Competitor vs. Distractors). The left panel represent the t-statistic for every 25ms time-bin. The time-cluster from 300ms to 1000ms is significant. The right panel indicates the distribution of t-sums from 10000 simulations. This distribution indicates that the time-cluster has a very small probability of occurring by chance.

Discussion

One notable difference between Experiment1 and Experiment 2 is that the divergence point between competitor and distractors in Experiment 2 occurs at the 300ms mark while in Experiment 1 it occurs at the 100ms mark. Although one could conclude from this that visual search is more efficient in Experiment 1 than in Experiment 2, it is very likely that this difference is purely methodological but nonetheless interesting. In Experiment 1, the divergence is measured 200ms after spoken target onset. Moreover, at the onset of the spoken target in Experiment 1, participants' eyes are already located at one of the 4 possible targets. In other words, they don't have to perform an initial saccade from the center of the screen to one of the four quadrants, which means that visual search has a head start of about 200ms. On the other hand, in Experiment 2 participants have to start the visual search at the center of the screen, which means that the visual search should account for the time it takes to initiate an extra saccade. This methodological difference is important because, after accounting for the extra saccade, the divergence in Experiment 2 occurs at about the 100ms mark rather than 300ms mark, making the divergence points across the two experiments equivalent.

The lack of a difference in divergence time between the two experiments is interesting in that it indicates that in Experiment 1 participants do not necessarily rely on the contents stored in working memory to guide their visual search. In Experiment 1, participants are given a 1.9s preview window, which should allow them to store the orthographic representations and their locations in working memory. Once participants are presented with the onset of the spoken target, they don't have to engage in an actual visual search by initiating saccades to the orthographic representations all over again. Rather, listeners can map the phonological representations to the words' locations stored in working memory and then initiate the saccade to the correct target. For example, this means that immediately when a listener hears the onset /k/, they can rule out the distractor words LUNCH and STAIR and the locations of those words and instead focus on the words CLOTH and CLOUD. Nonetheless, listeners don't necessarily engage their visual working memory. Instead, listeners perform the visual search as if they have not seen those words before. One reason why a more efficient, working memory-based strategy search is not selected is because visual working memory is actually very poor (Beck, Hollingworth, & Luck, 2012; S J Luck & Vogel, 1997; Steven J. Luck & Vogel, 2013; McCarley et al., 2003) and in hand-eye tasks participants prefer a lazy, serializing strategy for visual search (Ballard et al., 1995; Sprague et al., 2007). Ballard, Hayhoe, and Pelz (1995), gave participants a block task, in which they were asked to recreate the pattern of blocks in the model area using blocks from the resource area (see Figure 18 left panel). When performing this task, four strategies could be selected. The most "efficient" strategy (PD) would be to look at the model area, remember the pattern and location of the blocks, and then recreate the model area in the workspace area without ever looking at the model area again. On the other hand, the strategy that would require the largest number of eye-movements (MPMD) is to look at a block in the model area, pick a matching block from the resource area, confirm that the right block was picked by looking back at the model area, and lastly, move the block to the workspace area. Ballard, Hayhoe, Pelz (1995) found that participants generally rely on the memoryless strategy (MPMD). One of the reasons why participants rely on this strategy is because initiating multiple saccades to acquire information on-line might be considered a less effortful strategy than storing representations in working memory. When blocks are located further away from each other, so that participants have to turn their head rather than only move their eyes, the costs of the memoryless strategy increase, so participants rely more on the other, memory-reliant strategies.

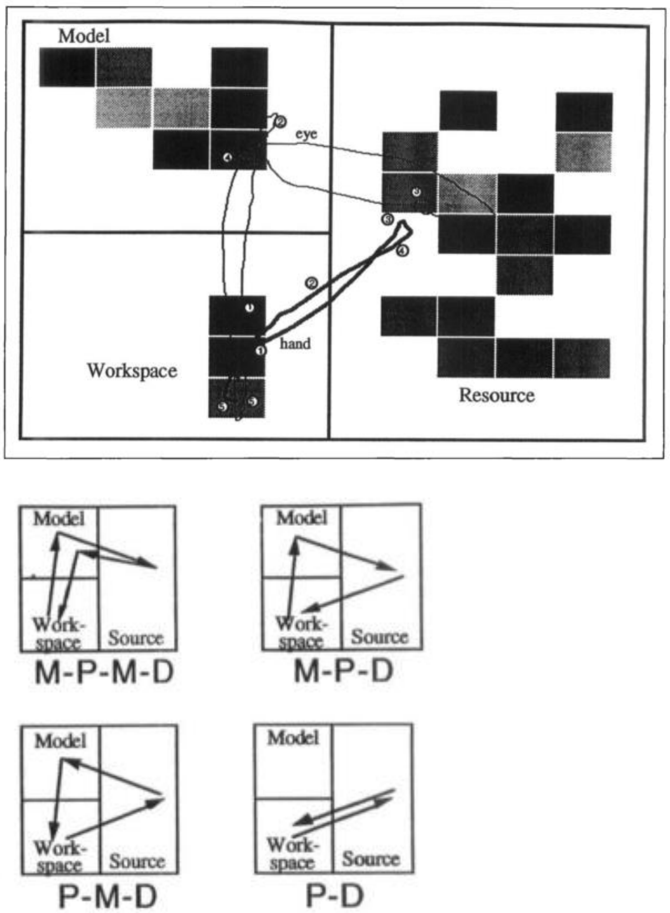


Figure 18: Displays used in the experiments by Ballard, Hayhoe, and Pelz (1995) and possible strategies that could be selected to perform the task.

A lack of a difference in divergence between Experiments 1 and 2 conflicts with the model proposed by Huettig, where the phonological representations are mapped in working memory to location representations and orthographic representations. The mapping doesn't have to occur in working memory. The mapping can occur on the physical canvas presented in the visual display without a need to store multiple representations in working memory.

This conclusion does not imply that working memory cannot be used to map representations to each other and guide rapid search. Work by Altmann (2004), using the Blank Screen Paradigm, indicates that working memory is being used by participants, but it is being used because there is no other alternative way that would help listeners guide their eye movements in the Blank Screen Paradigm. Moreover, as will be shown in Experiment 5, it seems that with a long enough preview, participants do store orthographic representations and their locations in working memory. This indicates that the role of working memory in binding different representations together is more complex and nuanced. There are several things that could be done in order to arrive at a better understanding how working memory affects the speed of search. One recommendation for a future study would be to collect measures of working memory performance and includes those measures as predictors for how fast participants show a competitor effect. Recent work by Huettig and Janse (2016) indicates that working memory performance is related to how quickly anticipatory fixations are initiated when processing spoken sentences. Other suggestions for how working memory representation affect visual search have been proposed below (see discussion in Experiment 4).

Results in Experiment 2 find similar competition effects as in Experiment 1, despite the fact that the spoken target is presented before the onset of the display. Competition effects in the absence of any concomitant speech input create a difficult scenario for the linking hypothesis to reconcile. As pointed out in the introduction, a more simple explanation for the competition effect found in Experiment 2, is that continuous-like competition effects are an artifact of how the data is aggregated rather than because of ongoing lexical competition. When participants' eyes land on a competitor word (relative to a distractor word) in the display, it takes them longer to disengage from that word and move their search on to the next candidate. Longer first fixations to competitor interest area relative to distractor interest areas will thus create the semblance of a competitor effect without any actual competition.

Perhaps the most interesting aspect in Experiment 2 is that it opens up the possibility to test how multiple sources of information stored in working memory guide visual attention. For instance, participants are given two auditory targets to store in working memory [e.g., SPOON and CUP] and are told that the visual search will involve only the second target (recall for the first target could be tested by asking subjects to name it at the end of the trial). The visual display, presented at a later time, contains the target [CUP], a lexical competitor to SPOON [SPOOL] and two unrelated distractors. This experimental manipulation would be similar to a scenario where participants hear a sentence in which multiple objects are mentioned, but only a few are relevant in the visual context (e.g., “Bring me a spoon for the tea cup”, where only the cup is visually available). In this scenario, when participants are told to store in WM the first target SPOON but ignore it for visual search purposes, will there still be residual activation from the ignored target and more looks to the competitor? If yes, this would indicate that almost anything that occupies our current working memory guides visual search. If there are no fixations to the competitor, this would indicate that we can suppress, for visual search purposes, representations that are stored in working memory.

Experiment 3

Experiment 3 is a follow-up on Experiment 2 and will test if an addition of cognitive load will interfere with the visual search of the target in the visual display. Previous work indicates that there are different levels of representations in working memory (Dahan & Tanenhaus, 2005; Huettig & Hartsuiker, 2008; Huettig et al., 2012; Huettig, Olivers, et al., 2011a), and the activation or storage of those representations depends on the visual display context (Huettig & McQueen, 2011), the duration of preview (Huettig & McQueen, 2007), and typicality with which agents in the display engage in specific behaviors (e.g., we expect a detective to use a spyglass and not jinx another person; see Coordination Interplay Account for sentence processing

Knoeferle & Crocker, 2006). Experiment 2 indicates that participants' visual search is affected by storing the linguistic representation in working memory. Nonetheless, central working memory capacity is a limited resource, with ~3-5 items capacity rather than the magic number 7 proposed by Miller 50 years ago (see review by Cowan, 2013). The question asked in this experiment is if working memory can be taxed in such a way as to eliminate the strength of the activated linguistic representations. In this experiment, participants are presented with the verbal instructions and an arithmetic problem, after which they are presented with the 4-word display. The cognitive load brought by the processing of the arithmetic problem should interfere with storing the linguistic representation of the target in working memory, thus leading to a weaker competition effect. Note that this does not presuppose that participants will completely eliminate the to-be-searched target from working memory. It simply specifies that the activation of linguistic representations in working memory will not be as strongly activated, and that visual search will not be as efficient in disengaging from the distractor words.

Methods

Materials

Same recordings used in Experiment 1 were used in Experiment 3.

Participants

A total of 21 Binghamton undergraduate students took part in Experiment 3 in order to fulfil a psychology course requirement. All participants were native speakers of English and had normal or corrected-to-normal vision and no hearing impairments. Two participants were bilingual, speaking mainly English at home, with the remaining being monolingual.

Procedure (see Figure 43c at the end of the document for trial structure)

Eye-tracking recording parameters and setup is the same as in Experiment 1.

At the start of each trial, participants were presented with a fixation cross in the center of the screen. Participants were asked to orient their eyes towards the cross when it appears. One second after the presentation of the fixation cross, the audio input is presented through the headphones instructing subjects which target to click (e.g., *Click on lamp*). Simultaneously with the presentation of the audio input, the fixation cross display is replaced with a display with an arithmetic problem presented in the center of the screen. Participants were required to solve the problem while they listen to the speech input and are told that they were asked to confirm the solution to the problem at the end of the trial. A recent meta-analysis by Block, Hancock, and Zakay (2010), found that manipulations of cognitive load that require subjects to engage in intentional and effortful processing (e.g., solve anagrams or arithmetic problems), seems to be very effective across a variety of tasks. Nonetheless, identifying arithmetic problems with just the right amount of difficulty for a subject is problematic. Based on a study by Brown (1997), the difficulty of the problems was set so that participants have to do addition or subtraction problems using 2-digit numbers (e.g., 15-73), making these high difficulty arithmetic problems. Half of the problems involved addition and half involved subtraction. Problems were created by randomly sampling from a uniform distribution of numbers between 10 and 99. At the onset of the display with the arithmetic problem, participants also heard the production (e.g., *Click on lamp)* letting them know which target they were required to select in a later display. The display with the problem was presented for 3000ms, after which the display with the four words appeared on the screen. Participants were instructed to look at the words immediately when they appear on the screen and to use the mouse to select the target word (e.g., *lamp*). Once the word was selected, a display with the correct/incorrect answer to the arithmetic problem was presented. In half of the trials, the correct solution appeared on the screen while in the other half the incorrect solution was presented. Incorrect solutions were created by adding or subtracting from the correct solution a number from a uniform distribution of numbers between 1 to 5. Participants were required to click *yes* if the solution is correct or *no* if it is incorrect. Once a response was made, the trial ended and the next trial began.

Results

Fixations for 0.3% of the trials for which participants selected a word other than the target were removed. Average mouse click time was 1704ms from the onset of the word display. All participants performed above chance on math problems, with an average accuracy across all participants of 82%. Average click time for math problems with correct responses was 1688ms and for problems on which participants got the math problem incorrect the average click time was 1783ms. An LMER model with subject and item random intercepts indicated that there is no difference in click times between trials on which participants got the math problem wrong and trials on which the math problem was solved correctly (estimate=-22.35, SE=42.43, t=-.0527, p=.599). Lack of an effect between correct and incorrect trials was also found for log-transformed click time responses (estimate=-.008, SE=.0232, t=-.351, p=0.726).

The top panels of Figure 19 plot the fixations proportions from the onset of the trial until 8000ms. The bottom panels plot a zoomed-in portion between the two vertical lines, which corresponds to the time 200ms from the onset of the word display until 5500ms. Bilingual participants performed similarly to monolingual participants, so the data used in figures and analyses includes all subjects.

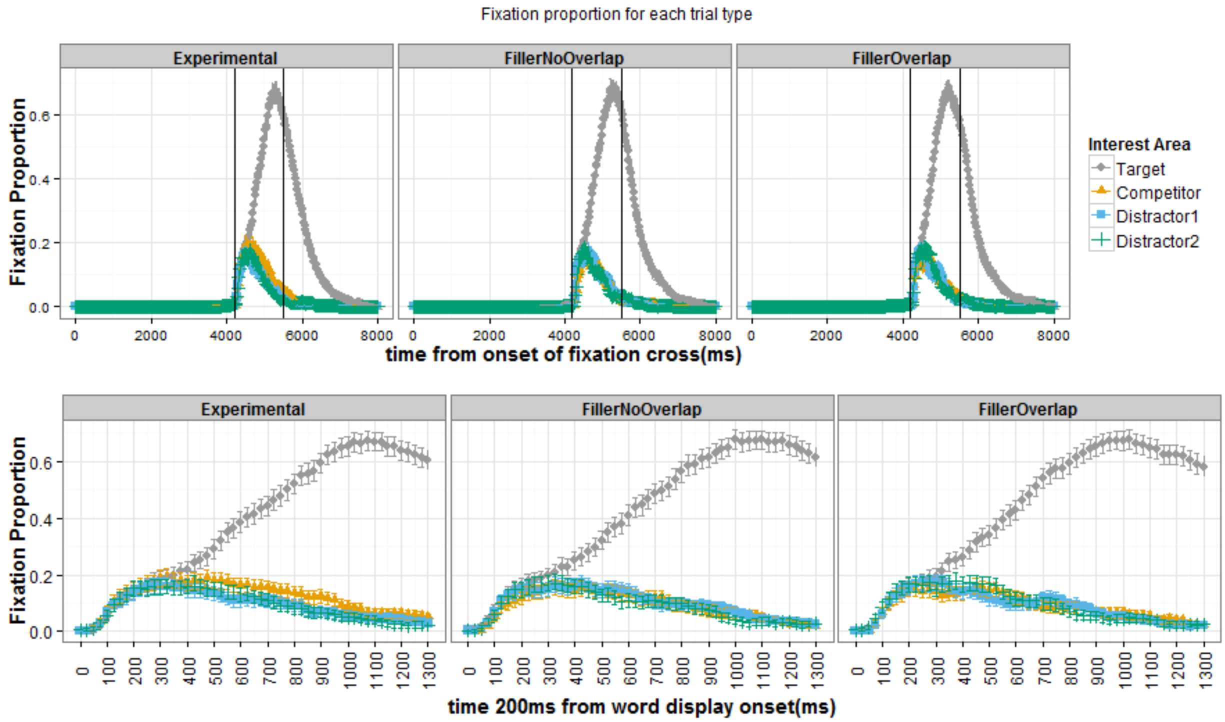


Figure 19: Fixation proportions for Experiment 3 from the start of the trial (top three panels) until 8000ms. The bottom panels represent the fixation proportions 200ms from the onset of the word display until 5500ms into the trial.

Comparing Figure 19 for Experiment 3 and Figure 13 for Experiment 2 one can notice that there was a reduced competitor effect in Experiment 3. Before committing to any formal models, one other exploratory graph is examined, where the fixation curves for Experimental trials are broken down by accuracy at the math problems (see Figure 20).

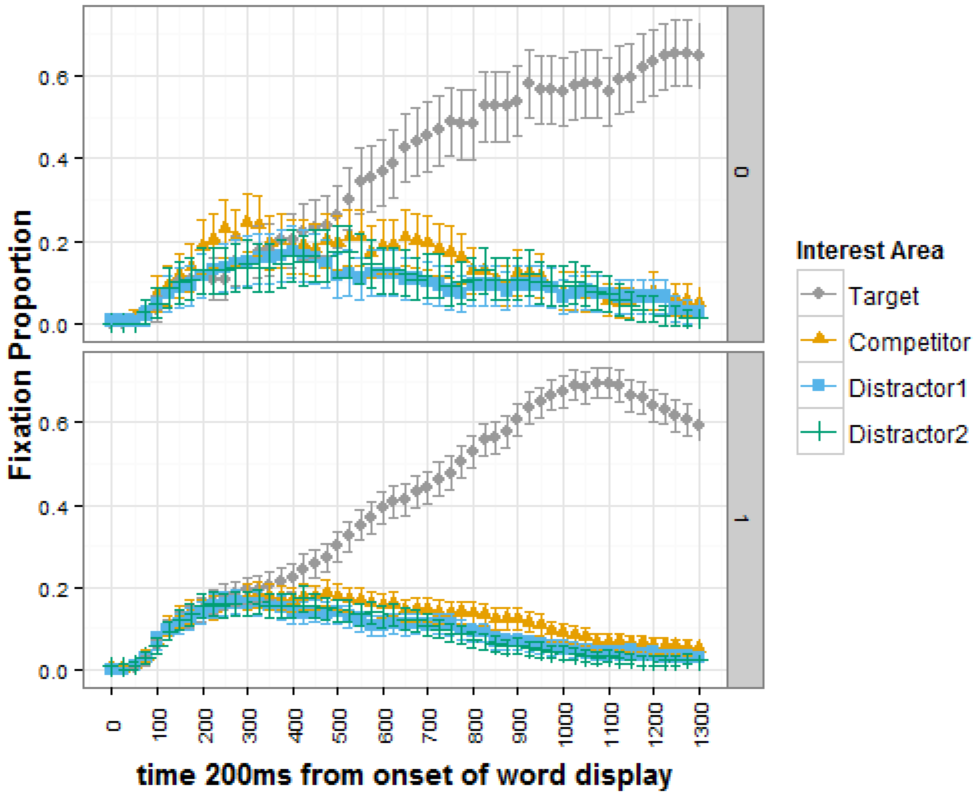


Figure 20: Fixations broken by accuracy at math problem. Top panel depicts fixation proportion for trials that participants got the math problems incorrect, while bottom panel represent fixation proportions for trials where participants got the math problems correct.

Figure 20 indicates an even more reduced competitor effect for trials participants got correct. The variability in fixations for trials participants got incorrect is larger, but the general pattern indicates that in incorrect trials participants had greater competitor effects than in trials that subjects got correct. Participants reported some of the problems to be very difficult to solve within a span of 3 seconds, so it is likely that for those problems they did not have a stored solution in working memory when the display with the four words appeared. In the following analyses, fixations between the target vs. competitor and competitor vs. average of two distractors are compared only for trials for which participants correctly identified the solution.

Target vs. Competitor

Fixations between the target and competitor are compared using GCA. Table 10 depicts the structure of all the models that have been tested, the percentage variance explained by each model, and the Aikake Information Criterion (AIC). Figure 21 presents a graphical representation of the models' fit to the data.

Table 10: Models tested for Target vs. Competitor comparison in Experiment 3

Model Type and Name	**Model Structure**	**R^2**	**AIC**
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.397	-1171.7
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.869	-4367.1
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.926	-5442.8
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.957	-6445.3
Model 5 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2|Subject:Condition)	.947	-5915.3
Model 6 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2|Subject)+ (1+T1+T2+T3|Subject:Condition)	.957	-6453.3

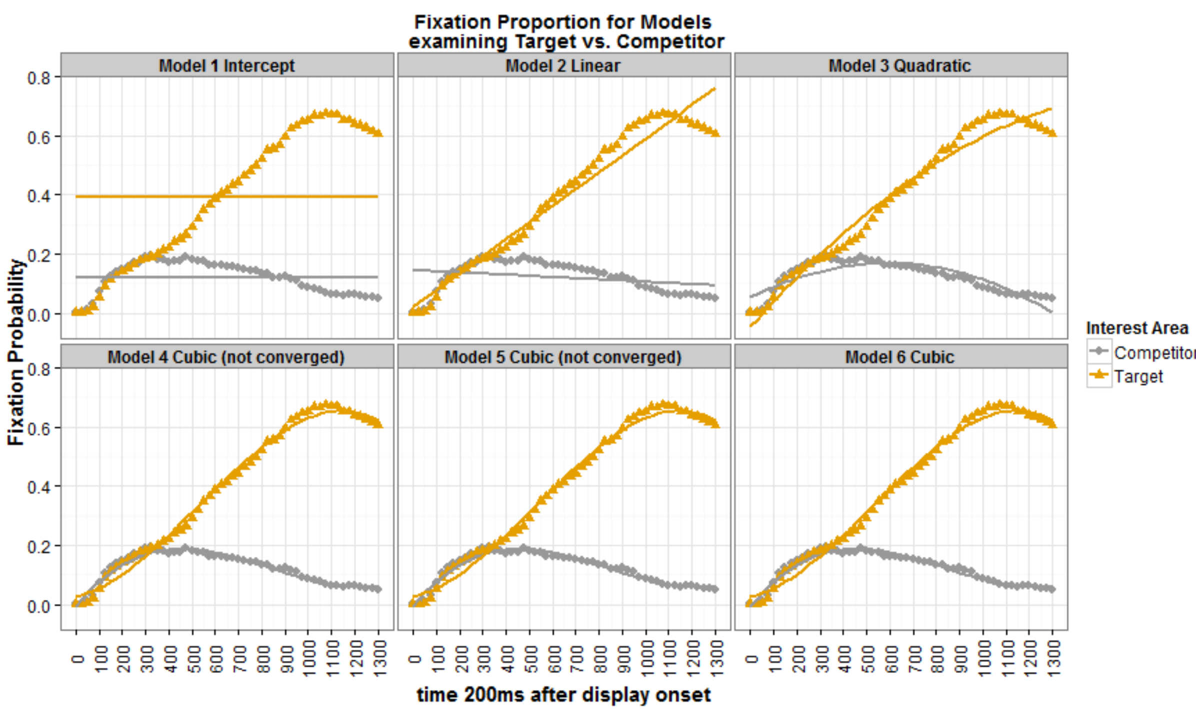


Figure 21: Predictions of the fitted models to the data in Experiment 3 (Target vs. Competitor). The lines represent the predicted values for the models, while the shapes are the mean data for each time point.

For GCA models, Model 6 has the best fit to the data when it comes to explained variance and AIC. The summary of Model 6 can be found in Table 11.

Table 11: Model 6 output for Target vs. Competitor in Experiment 3.

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.118	.011	10.84	<.0001
Condition (Target vs. Competitor)	.279	.014	19.68	<.0001
T1	-.111	.059	-1.865	.0696
T2	-.306	.066	-4.590	<.0001
T3	.180	.046	3.94	.0003
T1:Condition	1.718	.083	20.74	<.0001
T2:Condition	.068	.091	.751	.458
T3:Condition	-.428	.065	-6.613	<.0001

Similar to GCA models in the first two experiments, the interaction between linear intercept and condition seems to be the main driver of the differences in fixation proportions between the target and competitor. Model 6 in Table 11 also indicates that there is a significant interaction between the third polynomial and condition. Figure 21 indicates that this interaction is driven by trying to fit some of the endpoint fixations.

The bootstrapped cluster-based permutation analysis indicates that the competitor and the target diverge at the 425ms mark (see Figure 22). The cluster of summed t-tests (Sum=489; see cluster 1) is far outside of a 10000 sample distribution of permutation tests (Mean=.96; 95% CI [-117,120], p=0).

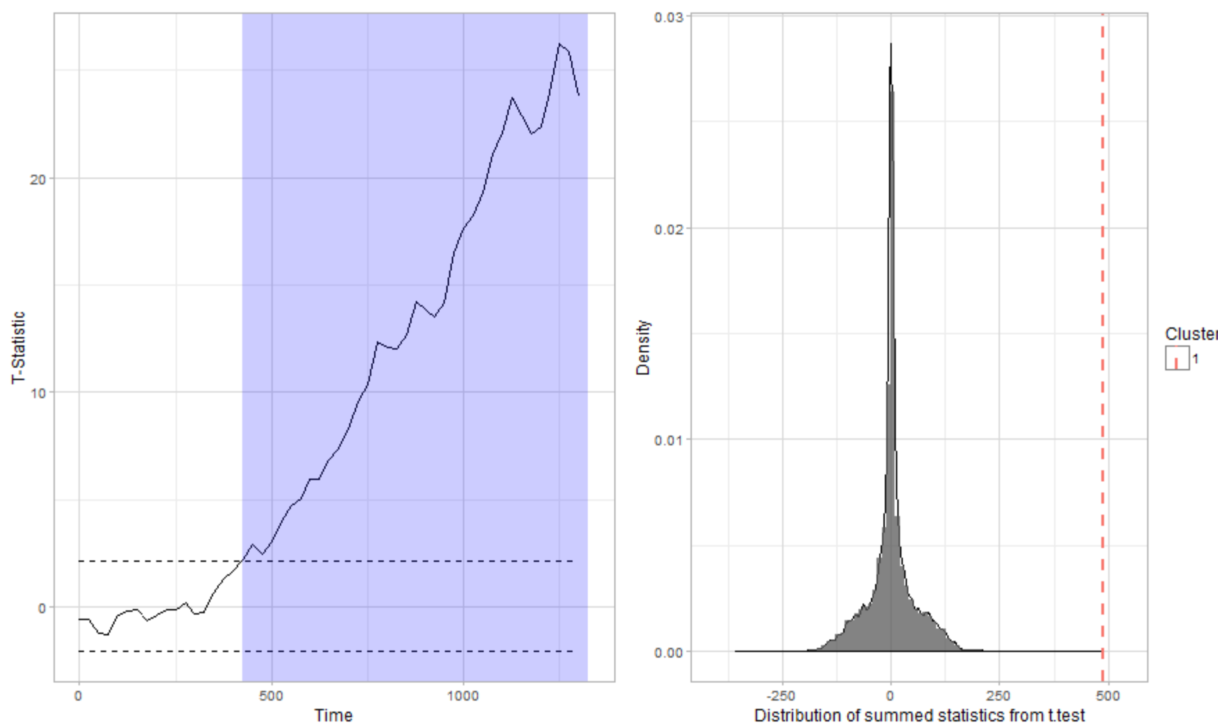


Figure 22: Permutation-cluster analysis in Experiment 3 (Target vs. Competitor). The left panel represent the t-statistic for every 25ms time-bin. The time-cluster from 425ms to 1325ms is significant. The right panel indicates the distribution of t-sums from 10000 simulations. This distribution indicates that the time-cluster has a very small probability of occurring by chance.

Competitor vs. Distractors

Fixations between competitor and the average of the two distractors are compared. First, several models were fitted to the data (see Table 12 and Figure 23) and the output for the models with the best fit (which also converged is presented).

Table 12: Models tested in Experiment 3, comparing the fixations between Competitor and the average of the two distractors.

Model Type and Name	**Model Structure**	**R^2**	**Aikake Information Criterion**
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.114	-5163.6
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.306	-5558.5
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.596	-6613.1
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.780	-7778.5

Model 5 (quartic)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+ (1+T1+T2+T3+T4|Subject)+ (1+T1+T2+T3+T4|Subject:Condition)	.830	-8222.9

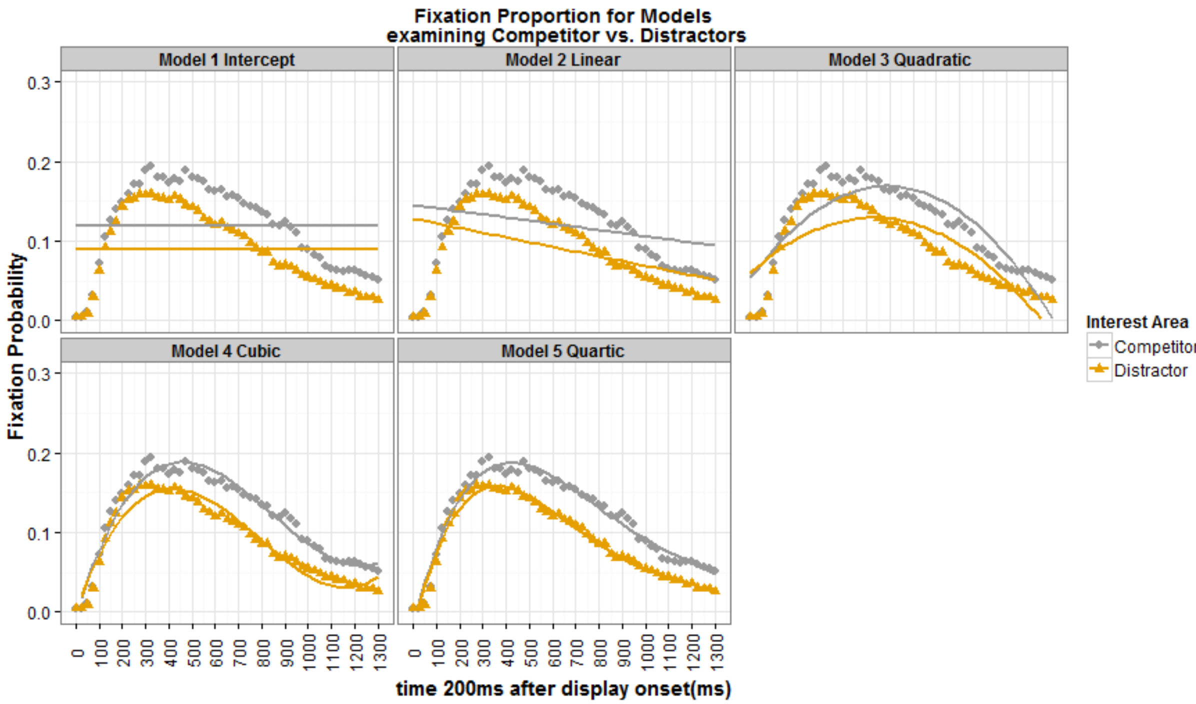


Figure 23: Predictions of the fitted models to the data in Experiment 3 (Competitor vs Distractors). The lines represent the predicted values for models, while the shapes are the mean data for each time point

For GCA models, Model 5 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 5 can be found in Table 13.

Table 13: Model 5 output for Competitor vs. Distractors in Experiment 3

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	0.1184	.005	23.62	<.0001
Condition (Competitor vs. Distractor)	-.030	.005	-5.54	<.0001
T1	-.111	.044	-2.49	.0180
T2	-.306	.034	-9.12	<.0001
T3	.180	.034	5.27	<.0001
T4	-.044	.026	-1.69	.0982
T1:Condition	-.074	.046	-1.59	.1232
T2:Condition	.079	.039	2.02	.0535
T3:Condition	.021	.044	0.48	.6345
T4:Condition	-.018	.035	-0.52	.6050

The GCA model in Table 13 indicates that only the Condition effect is significant, with competitors having a higher fixation proportion overall than the distractors. This is not particularly interesting, because the main effect of condition in the model does not allow us to evaluate changes over time. Moreover, the calculation of the S.E. for the main effect of condition is done after collapsing across time, leading to small S.E. and creating the impression that one deals with a large effect (based on the t-value), when the effect is actually very small (the coefficient is only .03). The interaction terms indicate a marginally significant effect between T2: Condition, with competitors having slightly higher peak than distractor. Otherwise, no other significant interactions have been found.

The bootstrapped cluster-based permutation analysis indicates that there is a difference between competitor and the distractor fixations, and that this difference emerges reliably at 700-1075ms time-cluster (see cluster 2, Figure 24). The cluster 2 summed t-tests (Sum=48) is outside of a 10000 sample distribution of permutation tests (Mean=.03; 95% CI [-21,21], p=.001). The other time-clusters are not significant and are all likely to have emerged by chance (all p>.15).

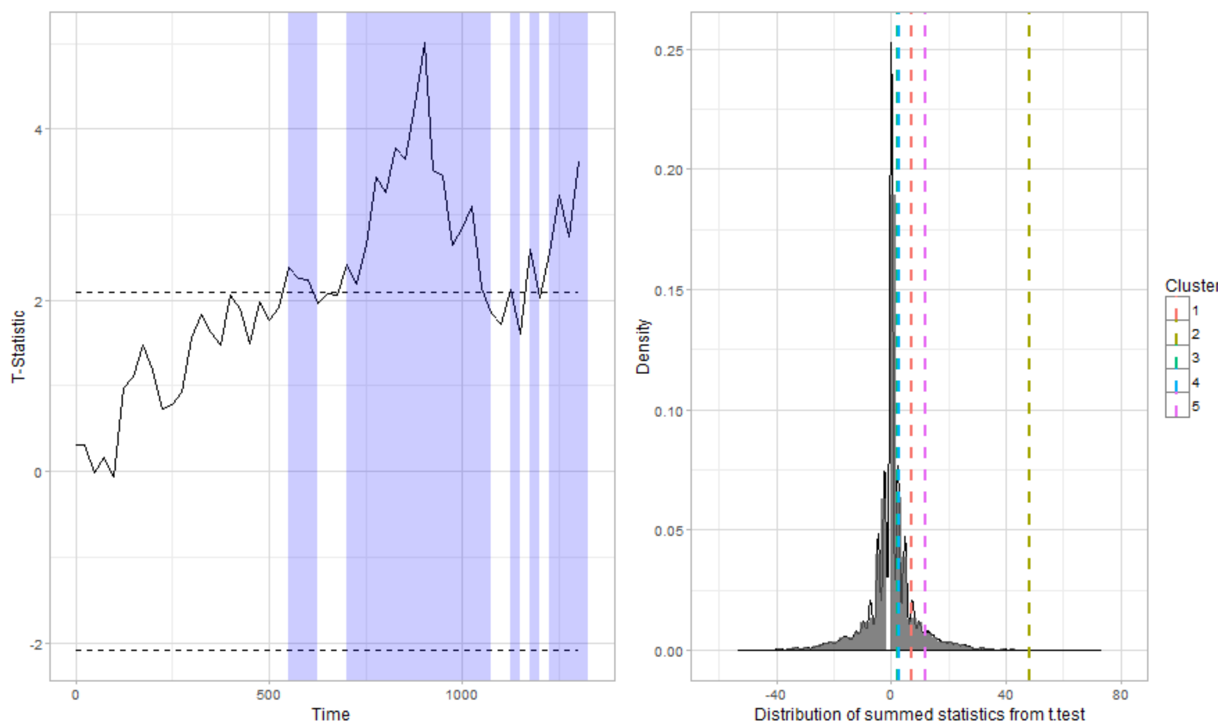


Figure 24: Permutation-cluster analysis outcome in Experiment 3 (Competitor vs. Distractors). The left panel represent the t-statistic for every 25ms time-bin. The time-cluster 2 from 700ms to 1075ms is significant. All the other clusters are not significant. The right panel indicates the

distribution of t-sums from 10000 simulations. This distribution indicates that the time-cluster 2 has a very small probability of occurring by chance.

Discussion

Overall, the patter of results in this experiment points to a competitor effect that is much smaller and delayed than in the previous two experiments. The GCA models indicate that no time-dependent competitor effect is observed and that the effect emerges only if one collapses across time-bins. The more sensitive permutation cluster analysis indicates that there is a small competitor effect emerging at the 700ms mark, which lasts until 1075ms. Note that the size of this effect is also smaller when you compare the t-sum statistic relative to what was observed in Experiments 1&2. The difference in the inference outcomes between GCA and permutation analyses likely emerges because the t-tests are essentially the foundation of permutation analyses, while GCA are performed through multi-level modelling. Multi-level models try to account for several sources of random variability, perhaps leading to more robust estimates in the face of Type-I error. Regardless of which inference model is selected as the true measure of an effect (GCA or permutation cluster) the fact remains that the competitor effect in Experiment 3 is smaller than previously seen effect.

One thing to note in the fixation curves in Figure 19 in Experiment 3 relative to Figure 13 in Experiment 2 is that the decline in the competitor effect is driven by a drop in fixation proportion to the competitor and a slight increase in fixation proportions to the distractors. I would like to note that fixation curves in VWP are co-dependent, meaning that a drop in fixations from one condition is generally associated with an increase in fixations to another condition. Hence, we can’t make any inferences on whether the smaller competitor effect is driven because competitors become less active or distractors are better at drawing fixations. Nonetheless, examining the filler trials across the two experiments indicates that distractors in filler trials for Experiment 3 drew more fixations than distractors for fillers in Experiment 2, which makes it more likely that the drop in competitor effect is driven by increased distractibility of irrelevant distractors. One explanation for why the competitor effect is smaller and more delayed is based on the Nillie Lavie load theory of selective attention (Forster & Lavie, 2008; Lavie, Hirst, de Fockert, & Viding, 2004; Lavie & Tsal, 1994; Weast & Neiman, 2010). Lavie et al. (2004) found that a working memory load or a dual-task coordination load increases distractor interference in visual search tasks. The load theory indicates that there is a passive and an active mechanism to selective attention. The passive component is perceptual and is responsible for the exclusion of irrelevant distractors from perception when the visual display has a lot of items in it (i.e., display with high perceptual load). For instance, a participant is asked to identify the circles in the visual display that has 4 circles and 4 squares. The participants' perceptual capacity is only 4 items, so the display with 8 items creates a scenario with high perceptual load. Under the passive mechanism, all the perceptual capacity will be used to process the targets (the circles) and no attention will be given to distractors. The passive mechanism helps to explain why distractors become more salient when the perceptual display has fewer items rather than many items (Lavie et al., 2004; Lavie & Tsal, 1994). The active mechanism underscores the presence of cognitive control when processing the distractors. Under the active mechanism, participants have the ability to suppress their attention to irrelevant distractors. Nonetheless, in conditions of high working memory load, the suppression mechanism is not as effective, hence leading to longer or increased number of fixations given to the irrelevant distractors. For instance, inhibition in an anti-saccade paradigm, where participants have to make saccade opposite from the location where a cue is flashed (generally, when a cue is flashed, bottom-up processing leads to an reflective initiation of a saccade to the location of the cue) is worse under conditions of high working memory load (Roberts, Hager, & Heron, 1994). An active mechanism would explain why the two unrelated distractors show generally higher fixation proportions in this experiment when compared to Experiment 2. The only issue that would be unexplained is why the competitor fixation curve doesn't show higher fixation proportions[4]. Competitor words are also distractors, though they are related to the target. If cognitive load affects how fast participants suppress their attention to irrelevant stimuli, it does not seem clear why the competitor fixation remains unaffected relative to Experiment 2, since it should be more difficult to disengage from the competitor in Experiment 3 rather than in Experiment 2. In other words, under a high cognitive load condition, all three non-target words in the display should receive higher fixations, thus keeping the competitor effect more or less intact.

One answer would be that cognitive load doesn't only affect how fast participants disengage from the distractors but also affects the type of representations that is stored in working memory when processing the auditory stimulus. When participants have to map the contents of working memory to visual word forms, the durations of the fixations they are going to make depends on the amount of overlap between the representations activated from visual input and the contents of working memory. In other words, the more representations (i.e., similarities, such as phonological, shape, semantic similarities) they find between a word in a display and the target stored in working memory, the longer they are going to linger their eyes on that word. Nonetheless, if the representations stored in working memory are disrupted by cognitive load, then the mapping has a lower degree of overlap, and hence the competitor would receive fewer fixations. At the moment, we have very little understanding how representations are bound in working memory and if working memory is indeed the place where the binding occurs. By showing that cognitive load disrupts the binding, we give more credence to the theory that it is working memory that serves the nexus for how different representations are brought together.

[4] Interestingly, Lavie et al. (2004) and Forster and Lavie (2008) found in a letter search task that distractors that are incongruent with the target are as good in their distractibility as distractors that are compatible with the target. Moreover, this lack of difference in distractibility was only present under high working memory load, similarly to what we observe in Experiment 3, where all three distractors show small or no differences in fixation curves among each other.

Nonetheless, further studies should examine how representations can be degraded in working memory.

Experiment 4

The purpose of Experiment 4 is to test how a non-linguistic representation stored in working memory (color) affects competition effects. Previous studies in the visual search literature found that objects or properties (e.g., color) stored in working memory can bias visual attention when those properties or objects are presented in a search display (Downing, 2000; Houtkamp & Roelfsema, 2006; Soto et al., 2005; Soto & Humphreys, 2007), though this bias is not always found (Downing & Dodds, 2004; Woodman & Luck, 2007a) and thought to be driven by how consistent is the stored object in working memory from trial to trial (Olivers, Meijer, & Theeuwes, 2006; Olivers, 2009). Olivers (2009) found that changing the object in working memory from trial to trial would lead to no memory based interference in a visual search task. In Soto and Humphreys (2007), it is reported that interference in visual search from items in working memory can happen even in the absence of an explicit memory task. In Experiment 2, they presented visual or verbal primes (e.g., “red square” or picture of a red square) followed by displays of four objects. Inside three of the four objects there was a vertical line, and the fourth object had a slanted line. Among the three objects with a vertical line, one of them had the same color or shape as the visual/verbal prime. Participants were asked to quickly identify the slanted line in the display (see Figure 25). They found working memory interference effects from both conjunction (i.e., shape) and color match for both verbal and visual primes, indicating that color and conjunction properties stored in working memory interact with visual search.

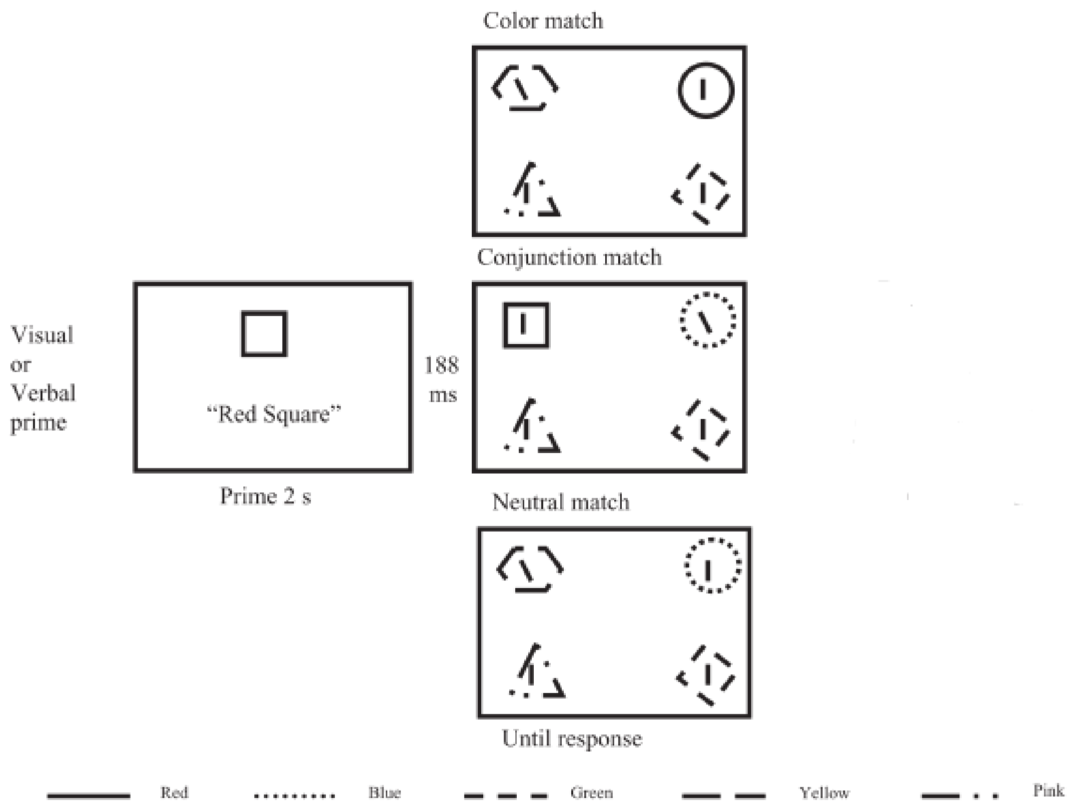


Figure 25: Example visual display from Soto & Humphreys (2007). Participants are given a visual or verbal preview of a prime and then are asked to identify the object with the slanted line inside it in a display with four objects that contain a vertical/slanted line.

Unlike Experiment 2&3, in which participants are given the target to be searched before the display appears, this experiment will explore how the addition of a working memory feature will affect the competition effect. In other words, participants will have to engage in a visual search driven by the target utterance and the color property of the target. First, participants will be given a color feature that they have to store in working memory. Afterwards, participants will be told that they will hear a target word which will appear in the display with three other words. In some cases, the target will appear in the same color as the color stored in working memory, and in other cases it will be a different color. Participants will have to make a response whenever the target is in the same color (yes) or different color (no). It is predicted that in experimental trials in which the target does appear in the primed color, participants will quickly fixate on it and show few fixations toward the competitor or distractors. In other words, the color feature stored in working memory will facilitate in the identification of the target and will lead to a smaller competition effect. On the other hand, if either the distractor or the competitor appears in the primed color, while the target is in a different color, the color will create visual search interference and one will observe enhanced competitor effects relative to the case where the target appears in the primed color.

Methods

Materials

Same target recordings from Experiment 1 were used in Experiment 4, except that the prompt before each target was "is it" rather than "click on." The prompt "is it" was spliced to each of the target word following the same procedures used in Experiment 1. The duration of "is it" was 435ms.

Participants

A total of 25 Binghamton undergraduate students took part in Experiment 4 in order to fulfil a psychology course requirement. All participants were native speakers of English and had normal or corrected-to-normal vision and no hearing impairments. Four participants were bilingual, speaking mainly English at home, with the remaining being monolingual. None of the participants reported issues with distinguishing between the colors blue, red, yellow, and green.

Procedure (see Figure 43d at the end of the document)

Eye-tracking recording parameters and setup is the same as in Experiment 1.

At the start of the trial, participants were presented with a display with a fixation cross in the center. They were asked to orient their eyes towards the fixation cross. One second after the onset of the display with the fixation cross, the display was changed to a display with a color patch (either in green, blue, yellow, or red) appearing in the center. The color of the patch changed randomly from trial to trial. Participants had a preview of the color patch for 1s+the duration of the is it prompt (435ms). 200ms before the presentation of the auditory target, the display changed to the word display. Each of the four words had a different color font that was either red, blue, yellow, or green. Participants were asked to identify if the color of the spoken target on the screen is the same as the color of the patch. Participants had to press *yes*, if the target is colored in the primed color or *no* of it is presented in a different color. For the 36 experimental trials, in 12 trials the target was in the primed color (Congruent), in other 12 the competitor was in the in the primed color (Incongruent Competitor), and for the remaining 12 trials one of the distractors was in the primed color (Incongruent Distractor). In the filler trials, the color was assigned so that in 1/3 of the trials the target appeared in the primed color. Color variable was counterbalanced, so that across participants a target would appear in either of the four colors.

Results

Fixations for 2.2% of the trials for which participants selected a word other than the target were removed (there were about 3-4 incorrect trials for each congruity condition). Average button press time for correct responses was 1543ms from the onset of the spoken target (Congruous=1198ms; Incongruous Competitor=1766ms; Incongruous Distractor=1666). An LMER model with subject and item random intercepts and by subject slope intercept for the congruity variable revealed a significant difference in RT for Congruous vs. Incongruous Competitor (estimate=567, SE=60, t=9.46, $p<.0001$), Congruous vs. Incongruous Distractor (estimate=460, SE=53, t=8.61, $p<.0001$) and Incongruous Distractor vs. Incongruous Competitor (estimate=107, SE=43, t=2.5, $p=.03$). RT for button presses indicate that participants were faster at identifying the color of the target when the color matched the color of the square. Moreover, participants were slower at identifying the color of the target when the competitor had the same color as the square rather than when the distractors matched in color with the square.

The top panels of Figure 26 plot the fixations proportions from the onset of the trial until 6000ms. Figure 26 depicts fixations for all three types of trials: experimental, filler no overlap, and filler overlap. Additionally, trials are broken down based on whether the target had the same color with the color patch (Congruous), the competitor had the same color as the patch (Incongruous Competitor), or one of the distractors was in the same color as the color patch

(Incongruous Distractor). The bars represent 95% confidence intervals. Bilingual participants performed similarly to monolingual participants, so the data used in figures and analyses includes all subjects.

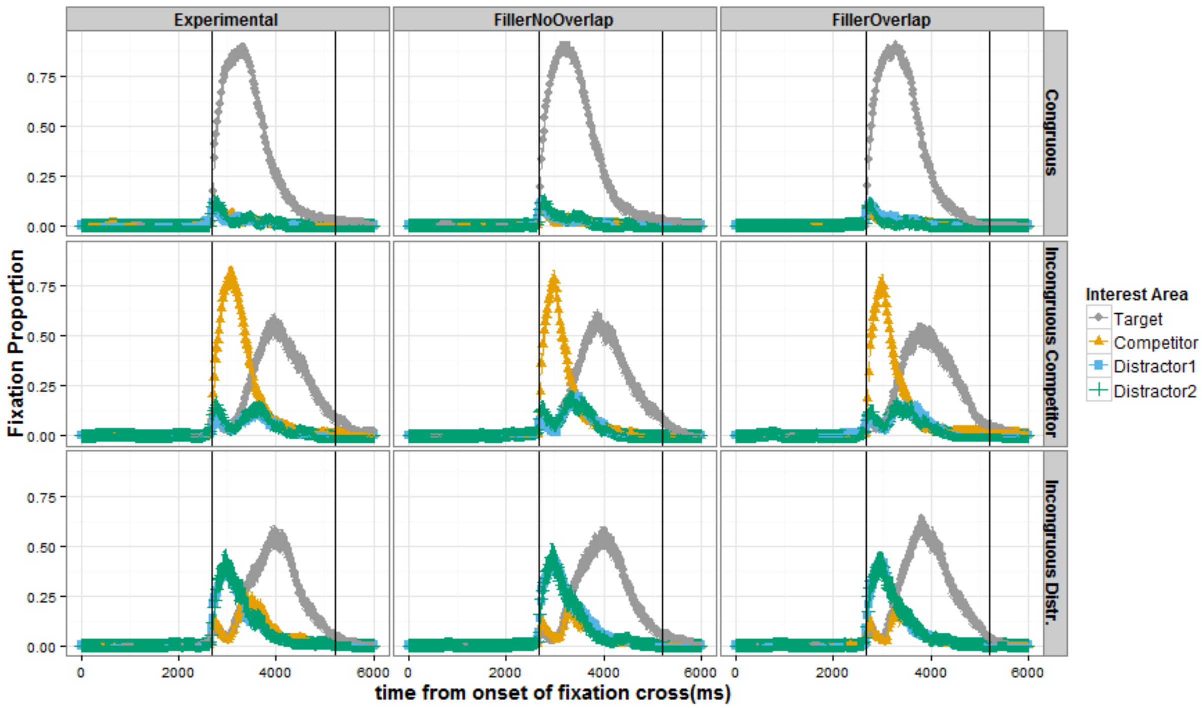


Figure 26: Fixation proportions for Experiment 4 from the start of the trial for each trial type, broken down by whether the target had the same color as the color patch (Congruous), the competitor had the same color as the patch (Incongruous Competitor), or one of the distractors had the same color as the patch (Incongruous Distr).

A zoomed-in portion of the trial between the two vertical lines in Figure 26 is presented below. The portion represents fixations 200ms from the onset of the word until 5200ms into the trial.

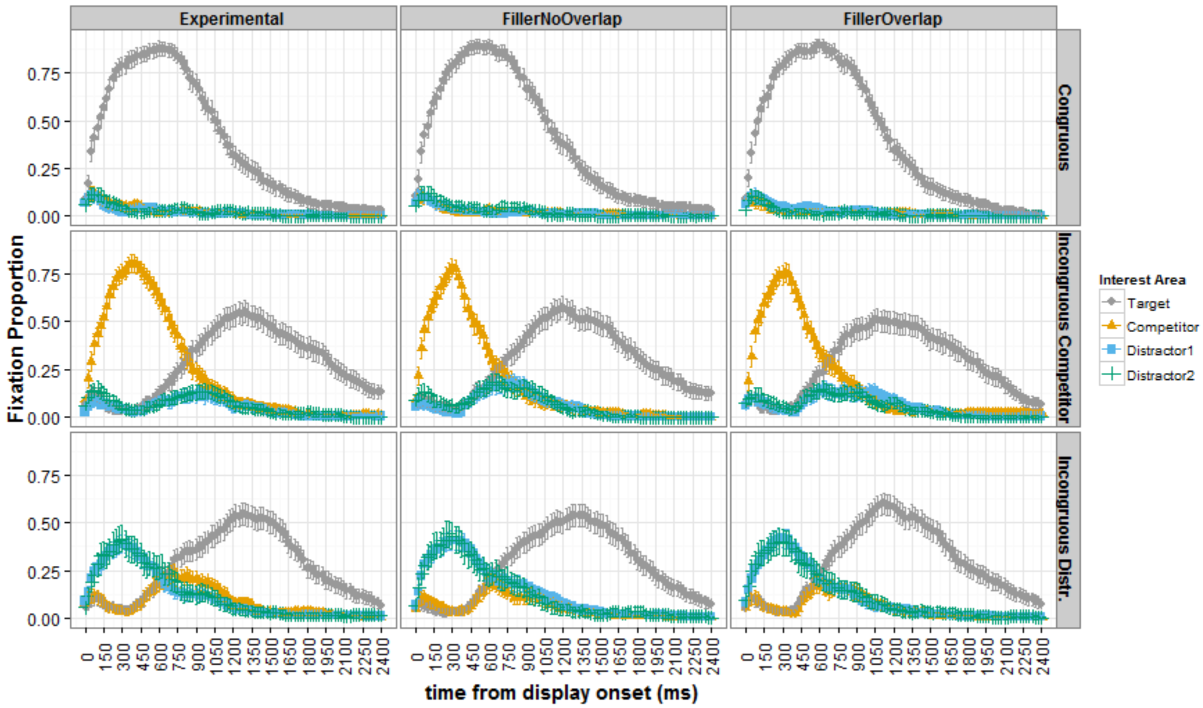


Figure 27: Fixation proportions for Experiment 4 from 200ms after target word onset until 5200ms into the trial.

Figure 27 indicates that initial fixations to the four interest areas are determined mainly by the color participants stored in working memory. If the target has the same color as the patch, participants show almost no fixations to the competitor or the two distractors. On the other hand, if the competitor or one of the distractors is presented in the same color as the color patch, then participants fixate initially more often to the competitor and the distractors, followed by more fixations to the target. In the following analyses, I fitted models comparing fixation proportions for targets and competitors across the congruity condition in experimental trials. To aid in the interpretation of which conditions were compared, the fixations in Figure 27 have been replotted. Figure 28 indicates the target fixations for the congruence conditions, followed by competitor fixations, followed by the two distractor conditions.

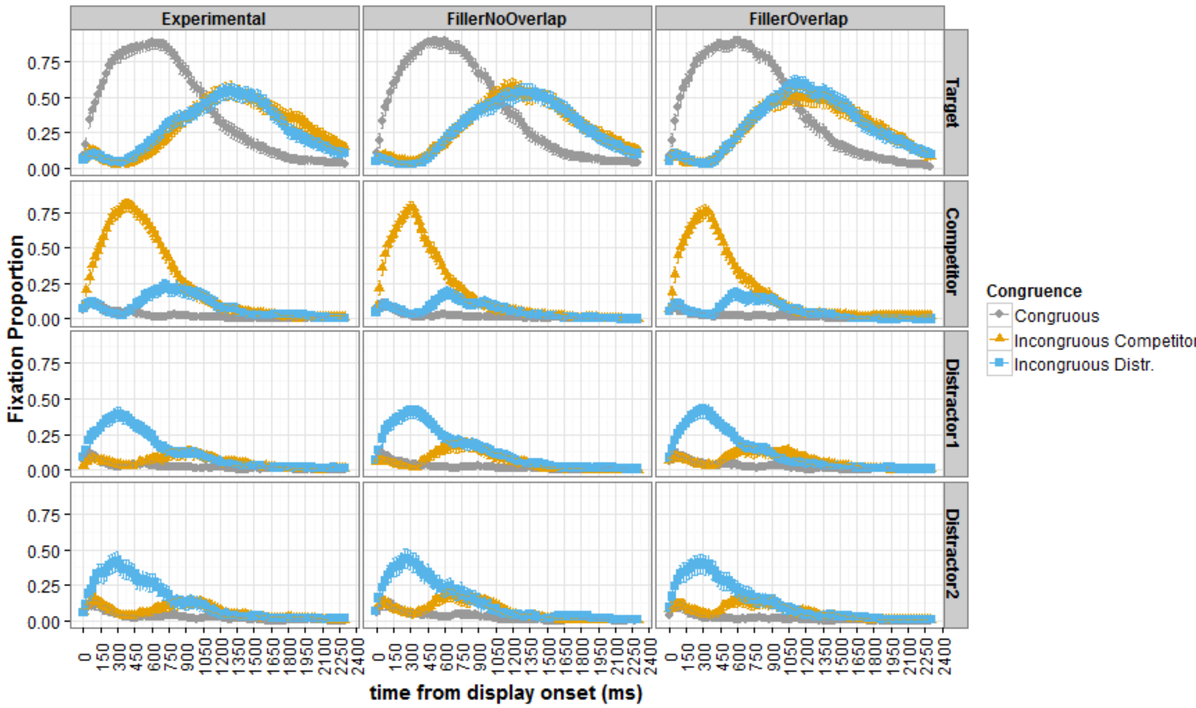


Figure 28: Target, competitor, and distractor fixations plotted based on congruence condition and trial type.

Target (congruous) vs. Target (incongruous comp.) vs. Target (incongruous distr.)

Fixations between the targets across congruity conditions are compared using GCA and random forests (RF). Table 14 depicts the structure of all the models that have been tested, the percentage variance explained by each model, and the Aikake Information Criterion (AIC). Figure 29 presents a graphical representation of the models' fit to the data.

Table 14: Models tested for Target (congruous) vs. Target (incongruous comp.) vs. Target (incongruous distr.) in Experiment 4

Model Type and Name	Model Structure	R^2	AIC
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.131	2339.2
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.467	-1132.3
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.703	-5245.2
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.834	-9318.7

Model 5 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2|Subject:Condition)	.815	-8673.9
Model 6 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2|Subject)+ (1+T1+T2+T3|Subject:Condition)	.834	-9305.9

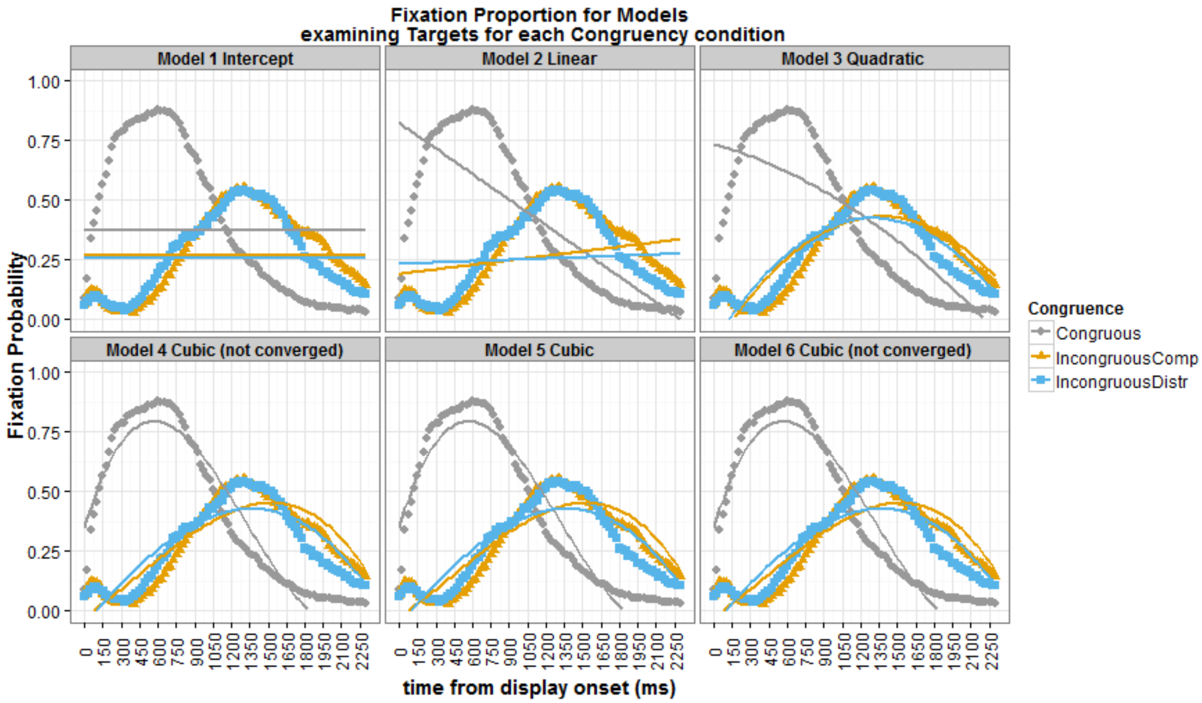


Figure 29: Predictions of the fitted models to the data in Experiment 4. The lines represent the predicted values for the models, while the shapes are the mean data for each time point.

For GCA models, although Model 6 has the best fit to the data when it comes to explained variance and AIC, the model does not converge. Model 5 was selected as the next best model. The summary of Model 5 can be found in Table 15.

Table 15: Model 5 output in Experiment 4

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.271	.019	13.94	<.0001
Congruous vs. Incon. Comp.	.102	.017	6.03	<.0001
Incon. Distr. vs. Incon. Comp.	-.014	.017	-0.83	.4081
T1	.468	.158	2.96	.0005
T2	-1.425	.161	-8.87	<.0001
T3	-.376	.083	-4.55	<.0001
T1:(Congruous vs. Incon. Comp.)	-3.131	.152	-20.55	<.0001
T1:(Incon. Distr. vs. Incon. Comp.)	-0.329	.152	-2.16	.0341
T2:(Congruous vs. Incon. Comp.)	1.004	.189	5.29	<.0001
T2:(Incon. Distr. vs. Incon. Comp.)	-0.096	.189	-0.51	.6134
T3:(Congruous vs. Incon. Comp.)	1.927	.037	52.36	<.0001
T3:(Incon. Distr. vs. Incon. Comp.)	0.224	.037	6.61	<.0001

The GCA model indicates that the fixation curves for targets in the incongruous competitor and incongruous distractor had very similar shapes. On the other hand, the performance of the congruous target relative to the incongruous competitor shows major differences across all polynomial interactions.

Competitor (congruous) vs. Competitor (incongruous comp.) vs. Competitor (incongruous distr.)

Fixations between competitors for each congruence condition are compared. First, several models were fitted to the data (see Table 16 and Figure 30) and the output for the models with the best fit (which also converged is presented).

Table 16: Models tested in Experiment 4, comparing the fixations for competitors across three congruency conditions

Model Type and Name	**Model Structure**	**R^2**	**AIC**
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.214	-4195.8
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.646	-10061.3
Model 3 (linear)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.631	-9837.9
Model 4 (linear)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.646	-10065.1
Model 5 (linear)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2|Subject:Condition)	.623	-9837.9

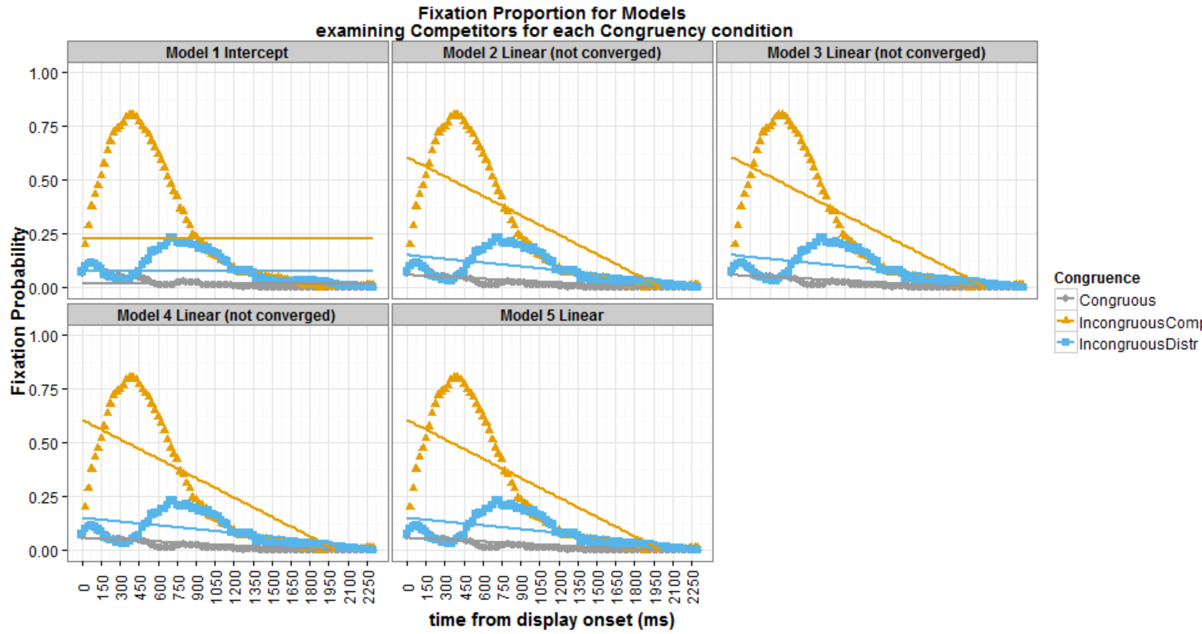


Figure 30: Predictions of the fitted models to the data in Experiment 4. The lines represent the predicted values for models, while the shapes are the mean data for each time point

For GCA models, Model 5 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 5 can be found in Table 17.

Table 17: Model 5 output in Experiment 4

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.076	.007	10.13	<.0001
Congruous vs. Incon. Comp.	-.056	.011	-5.18	<.0001
Incon. Distr. vs. Incon. Comp.	.151	.011	13.94	<.0001
T1	-.437	.025	-17.29	<.0001
T1:(Congruous vs. Incon. Comp.)	.216	.036	6.035	<.0001
T1:(Incon. Distr. vs. Incon. Comp.)	-1.800	.036	-50.35	<.0001

GCA indicates that the main driver of differences in fixation curves between competitor in the incongruous distr. condition and competitor in the incongruous comp. condition is the interaction with the linear polynomial.

Discussion

Experiment 4 indicates that competitor effects are enhanced when competitors match the color of the patch, supporting the hypothesis that non-linguistic information stored in working memory will guide visual attention away from competitors (in congruous condition) or to the competitors (in the incongruous comp. condition). The results also show that attention initially (0-

700ms range) is primarily being captured by the color property rather than by the processing of the spoken target. For instance, in both cases where the competitor or the distractors had the font color matching that of the colored square, participants ignored the target that was presented through the auditory input and looked at non-targets. The key feature of this experiment is that visual search is based on a conjunction of features (color and visual word-form) and initial saccades are driven by attentional capture of the color property. There are a few reasons why participants are using color as their first property in the conjunction search. First, color is immediately available (both because it is pre-activated in working memory and also because it is visually presented), while the spoken input is delivered incrementally. This implies that participants don't have to wait for the speech signal to start the search process, they can simply initiate the visual search based on color, and use the auditory input at a later time to decide if the current word they are fixating matches the spoken input. Second, color is a visual property, which likely makes it more salient than abstract lexical representations. Third, routine visual search seems to rely on a pre-attentive, largely parallel processing of basic visual features (such as color, motion), followed by a later stage that performs more complex processing (e.g., object recognition, reading; see Guided Search 2.0 model by Wolfe, 1994). Differences in timing how information is presented and stored in working memory and also in the type of information that is required to be accessed and stored allows for several other questions to be asked.

First, can non-visual properties, such as semantic properties also interfere with lexical competition patterns found in the VWP? For instance, instead of asking participants to confirm if the target is red, they can be asked to confirm if the target is a living or non-living thing. In other words, the conjunction search will not be based on a salient visual property such as color, but rather on a semantic property. If participants are then presented with a display such as [Target=HAMSTER, Phonological Competitor=HAMMER, Semantic Competitor: IGUANA, Distractor: PHONE], will they attend to the phonological competitor, which should get activated based on lexical activation from the spoken input, or will they attend to the semantic competitor iguana? Moreover, will a competitor be stronger if it shares its phonology and semantic category with the target (e.g., [Target=BUCKLE, Phonological Competitor=BUCKET, Distractor: IGUANA, Distractor: HAMSTER]) or weaker if it shares only the phonology ([Target=HAMSTER, Phonological Competitor=HAMMER, Distractor: BUCKET, Distractor: PHONE]). These questions are important because they will reveal to what extent working memory affects visual attention during speech processing in the VWP. Color clearly guides attention to the point where it overwhelms language-mediated eye-movement, but are non-visual representations/features capable of doing this?

Second venue for future work could explore if the contents of working memory can be used to guide visual search away from specific objects when participants are told that the contents of working memory serve no purpose in identifying the target. For instance, Experiment 4 could be modified so that instead of asking participants to identify if the font color of the target matches the color of the square patch, they could be asked to remember the color for later recall and during visual search to simply click on the target they hear through the headphones. All other aspects of the task would remain intact, including having the words presented in one of the four color fonts. In this version of the task, color serves no purpose in visual search. Findings from visual search literature indicate that observes have difficulty in ignoring the contents stored in working memory, even when they are detrimental to visual search (see review by Soto, Hodsoll, Rotshtein, & Humphreys, 2008; Soto & Humphreys, 2007; but see Woodman and Luck, 2007, who argue against automatic capture of attention from WM)[5]. Studying how pre-activated representations in working memory affect eye-movements in the VWP allows one to have a more nuanced understanding of sentence processing. For instance, if a mechanic is saying to his assistant “Give me the round thingy…actually, forget about it, I need the nails first”, will visual

[5] Note that attentional capture by working memory representations is likely intertwined with the load theory proposed by Lavine (see Experiment 3 discussion above). Distractors that match the WM representation are more likely to be looked at if perceptual load is low (i.e., there are few items in the display) and cognitive load is low (like we observed in Experiment 2).

attention when processing the spoken word *nails* be guided to a box of nails, or will there be residual activation from having just been told to look for something round, thus having more looks to a box of nuts? Currently, we have very little understanding how attentional capture due to working memory affects language-mediated eye-movements. Experiment 4 provides a solid start, by indicating that attentional capture is happening when representations involve color and when the search requires the use of WM information.

I argued in the discussion section of Experiment 2 that participants are unlikely to engage visual working memory when the visual display is constantly present, thus indicating that competition processes that are observed in VWP are due to just-in-time visual search rather than pre-activated working memory representations. I want to point out that I don't argue that working memory plays no role in visual search. By asking participants to actively store a search template in working memory, eye-movements are clearly drawn to word-forms that match the stored template (Experiments 2, 3, and 4 are a clear indication that this is the case). What I do want to point is that if the display is complex or is always present in the visual field, participants have no need to store representations in working memory.

Experiment 5

The question asked in this experiment is very simple: how do eye movements map in manipulations of set size? Despite the simplicity of this manipulation, there has been very little investigation regarding the efficiency with which visual attention in VWP is guided based on the set size of the display and a great deal of interest on the part of other researchers in finding if the interaction between language and eye movements is fast an efficient with increased complexity of the display (Allopenna et al., 1998; Ferreira et al., 2013; Huettig et al., 2012; Huettig, Olivers, & Hartsuiker, 2011b; Huettig, Rommers, et al., 2011; Ramesh K. Mishra et al., 2013). The only preliminary data concerning the effects of set size on competition are several experiments from a conference report by Sorensen and Bailey (2006), which has received only four citations since its publication[6][7][8]. Sorensen and Bailey (2006) extended the VWP manipulation reported in Huettig and Altmann (2005), who examined semantic competitor effects. They presented 2x2, 3x3, and 4x4 picture displays and found that semantic competitor effects for displays with more than 4 objects are not closely time-locked to the spoken utterance, and in some cases they even found absence of semantic competition effects if the preview duration was short (e.g., 1s of preview). Absence of competitor effect for large set sizes with short preview times indicate that participants are not able to store the visual information in short term memory that would later enable the mapping of language processing to the visual array. In essence, participants would have to engage in a visual search task once the information from the visual display decayed. Experiment 5 aims to explore the influence of set size by presenting 4-word or 6-word display in which the display will contain both the target and a competitor. The expectation is that in 6-word display condition the competitor effect will be smaller and delayed (i.e., the link between eye-movements and the spoken utterance will not time-locked) relative to the 4-word condition, indicating that VWP effects are contingent on display set size.

Methods

Materials

Same recordings used in Experiment 1 were used in Experiment 5. Experiment 5 had two types of displays: 4-word displays and 6-word displays. The 4-word displays were the same as the ones used in Experiments 1-4. In the 6-word displays there were two more distractors present in addition to the words in 4-word displays (see last 2 columns in Appendix A). For experimental trials with 6-word displays, stimuli were normed so that all six words in a display do not differ

[6] As of 4/9/2015, using google scholar

[7] Questions about the visual complexity of the display have been also examined by Ferriera et al., (2013) in the context of syntactic ambiguity rather than competition. They found that complex visual displays did not constrain the interpretations of garden path sentences. In other words, the typical garden-path effects are found only when the display contains few pictures.

[8] UPDATE: when revising this document, after the data has been collected, analyzed, and written up, I found another article by Andersson, Ferreira, & Henderson (2011) in which they had very complex displays with dozens of objects. I discuss this article in the discussion section of this experiment.

among each other in terms of a number of lexical/semantic dimensions (all pair-wise comparisons have p>.2; see descriptive statistics in Table 1 above).

Participants

A total of 42 Binghamton undergraduate students took part in Experiment 5 in order to fulfill a psychology course requirement. There were 21 participants in Group 1 and 21 in Group 2. All participants were native speakers of English and had normal or corrected-to-normal vision and no hearing impairments. There were four bilingual participants in Group 1 and three in Group 2. The rest of the subjects reported being monolingual.

Procedure (see Figure 43d at the end of the document)

Eye-tracking recording parameters and setup is the same as in Experiment 1, except that the interest area size was decreased to 200x210 pixels.

At the start of each trial, participants were presented with a fixation cross in the center of the screen for one second. Participants were asked to orient their eyes towards the cross when it appears. After the presentation of the fixation cross, a display with four/six printed words (6 words for subjects in Group 1 and 4 words for subjects in Group 2) was presented. Participants were instructed to look at the words immediately when they appear on the screen and were told that they will have to hear later a target word and will have to select it among the words in the display. The onset of the auditory input was presented 2000ms after the presentation of the display with the words. The duration of the preview has been selected to allow participants sufficient time to look at all the locations of the words in the display. After the participants heard the target, they had to click on the corresponding word in order to complete the trial. In this experiment, participants are given a preview of 2882ms before they get the presentation of the spoken target.

Results

Fixations for 0.3% of the trials in Group 1 and .01% of the trial in Group 2 for which participants selected a word other than the target were removed. Average mouse click time was

1294ms for Group 1 and 1315ms for Group 2, as measured from the onset of the spoken target.

Figure 31 plots the fixations for Group 1 and Figure 32 plots the fixations for Group 2.

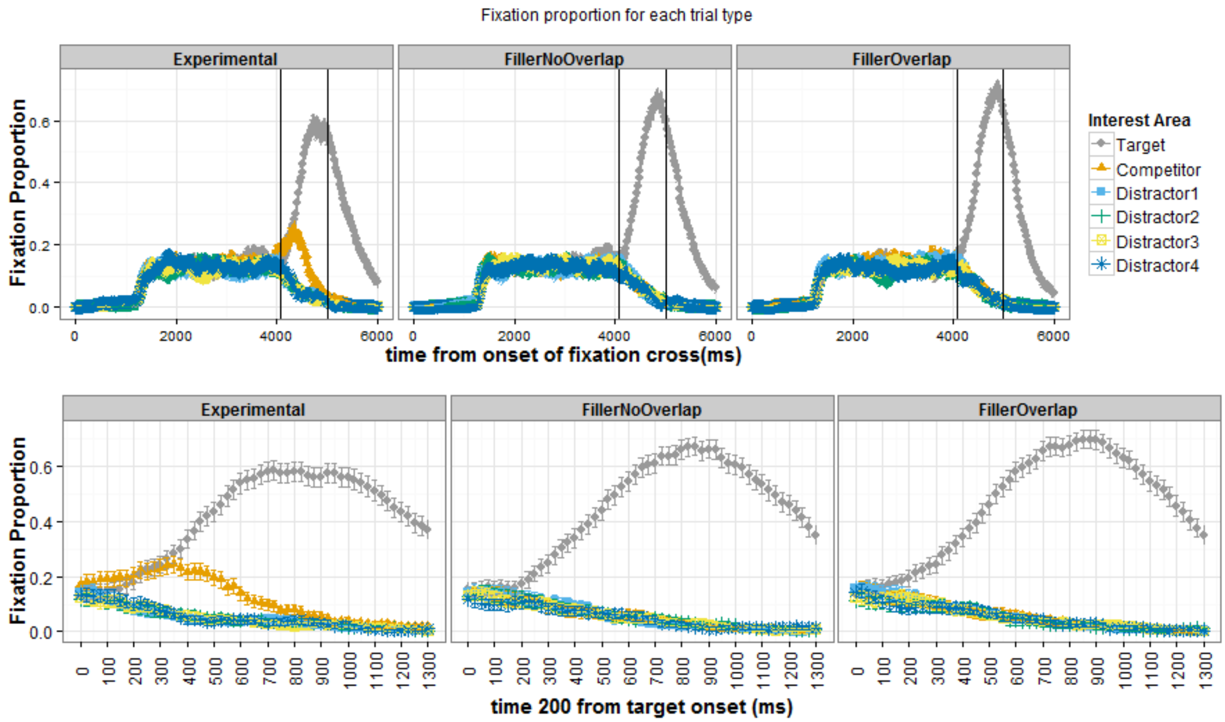


Figure 31: Fixation proportions for Experiment 5 from the start of the trial (top three panels) until 6000ms for Group 1. The bottom panels represent the fixation proportions 200ms from the onset of the target word until 5000ms.

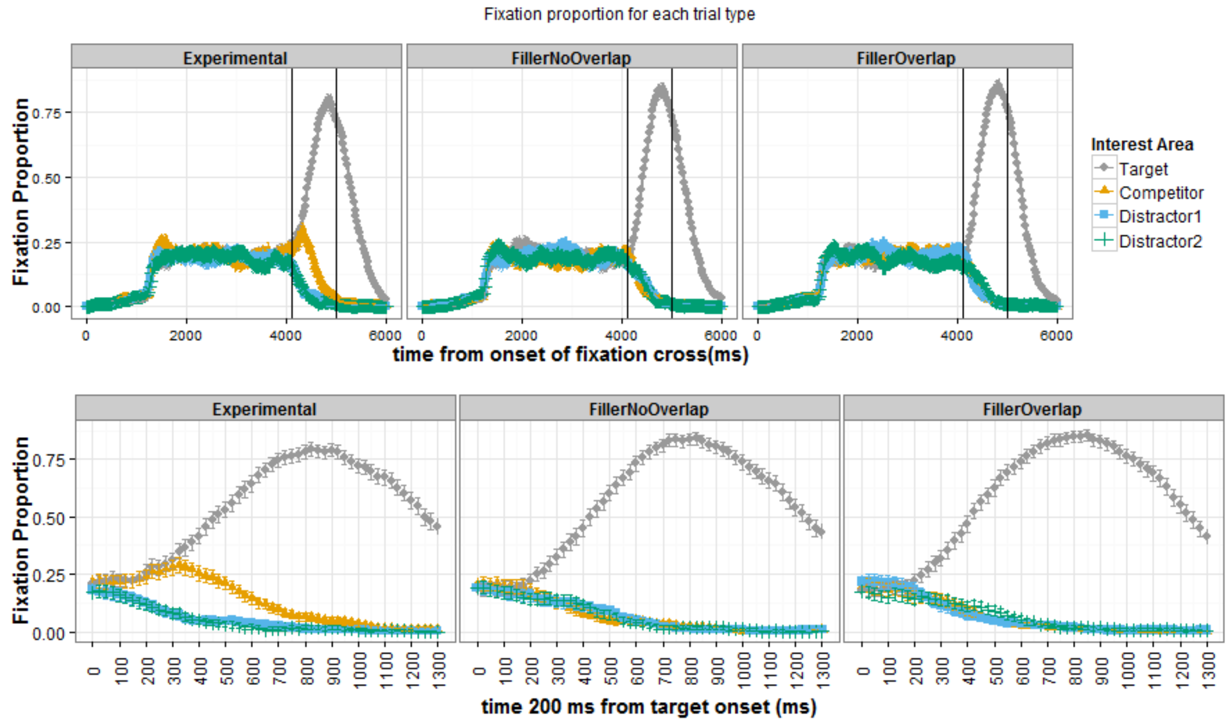


Figure 32: Fixation proportions for Experiment 5 from the start of the trial (top three panels) until 6000ms for Group 2. The bottom panels represent the fixation proportions 100ms from the onset of the target word until 5000ms.

Both figures indicate the now familiar competitor effect.

Target vs. Competitor Group 1

Fixations between the target and competitors are compared using GCA and random forests (RF). Table 18 depicts the structure of all the models that have been tested, the percentage variance explained by each model, and the Aikake Information Criterion (AIC). Figure 33 presents a graphical representation of the models' fit to the data.

Table 18: Models tested for Target vs. Competitor comparison in Experiment 5, Group 1

Model Type and Name	Model Structure	R^2	AIC
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.516	-1585
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.887	-3623.4
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.945	-4513.6
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.967	-5102.8

Model 5 (quartic)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+ (1+T1+T2+T3+T4|Subject)+ (1+T1+T2+T3+T4|Subject:Condition)	.978	-5533.4

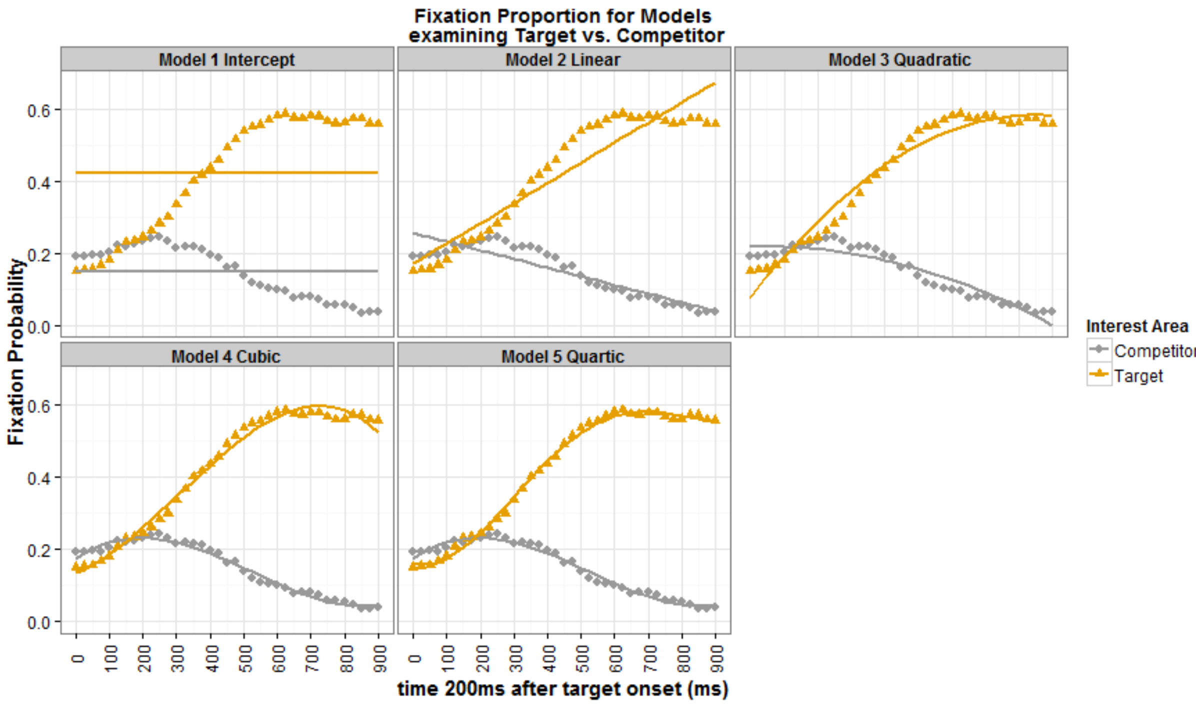


Figure 33: Predictions of the fitted models to the data in Experiment 5, Group 1 (Target vs. Competitor). The lines represent the predicted values for the models, while the shapes are the mean data for each time point.

For GCA models, Model 5 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 5 can be found in Table 19.

Table 19: Model 5 output for Target vs. Competitor in Experiment 5,Group 1

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.147	.011	13.46	<.0001
Condition (Target vs. Competitor)	.274	.013	20.55	<.0001
T1	-.391	.058	-6.69	<.0001
T2	-.117	.045	-2.56	.0144
T3	.119	.027	4.39	<.001
T4	.001	.026	0.05	.9640
T1:Condition	1.298	.075	17.33	<.0001
T2:Condition	-.159	.062	-2.55	.0154
T3:Condition	-.271	.037	-7.22	<.0001
T4:Condition	.078	.036	2.18	.0356

GCA shows the same pattern of results that have been observed in Experiments 1-3. The linear interaction seems to be the main driver between fixation curves for target vs. competitor.

The bootstrapped cluster-based permutation analysis indicates that the competitor and the target diverge at the 275ms mark (see Figure 34). The cluster of summed t-tests (Sum=415; see cluster 2) is far outside of a 10000 sample distribution of permutation tests (Mean=-.02; 95% CI [-86,87], p=0).

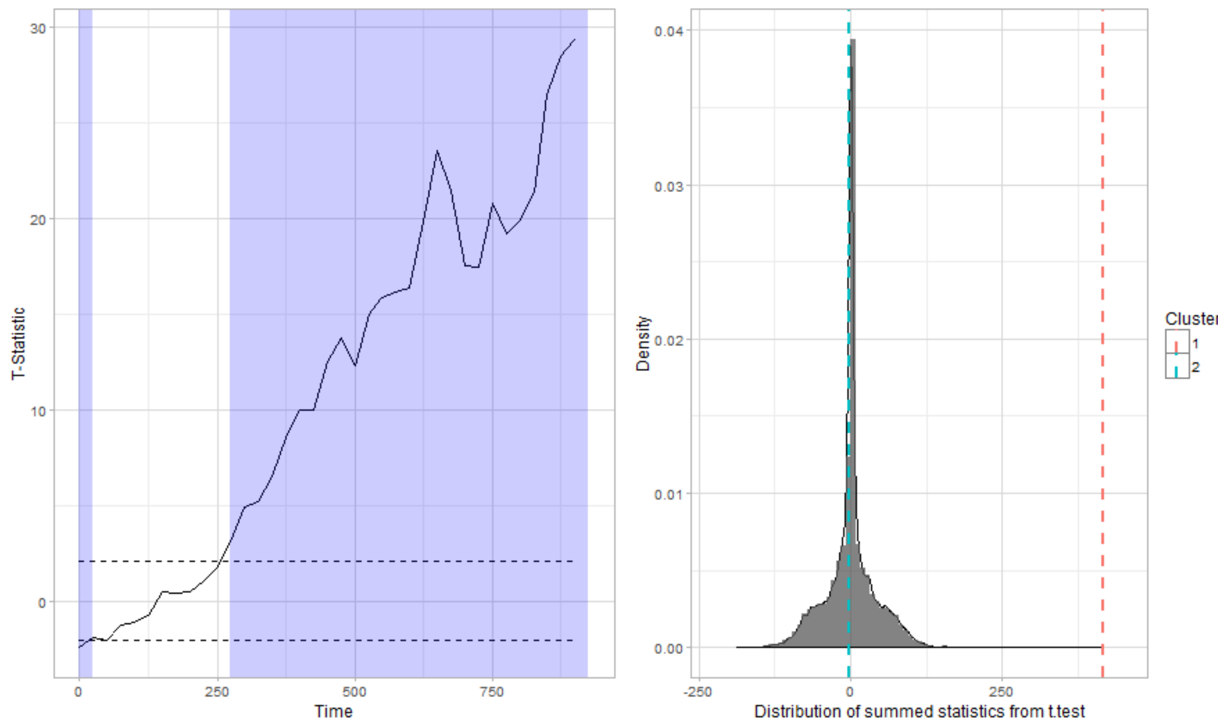


Figure 34: Permutation-cluster analysis outcome for Experiment 5, Group 1 (Target vs. Competitor). The left panel represent the t-statistic for every 25ms time-bin. The time-cluster 2 from 275ms to 925ms is significant. The right panel indicates the distribution of t-sums from 10000 simulations. This distribution indicates that the time-cluster 2 has a very small probability of occurring by chance.

Target vs. Competitor Group 2

Fixations between the target and competitors are compared using GCA and random forests (RF). Table 20 depicts the structure of all the models that have been tested, the percentage variance explained by each model, and the Aikake Information Criterion (AIC). Figure 35 presents a graphical representation of the models' fit to the data.

Table 20: Models tested for Target vs. Competitor comparison in Experiment 5, Group 2

Model Type and Name	Model Structure	R^2	AIC
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.495	-930.1
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.923	-3921.1
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.948	-4395.9
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.979	-5743.1
Model 5 (quartic)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+ (1+T1+T2+T3+T4|Subject)+ (1+T1+T2+T3+T4|Subject:Condition)	.984	-5960.1

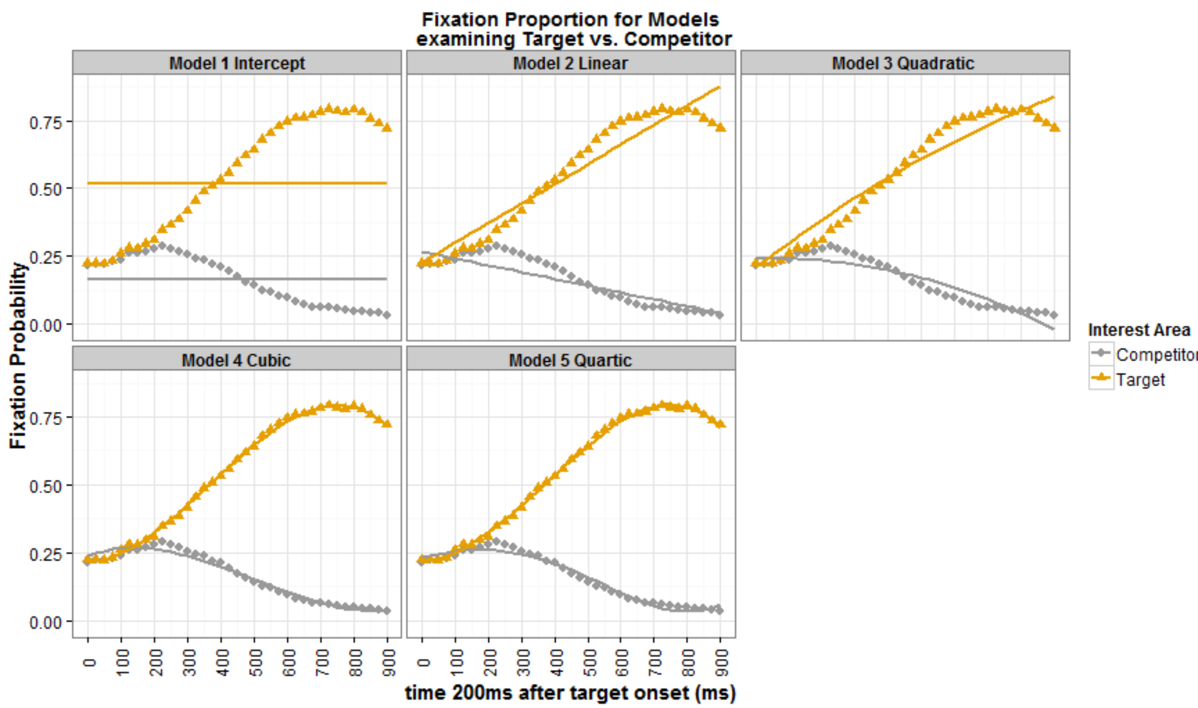


Figure 35: Predictions of the fitted models to the data in Experiment 5, Group 2 (Target vs. Competitor). The lines represent the predicted values for the models, while the shapes are the mean data for each time point.

For GCA models, Model 5 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 5 can be found in Table 21.

Table 21: Model 6 output for Target vs. Competitor in Experiment 5,G2

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.167	.009	17.44	<.0001
Condition (Target vs. Competitor)	.350	.013	27.41	<.0001
T1	-.473	.063	-7.47	<.0001
T2	-.184	.045	-4.11	.0002
T3	.158	.028	5.58	<.0001
T4	.057	.021	2.67	.0109
T1:Condition	1.84	.072	25.46	<.0001
T2:Condition	0.06	.062	0.95	.3507
T3:Condition	-.493	.039	-12.50	<.0001
T4:Condition	-.075	.030	-2.48	.0177

The bootstrapped cluster-based permutation analysis indicates that the competitor and the target diverge at the 225ms mark (see Figure 36). The cluster of summed t-tests (Sum=538; see cluster 1) is far outside of a 10000 sample distribution of permutation tests (Mean=.29; 95% CI [-99,101], p=0).

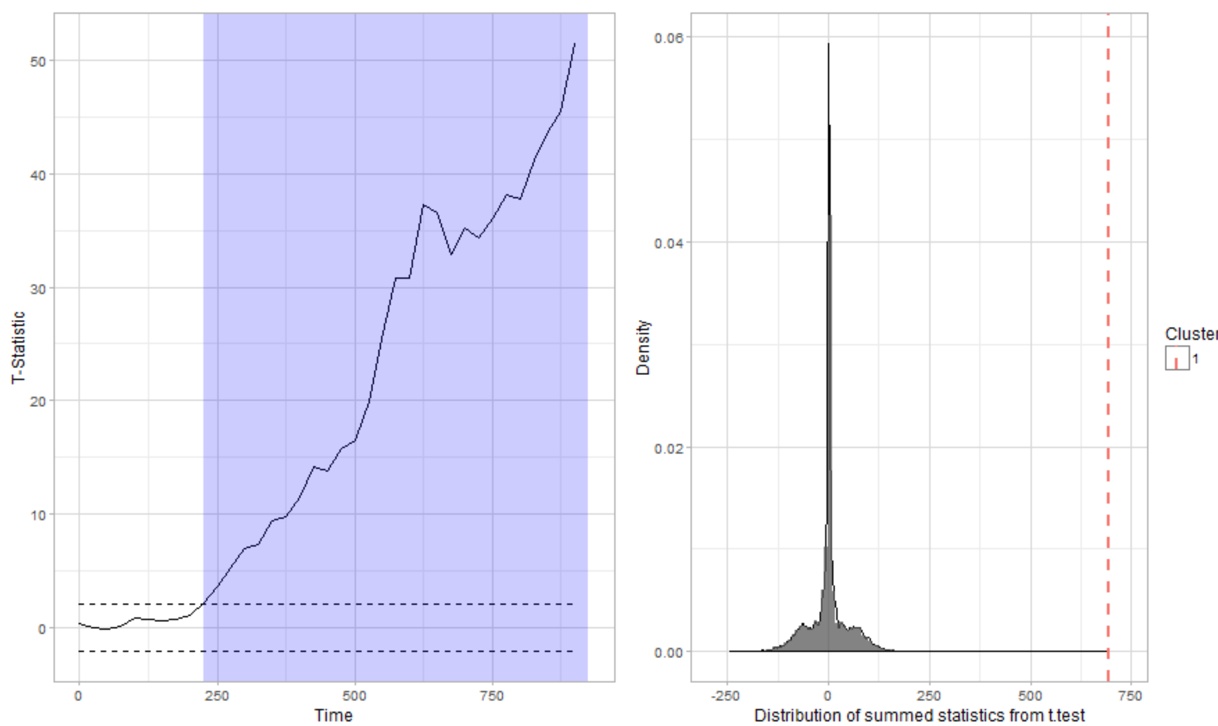


Figure 36: Permutation-cluster analysis outcome for Experiment 5, Group 2 (Target vs. Competitor). The left panel represent the t-statistic for every 25ms time-bin. The time-cluster from 225ms to 925ms is significant. The right panel indicates the distribution of t-sums from 10000 simulations. This distribution indicates that the time-cluster has a very small probability of occurring by chance.

Competitor vs. Distractors Group 1

Fixations between competitor and the average of the two distractors are compared using the same approach as above. First, several models were fitted to the data (see Table 22 and Figure 37) and the output for the models with the best fit (which also converged is presented).

Table 22: Models tested in Experiment 5, comparing the fixations between Competitor and the average of the four distractors.

Model Type and Name	Model Structure	R^2	AIC
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.382	-3965.7
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.778	-5378.3
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.857	-5475.1
Model 4 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1 |Subject)+ (1|Subject:Condition)	.783	-5611.2
Model 5 (quartic)	Fixation Proportion ~ Condition*(T1+T2)+ (1 |Subject)+ (1+T1|Subject:Condition)	.811	-5874.7

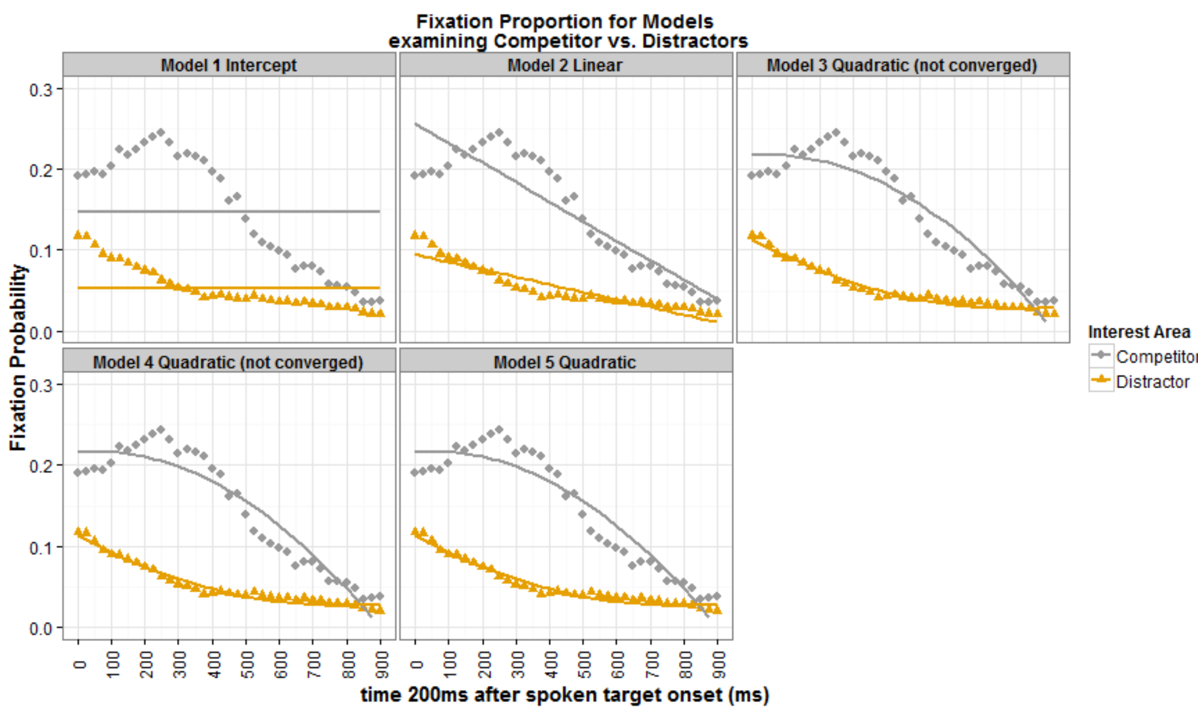


Figure 37: Predictions of the fitted models to the data in Experiment 5, Group 1 (Competitor vs. Distractors). The lines represent the predicted values for models, while the shapes are the mean data for each time point.

For GCA models, Model 5 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 5 can be found in Table 23.

Table 23: Model 5output for Competitor vs. Distractors in Experiment 1, Group 1

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.148	.005	31.64	<.0001
Condition (Competitor vs. Distractor)	-.095	.007	-14.43	<.0001
T1	-.391	.026	-15.24	<.0001
T2	-.117	.008	-14.44	<.0001
T1:Condition	.239	.036	6.61	<.0001
T2:Condition	.171	.011	14.97	<.0001

The bootstrapped cluster-based permutation analysis indicates that the competitor and the distractors diverge at the 0ms mark (see Figure 38). The cluster of summed t-tests (Sum=218; see cluster 1) is far outside of a 10000 sample distribution of permutation tests (Mean=.24; 95% CI [-62,65], p=0).

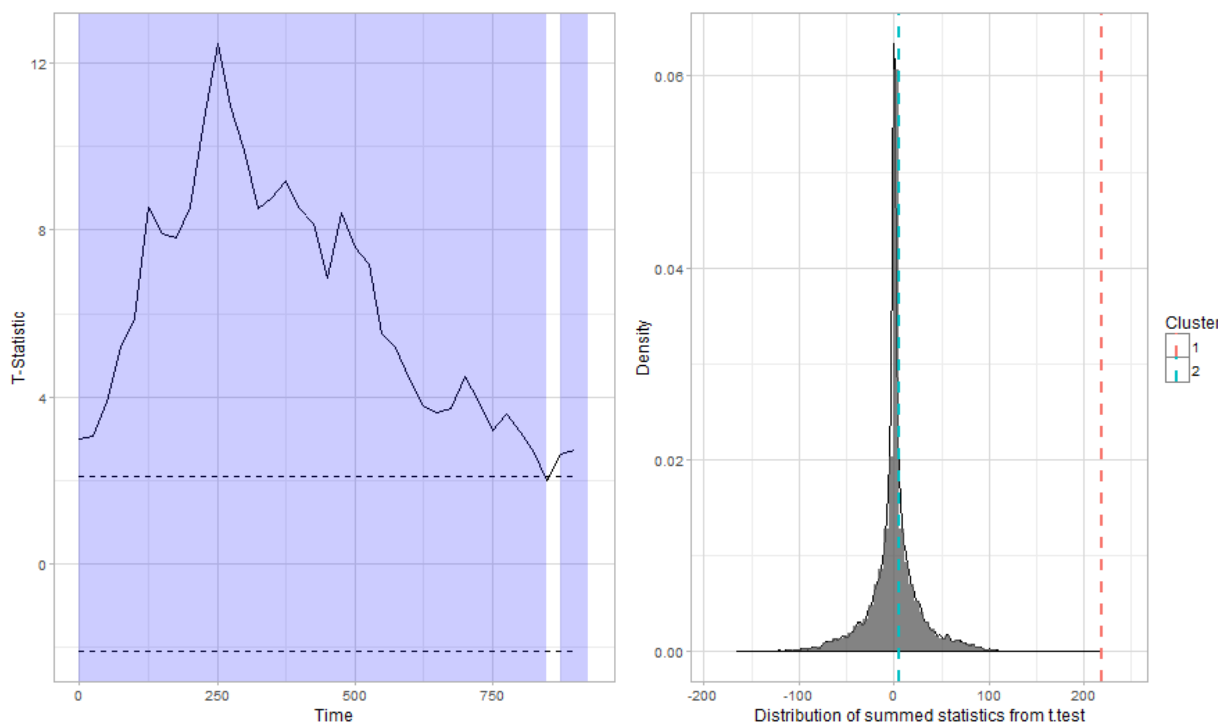


Figure 38: Permutation-cluster analysis outcome for Experiment 5, Group 1 (Competitor vs. Distractors). The left panel represent the t-statistic for every 25ms time-bin. The time-cluster from 0ms to 850ms is significant. The right panel indicates the distribution of t-sums from 10000 simulations. This distribution indicates that the time-cluster has a very small probability of occurring by chance. The plot on the left seems to suggest that competition is starting before 0 (i.e., before 200ms after spoken target onset). A reanalysis of the data in which I extended the time-period from target onset until 1100ms after target onset suggests that the competitor and the

distractors diverge at the 175ms. In other words, in the plot on the left, at about time -25ms before 0, the competitor and the distractor start to diverge.

Competitor vs. Distractors Group 2

Fixations between competitor and the average of the two distractors are compared using the same approach as above. First, several models were fitted to the data (see Table 24 and Figure 39) and the output for the models with the best fit (which also converged is presented).

Table 24: Models tested in Experiment 5, comparing the fixations between Competitor and the average of the two distractors

Model Type and Name	**Model Structure**	**R^2**	**AIC**
Model 1 (Intercept)	Fixation Proportion ~ Condition+ (1|Subject)+(1|Subject:Condition)	.323	-3469.5
Model 2 (linear)	Fixation Proportion ~ Condition*T1+ (1+T1|Subject)+ (1+T1|Subject:Condition)	.772	-5146.3
Model 3 (quadratic)	Fixation Proportion ~ Condition*(T1+T2)+ (1+T1+T2|Subject)+ (1+T1+T2|Subject:Condition)	.874	-5963.7
Model 4 (cubic)	Fixation Proportion ~ Condition*(T1+T2+T3)+ (1+T1+T2+T3|Subject)+ (1+T1+T2+T3|Subject:Condition)	.923	-6631.6
Model 5 (quartic)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+ (1+T1+T2+T3+T4|Subject)+ (1+T1+T2+T3+T4|Subject:Condition)	.942	-6960.1

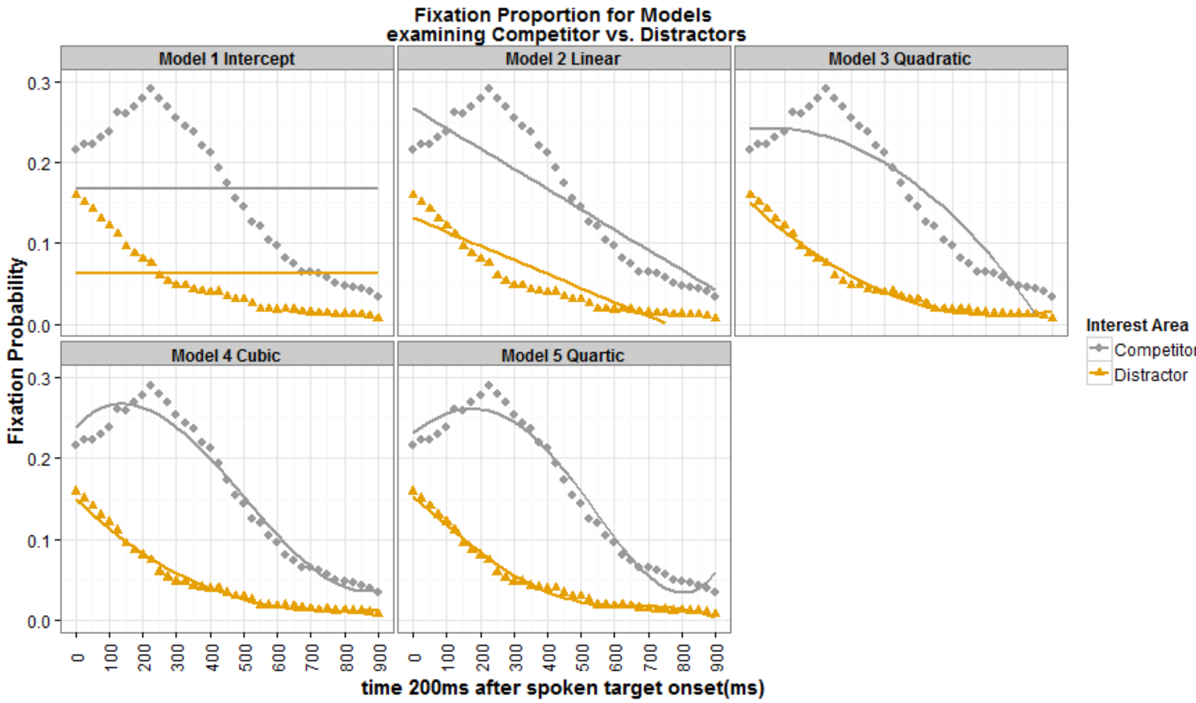


Figure 39: Predictions of the fitted models to the data in Experiment 5, Group 2 (Competitor vs. Distractors). The lines represent the predicted values for models, while the shapes are the mean data for each time point

For GCA models, Model 5 had the best fit to the data, both in terms of its explained variance and AIC value. The output of Model 5 can be found in Table 25.

Table 25: Model 5 output for Competitor vs. Distractors in Experiment 5,Group 2

Predictor	**β**	**S.E.**	**t**	**p**
(Intercept)	.167	.006	26.515	<.0001
Condition (Target vs. Competitor)	-.105	.008	-13.07	<.0001
T1	-.473	.040	-11.78	<.0001
T2	-.184	.032	-5.82	<.0001
T3	.158	.022	7.23	<.0001
T4	.057	.018	3.19	.0028
T1:Condition	.145	.056	2.57	.0142
T2:Condition	.312	.044	7.04	<.0001
T3:Condition	-.168	.030	-5.60	<.0001
T4:Condition	-.083	.025	-3.34	.0018

The bootstrapped cluster-based permutation analysis indicates that the competitor and the distractors diverge at the 25ms mark (see Figure 40). The cluster of summed t-tests (Sum=217;

see cluster 1) is far outside of a 10000 sample distribution of permutation tests (Mean=.03; 95% CI [-56,57], p=0).

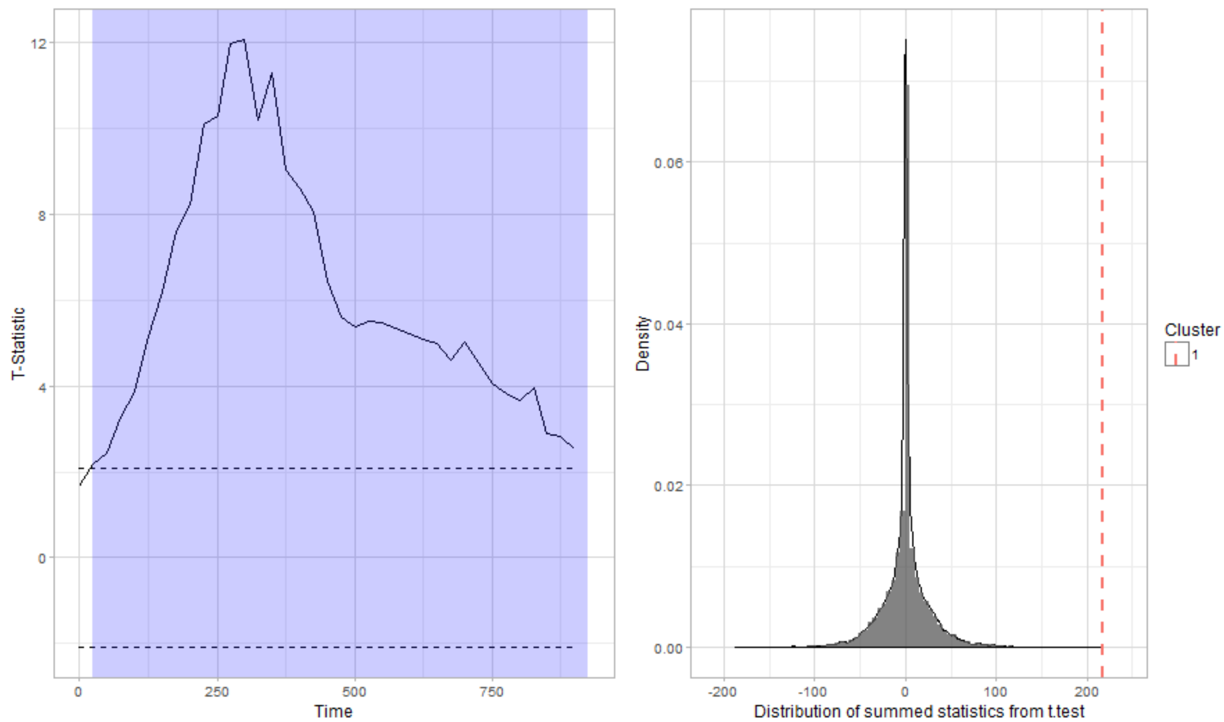


Figure 40: Permutation-cluster analysis outcome for Experiment 5, Group 2 (Competitor vs. Distractors). The left panel represent the t-statistic for every 25ms time-bin. The time-cluster from 25ms to 925ms is significant. The right panel indicates the distribution of t-sums from 10000 simulations. This distribution indicates that the time-cluster has a very small probability of occurring by chance.

Discussion

Experiment 5 Group 2 provides a replication of Experiment 1, by indicating that during the preview period (2.9s), participants look indiscriminately at all the words in the display. Experiment 5 also indicates that the competitor effect is present in both 4-word and 6-word displays. Based on work reported by Sorensen and Bailey (2006), it was expected that there would be a smaller competitor effect and that the effect will not be time-locked to the spoken target. This prediction was not confirmed: competition was observed in both types of displays. These results indicate that set size does not impact competition effects.

The cluster permutation analyses indicate that in Experiment 5, participants initiate saccades much faster relative to what is observed in Experiment 1. For instance, the competitor effects for Experiment 5, Group 2 start 75ms sooner relative to the pattern of results observed in Experiment 1. Why are participants faster in disengaging from distractor words in Experiment 5? The only methodological difference between Experiment 5 (Group 2) and Experiment 1 is that in Experiment 1 participants had a preview which was shorter by a second. One possibility is that a longer preview period allows the participants to store in working memory the visual word forms and their locations. If this is the case, participants in Experiment 5 used the contents of working memory to inhibit saccades to distractor words, thus leading to an early competitor effect[9]. Nonetheless, one issue with the early competitor effect is that it shows up for both Group 1 and Group 2. Group 1 had a more complex display and the visual search literature (Horowitz & Wolfe, 1998; S J Luck & Vogel, 1997; McCarley et al., 2003) and several studies from the VWP literature indicate that complex displays should lead to delayed time-locking for language-driven eye-movements rather than earlier time-locking. For displays with larger set sizes, the competition effects should show much later than for displays with smaller sizes. For instance, for displays that contain dozens of objects, the close linking between speech as visual objects falls behind and participants don't fixate on all the objects (Andersson et al., 2011; see example display in Figure 41).

[9] It is important to note that this pattern of results has been discovered only after the experiment has been completed and the results analyzed. It was not part of the initial prediction proposed for this dissertation.

Figure 41: An example of a display used by Andersson, Ferreira, and Henderson (2011). Participants would hear a sentence that refers to multiple objects in the display. An example sentence would be, “I love going to garage sales. I like [the sailboat that’s old and dust-covered], [the plane], [the sombrero], and [the surprisingly mint uniform]. However, I’ll skip buying anything.”

Work by Hintz and Huettig (2015) also found that in more complex displays, there is delayed language-mediated mapping, even though participants were provided with a 3s preview (see Figure 42). The displays used by Hintz and Huettig (2015) had the same items sets as the ones used in Huettig and McQueen (2007) except that the objects were embedded in the context of several characters interacting with them. Hintz and Huettig (2007) found semantic and shape competitor effect only about 1.5 seconds after the acoustic onset of the target word, when participants are given a look-and-listen task. Moreover, they found that there was no phonological competitor effect. This contrasts with the findings provided by Huettig and McQueen (2007), where the competitor effects were found much sooner (about 200-300ms after spoken target onset).

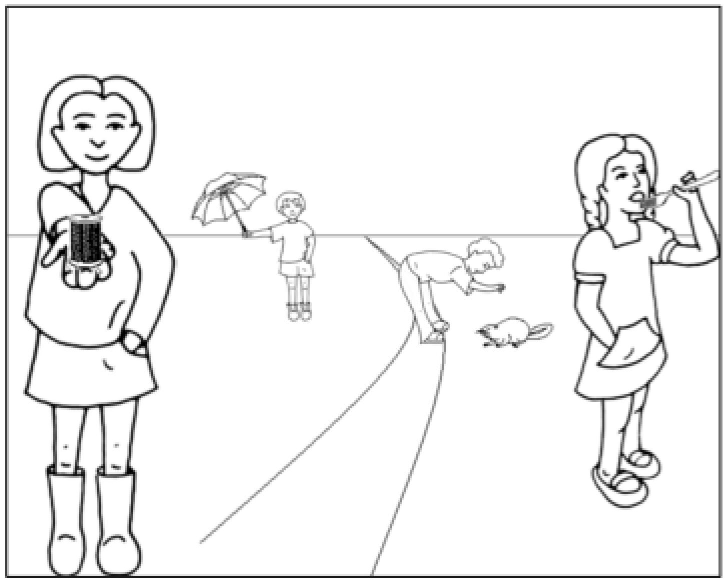


Figure 42: Example display from Hintz and Huettig (2015). Participants would hear the target beaker in Dutch. The display would contain the objects klos (bobbin, shape competitor), bever (beaver, phonological competitor), vork (fork, shape competitor), paraplu (umbrella; distractor)

One reason why the early competitor effect shows up is because the preview period (2.9s) was long enough to allow participants to store the words in the display in working memory. This explanation nonetheless goes counter to research indicating that visual or verbal working memory capacity is about 4 items (Cowan, 2010) and that participants prefer to not store information in working memory if the display is always present in their vision. The duration of the preview has not received enough examination in the VWP literature, so at this point it is difficult to say how preview affects the type of representations that are accessed before the spoken input is presented. Work by Huettig & McQueen (2007), indicates that phonological representations are not accessed if the preview for visual displays with objects is short (less than 250ms). On the other hand, if the display contains words, the short duration of the preview does not seem to activate visual representation (e.g., activating the visual property roundness when processing the word BALL). It would be interesting if future work could examine the preview durations under more systematic conditions (e.g., having conditions with no preview, 200ms preview, 500ms preview, 1000ms preview, 3000ms preview, and 5000ms preview).

Another reason why the competitor effect emerges early is because the display with larger item sets induce a high perceptual load, which according to load theory leads to less attentional capture on the part of distractors (Forster & Lavie, 2008; Lavie et al., 2004; Lavie & Tsal, 1994). Unfortunately, it is difficult to know if this is the case in this experiment. Most of the work in load theory have examined the distractibility of distractors using very simple stimuli (e.g., shapes or letters), so it is known how strongly words engage attentional capture during bottom-up visual processing.

General Discussion

The aim of this dissertation was to examine some task parameters in the Visual World Paradigm (VWP) in the hope of answering these questions:

1) What is the influence of preview? Can competition effects be observed through implicit naming during the preview period?

2) Can competition effects during visual search be guided by linguistic representations stored in working memory? If yes, can those representations be disrupted though a cognitive load manipulation?

3) Are competition effects affected by non-linguistic representations stored in working memory?

4) Does display size matter? Are fixations time-locked to the spoken input in displays with larger number of words?

Preview and implicit naming

As outlined through several studies in the introduction (Huettig & McQueen, 2007; Meyer et al., 2007; Meyer & Damian, 2007; Zelinsky & Murphy, 2000), the preview in the VWP allows participants to engage in implicit naming of the objects/words before the presentation of the spoken target (see also evidence for activation of phonological information during visual word identification in Lukatela, Eaton, Sabadini, & Turvey, 2004; Pollatsek, Lesch, Morris, &

Rayner, 1992). This implicit naming process could lead to competition effects being observed between two phonological onset competitors without the presence of any auditory input.

Experiments 1 and 5 in this volume indicate that concerns due to implicit naming in the VWP can be put to rest when it comes to examining how visual attention couples with spoken word processing. In both experiments, participants were given a significant preview (1.9s in Experiment 1 and 2.9s in Experiment 5) that allowed them to read and process all words in the display. In experimental trials, despite having two words in the display that are similar to each other, participants did not show a higher number of fixations to those words relative to the distractor words. Instead, throughout the preview period, all words in the display had equivalent fixation proportions. Only during the presentation of the spoken target, the visual target and competitor were fixated with a higher probability than the two (Experiment 1) or four distractors (Experiment 5).

These results indicate that in the word-display version of the VWP, when the spoken target is presented in a neutral context, the preview does not matter. This doesn't mean that the preview doesn't matter in all circumstances. As pointed out by Huettig and McQueen (2007), a short preview duration does not allow for the activation of the phonological codes during object identification. Instead, visual similarity between objects is the main driver of competition effects. Ferreira et al. (2013) point out that the preview plays an important role in letting listeners build expectations about possible utterances they are going to hear. A longer preview was observed to lead to larger garden-path effects in the VWP because participants could guess what the spoken utterance might be (see Experiment 3 in Ferreira et al., 2013). In the current experiments, a guessing strategy would not work, because of the presence of the filler overlap trials, which eliminates the bias to look at two competitor words because one of them is likely to be the target. Hence, in an experimental context that does not bias's participants' expectations about which word they have to select, fixations seem to be guided mainly by the processing of the spoken input. It is important to note that the results of Experiment 1 and 5 should not be used to argue that participants are not engaging in implicit naming, but rather that, even if participants implicitly name the words in the display, implicit naming does not guide competition effects—the processing of the spoken target guides those effects.

Linguistic representations in working memory

Experiment 2 indicates that linguistic representations stored in working memory show similar competition effects that are observed when the spoken target is presented concomitantly with the display or after a preview period. Working memory was defined as the active storage of a stimulus representation after it has been presented and then removed. The results of experiment provide a theoretical replication of some previous studies in visual search literature that found top-down guidance of attention from visual stimuli stored in working memory (Soto et al., 2005; Soto & Humphreys, 2007; Meyer & Telling, 2007; nontheless see also Woodman & Luck, 2007 and Olivers, 2009, who found failure to find working memory effects across a variety of manipulations). The presence of competitor effect in Experiment 2, does not invalidate the use of the VWP to study speech processing. Rather, these results indicate that visual search is influenced by phonological representation of the spoken target.

One explanation for why competitor effects are observed in Experiment 2 is because looks to the target and competitor are contingent on making a response. Relating the visual information and the spoken information in working memory requires participants to look at relevant parts of the display that would help them make a correct response. In other words, the task required participants to overtly move the mouse to the location of the target, hence fixations were observed because of these overt movements. Salverda, Brown, & Tanenhaus (2011) suggest that in in majority of the visual search literature, task-relevant goals account for a large number of saccades. In other words, the number of saccades and the timing of fixations are influenced by the cognitive goals of the task rather than purely by processing that is ascribed to look-and-listen tasks. Salverda, Brown, & Tanenhaus (2011) have consolidated this view by reformulating the linking hypothesis from one that was purely focused on comprehension to one that is based on goals and instructions. According to the goal-based linking hypothesis, "the view assumes that the instruction in a task-based visual world experiment triggers a basic language-vision routine, for instance mapping a word onto a potential referent for purposes of visually guided reaching" (p. 178). This explanation is intuitive and appealing on a few fronts. First, it is unlikely that for every spoken word that a listener hears during a dialogue, a phonological representation is created and stored in working memory together with related competitors. Second, several studies examining visual processing in natural environments (Ballard et al., 1995; Land, Mennie, & Rusted, 1999; Sprague et al., 2007), indicate that visual attention in an environment is generally oriented towards locations that are relevant in performing a particular task (e.g., looking at a kettle's handle before pouring tea). The goal-based view of the linking hypothesis also suggests an interesting manipulation that could be performed in future studies. If task-goals are the main driver of saccades, will competition effects be observed in circumstances when participants are asked to look-and-listen, without engaging in any overt task? The goal-based view indicates that competition effects are likely to be diminished because participants are not explicitly told to store any information in working memory because there is no requirement for them to look at the words that they previously heard.

Experiment 3 indicated that the addition of a cognitive load lead to some disruption in storing the phonological representation of the spoken word in working memory and to smaller competitor effect when participants had to search for the visual target. This is the only known manipulation that I'm aware which indicates that representations stored in working memory in the VWP or visual search can be successfully disrupted by solving difficult arithmetic problems.

Overall, these two experiments indicate that representations stored in working memory are enough to guide competition effects. The manipulation Experiment 2 opens the venue to other experiments that could manipulate how conflicting linguistic information affects the ongoing competition observed in typical VWP studies. Moreover, the fact that fixations can be guided by working memory representations alone, can make the VWP an attractive task to use for nonpsycholinguistics experiments, such as semantic priming, categorization, memory recognition and recall tasks.

Non-linguistic representations in working memory

While Experiment 2 indicates that a linguistic working memory representation can lead to competitor effects that are typically observed in the VWP, the results of Experiment 4 indicate that non-linguistic representations can also influence the guidance of selective attention during the search of the visual target. Previously, the effects of non-linguistic representations in working memory have been examined only in the context of the visual search task (Downing & Dodds, 2004; Downing, 2000; Soto et al., 2005; Soto & Humphreys, 2007). Experiment 4 combines the two paradigms, visual search and visual world, together, and indicates that perceptual features (such as color) can actively disrupt or enhance competition effects.

In Experiment 4, if the color of the target was congruent to the color patch, competition effects were essentially diminished, with the target receiving most of the fixations. On the other hand, if the font color of the competitor or one of the distractors was the same as the color of the patch, then targets had fewer fixations initially, followed by more fixations as participants realized they were not looking at the correct word. The visual search process in this experiment was guided primarily by the color stored in working memory rather than the ongoing processing of the spoken word. In other words, the color feature was the most salient property that guided eye-movements rather than the auditory input.

Display size

The goal of Experiment 5 was to examine if visual attention during speech processing is influenced by display size. Sorensen and Bailey (2006) report that larger display sizes lead to diminished semantic competitor effects for object displays. Moreover, the competitor effects not only get smaller (or non-existent for displays with short preview) as the display size increases, but they are also not time-locked to the spoken utterance. Their experiments indicate that integration of the spoken information together with the visual information breaks apart when one deals with more complex displays, suggesting that some of the competitor effects observed in the VWP are task dependent.

The results of Experiment 5 indicate that display size does not play role in guiding the phonological/orthographic competitor effects. If participants are given a long enough preview (i.e., 2.9s) to allow them to identify each word on the screen, then competitor effects are closely time-locked to the spoken target regardless of display size.

One criticism that could be brought for the methodology of this experiment is that the display size manipulation was not strong enough, hence the reason why similar competitor effects were observed for both 4-word and 6-word displays. This criticism does not receive support in the literature. As pointed by Luck and Vogel (2013), the visual working memory across a variety of visual search manipulations diminishes right after the set size of objects during preview is larger than 4, indicating that mapping of the spoken referent should be more difficult in the 6-word display than in the 4-word display.

One possibility for why no decoupling was observed in Experiment 5 between the speech input and visual attention is due to task instructions. In Sorensen and Bailey (2006), participants were not given an overt task, but rather had to listen to a sentence and look at a display. Altmann and Kamide (1999) previously reported that eye-movements are more rapidly and more frequently deployed in experiments in which participants had to perform an overt task (e.g., move an object) rather than look-and-listen tasks, which involve no overt responses.

Future work

In this section, I have included a list of proposed experiments/ideas that could be implemented to further explore several aspects of the VWP. In particular, this section aims to consolidate some of the questions that have been put forth in the discussions portions of each experiment. The proposed manipulations in this section are by no means exhaustive.

Question 1: Do listeners extract visual representations during preview and store them in working memory?

Working memory model proposed by Huettig et al. (2011) presupposes that participants can extract different types of representations from the visual display and index those representations to locations in working memory. Nonetheless, the memory-less accounts of visual search and the hand-eye coordination work indicates that observers prefer to engage in just-in-time visual processing, if they expect that the display will be always available. Therefore, it is not clear if participants automatically are storing in working memory semantic, phonological, or visual information during the preview period. One way to resolve this issue is to acknowledge that if representations are stored automatically in working memory, then visual search should be more rapid (see discussion in Exp2.) because participants can use working memory representations to inhibit fixations to locations that mismatch the auditory target. There are a few ways one could test if different representations are pre-activated and stored in working memory.

- Present a display with 4 items for a few seconds. After the presentation of this display, provide a display with one item in the center. The item in the center shares some of its representations with one of the four items in the previous display. Example: Have a display with four objects [Coiled rope, Umbrella, Envelope, Phone] followed by a picture of an uncoiled snake (snake and rope should be related due to their coilness [Dahan & Tanenhaus, 2005]). Participants can be asked to confirm if the second object is animate or not. In this example, will semantic categorization be faster because participants pre-activated some information about snake? Note that this task would be something similar to a priming experiment, except that there are four primes and only one target.
- The other option would be to ask participant to not only look at the display during the preview period, but to actively engage in memorizing the objects in the display to the best of their ability. One could do that by presenting the display with the objects/words for 2 seconds. Each object/word will be located in one quadrant of the display. Following this initial preview, the objects will appear at the bottom of the screen. Participants will then

have to move those objects to their respective locations on the screen. This will ensure that they have memorized the locations of the objects/words. After that, the trial will follow the structure of the trial used in the VWP. The expectation here would be that participants should show earlier divergence between competitor and distractor fixation curves because they remember those objects and tied them to their specific locations. They should be able to inhibit the locations of mismatching visual referents once they process the onset of the auditory target.

Question 2: Does preview length matter?

The influence of preview length has not been systematically explored in the VWP. Work by Huettig and McQueen (2007) indicates that preview does affect the type of representations that are being accessed. For instance, if visual objects are presented for only 250ms, then participants are not given enough time to extract the label of the objects through implicit naming, and subsequent processing of the auditory input shows no phonological competition. Preview duration matters because it can affect whether participants decide to memorize the objects and their locations in working memory. As pointed above, storing the pre-activated representation in working memory can lead to differences in how early competitor effects are being observed.

Question 3: Does perceptual load matter?

Load theory (Forster & Lavie, 2008; Lavie et al., 2004; Lavie & Tsal, 1994; Weast & Neiman, 2010) suggests that distractors are more salient when the perceptual load is small. In other words, having smaller number of items in the display would lead to more fixations towards distractors relative to having displays with larger item sizes. Load theory has been used extensively to explain attentional capture in selective attention paradigms. Nonetheless, the implications of this theory for VWP have been completely unexamined. Future work could assess the distractibility of distractors by manipulating the set size of the items in the display and by eliminating the preview period (so as not to confound the effects of pre-activated working memory representations). The active component of the load theory deals with cognitive load rather than perceptual load. Cognitive load influences the ability of the observer to disengage from the distractor, with larger load leading to more prolonged looks to the distractors. In Experiment 3, I tested how cognitive load affects the mapping of pre-activated linguistic knowledge. One could easily adapt Experiment 3 so that it resembles more closely the VWP and thus study how load affects concomitant processing of visual + auditory information. For instance, one could present the math problem for 2s. After that, the display changes to the display with four objects and the spoken target is presented. In this scenario, will participants show more fixations towards distractors, as predicted by load theory?

Conclusions

The work in this dissertation indicates that visual attention can be guided by an interplay of spoken word processing, visual processing, and working memory, potentially opening the use of the VWP for a variety of non-psycholinguistic tasks, such as semantic priming, categorization, and memory recognition. Experiment 1 indicates that in a typical VWP setup, eye-movements are mediated by the processing of the spoken input. Experiment 2 shows that competitor-like effects can be observed during visual search when the auditory target is present before the visual display is presented to participants. Experiment 3 indicates that a cognitive load makes irrelevant distractors more salient and leads to no competitor effect. Experiment 4 shows that a visual working memory representation can lead to strong attentional capture, allowing participants to use spoken information as a secondary decision input (does the word match the color rather than does the color match the word?). Experiment 5 indicates that display size does not delay time-locking of the spoken input to visual processing.

References

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*(38), 419–439. doi:10.1006/jmla.1997.2558

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The “blank screen paradigm.” *Cognition*, *93*, 79–87. doi:10.1016/j.cognition.2004.02.005

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264. doi:10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502–518. doi:10.1016/j.jml.2006.12.004

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55–71. doi:10.1016/j.cognition.2008.12.005

Anderson, S. E., Chiu, E., Huette, S., & Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica*, *137*(2), 181–9. doi:10.1016/j.actpsy.2010.09.008

Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you’re saying: the integration of complex speech and scenes during language comprehension. *Acta Psychologica*, *137*(2), 208–16. doi:10.1016/j.actpsy.2011.01.007

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory Representations in Natural Tasks. *Journal of Cognitive Neuroscience*, *7*, 66–80. doi:10.1162/jocn.1995.7.1.66

Bartolotti, J., Daniel, N. L., & Marian, V. (2013). Spoken Words Activate Cross-Linguistic Orthographic Competitors in the Absence of Phonological Overlap. *CogSci 2013 Annual Meeting*, 1827–1832.

Bates D, Maechler M, Bolker B and Walker S (2015). _lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1-8, <URL: http://CRAN.R-project.org/package=lme4>.

Beck, V. M., Hollingworth, a., & Luck, S. J. (2012). Simultaneous Control of Attention by Multiple Working Memory Representations. *Psychological Science*, *23*, 887–898. doi:10.1177/0956797612439068

Becker, M. W., & Pashler, H. (2002). Volatile visual representations: failing to detect changes in recently processed information. *Psychonomic Bulletin & Review*, *9*(4), 744–50. doi:10.3758/BF03196330

Biau, G. (2010). Analysis of a Random Forests Model. *The Journal of Machine Learning Research*, *13*, 1063–1095. Retrieved from http://arxiv.org/abs/1005.0208

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, *27*(April 2015), 539–571. doi:10.1080/01690965.2011.555268

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–90. doi:10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*. doi:10.3758/s13428-013-0403-5

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), 687–696. doi:10.1037/0278-7393.30.3.687

Chiu, E. M., & Spivey, M. J. (2014). Timing of speech and display affects the linguistic mediation of visual search. *Perception*, *43*, 527–548. doi:10.1068/p7593

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

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Curr Dir Psychol Sci*, *19*(1), 51–57. doi:10.1177/0963721409359277

Cowan, N. (2013). The magical. *Journal of Chemical Information and Modeling*, *53*, 1689–1699. doi:10.1017/CBO9781107415324.004

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cognitive Psychology*, *42*, 317–367. doi:10.1006/cogp.2001.0750

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459. doi:10.3758/BF03193787

Dink, J. W., & Ferguson, B. F. (2015). *eyetrackingR*: An R Library for Eye-tracking Data Analysis. Retrieved from http://www.eyetrackingr.com

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science, 11,* 467–473. doi:10.1111/1467-9280.00290

Downing, P. E., & Dodds, C. M. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*(5), 689–704. doi:10.1080/13506280344000446

Durda, K., Buchanan, L. (2006). *WordMine2* [Online] Available: http://web2.uwindsor.ca/wordmine

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, *69*(3), 165–182. doi:10.1016/j.jml.2013.06.001

Forster, S., & Lavie, N. (2008). Failures to ignore entirely irrelevant distractors: the role of load. *Journal of Experimental Psychology*. *Applied*, *14*(1), 73–83. doi:10.1037/1076-898X.14.1.73

Gauvin, H. S., Hartsuiker, R. J., & Huettig, F. (2013). Speech monitoring and phonologically-mediated eye gaze in language perception and production: a comparison using printed word eye-tracking. *Frontiers in Human Neuroscience, 7*(December), 818. doi:10.3389/fnhum.2013.00818

Görges, F., Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2013). Activation of phonological competitors in visual search. *Acta Psychologica*, *143*(2), 168–175. doi:10.1016/j.actpsy.2013.03.006

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. doi:10.1016/j.tics.2005.02.009

Hintz, F., & Huettig, F. (2015). The complexity of the visual environment modulates language-mediated eye gaze. In R. K. Mishra, K. Srinivasan, R., & F. Huettig (Eds.), *Attention and Vision in Language Processing* (pp. 39–55). Brlin: Springer. doi:10.1007/978-81-322-2443-3_3

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394,* 575–577. doi:10.1038/29068

Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(2), 423–442. doi:10.1037/0096-1523.32.2.423

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*, 23–32. doi:10.1016/j.cognition.2004.10.003

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness, (771335669). doi:10.1080/13506280601130875

Huettig, F., & Hartsuiker, R. J. (2008). When you name the pizza you look at the coin and the bread: eye movements reveal semantic activation during word production. *Memory & Cognition*, *36*(2), 341–360. doi:10.3758/MC.36.2.341

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(December 2015), 80–93. doi:10.1080/23273798.2015.1047459

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. doi:10.1016/j.jml.2007.02.001

Huettig, F., & McQueen, J. M. (2011). The nature of the visual environment induces implicit biases during language-mediated visual search. *Memory & Cognition*, *39*(6), 1068–84. doi:10.3758/s13421-011-0086-z

Huettig, F., Mishra, R. K., & Olivers, C. N. L. (2012). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology*, *3*(January), 1–11. doi:10.3389/fpsyg.2011.00394

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011a). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*, 138–150. doi:10.1016/j.actpsy.2010.07.013

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011b). Looking, language, and memory: bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*(2), 138–50. doi:10.1016/j.actpsy.2010.07.013

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. doi:10.1016/j.actpsy.2010.11.003

Kalénine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1274–1295. doi:10.1037/a0027626

Kuhn, M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang and Can Candan. (2016). caret: Classification and Regression Training. R package version 6.0-64. https://CRAN.R-project.org/package=caret

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cognitive Science*, *30*, 481–529. doi:10.1207/s15516709cog0000_65

Knoeferle, P., & Crocker, M. W. (2007a). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*(0), 519–543. doi:10.1016/j.jml.2007.01.003

Knoeferle, P., & Crocker, M. W. (2007b). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*, 519–543. doi:10.1016/j.jml.2007.01.003

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–90. doi:10.3758/s13428-012-0210-4

Kuznetsova, A.,Brockhoff P. B., & Christensen, R. (2015). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-29. http://CRAN.R-project.org/package=lmerTest

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(October 2015), 1311–1328. doi:10.1068/p2935

Lavie, N., Hirst, a, de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General; Journal of Experimental Psychology: General*, *133*(3), 339. doi:10.1037/0096-3445.133.3.339

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197. doi:10.3758/BF03213897

Lee, C., Middleton, E., Mirman, D., Kalénine, S., & Buxbaum, L. J. (2013). Incidental and context-responsive activation of structure- and function-based action features during object identification. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(1), 257–70. doi:10.1037/a0027533

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(1996), 279–281. doi:10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. doi:10.1016/j.tics.2013.06.006

Lukatela, G., Eaton, T., Sabadini, L., & Turvey, M. T. (2004). Vowel duration affects visual word identification: evidence that the mediating phonology is phonetically informed. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(1), 151–62. doi:10.1037/0096-1523.30.1.151

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(February 2016), 177–190. doi:10.1016/j.jneumeth.2007.03.024

McCarley, J. S., Wang, R. F., Kramer, A. F., Irwin, D. E., & Peterson, M. S. (2003). How much memory does oculomotor search have? *Psychological Science*, *14*, 422–426. doi:10.1111/1467-9280.01457

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology (2006)*, *60*(February 2015), 661–671. doi:10.1080/17470210601183890

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, *14*(4), 710–716. doi:10.3758/BF03196826

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, *35*(3), 494–503. doi:10.3758/BF03193289

Meyer, A. S., Ouellet, M., & Häcker, C. (2008). Parallel processing of objects in a naming task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(4), 982–987. doi:10.1037/0278-7393.34.4.982

Mirman, D., Dixon, J. a, & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. doi:10.1016/j.jml.2007.11.006

Mirman, D., & Graziano, K. M. (2012). Individual differences in the strength of taxonomic versus thematic relations. *Journal of Experimental Psychology: General*, *141*(4), 601–609. doi:10.1037/a0026451

Mishra, R. K., Olivers, C. N. L., & Huettig, F. (2013). Spoken language and the decision to move the eyes: To what extent are language-mediated eye movements automatic? *Progress in Brain Research*, *202*, 135–149. doi:10.1016/B978-0-444-62604-2.00008-3

Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(5), 1275–1291. doi:10.1037/a0013896

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(5), 1243–1265. doi:10.1037/0096-1523.32.5.1243

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 148–162. doi:10.1037/0096-1523.18.1.148

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rensink, R. a. (2000). The Dynamic Representation of Scenes. *Visual Cognition*, *7*, 17–42. doi:10.1080/135062800394667

Rensink, R. a. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277. doi:10.1146/annurev.psych.53.100901.135125

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General*, *123*(4), 374–393. doi:10.1037/0096-3445.123.4.374

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object Shape and Orientation Do Not Routinely Influence Performance During Language Processing. *Psychological Science*, (September). doi:10.1177/0956797613490746

Sajin, S. M., & Connine, C. M. (2014). Semantic richness: The role of semantic features in processing spoken words. *Journal of Memory and Language, 70*, 13–35. doi:10.1016/j.jml.2013.09.006

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. doi:10.1016/j.actpsy.2010.09.010

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(5), 1108–1117. doi:10.1037/a0019901

Sandri, M., & Zuccolotto, P. (2006). Variable selection using random forests. In Z. Sergio, C. Andrea, R. Marco, & V. Maurizio (Eds.), *Data Analysis, Classification and the Forward Search* (pp. 263–270). Springer. doi:10.1007/3-540-35978-8

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493. doi:10.1016/j.tics.2003.09.010

Sorensen, D. W., & Bailey, K. G. D. (2006). The world is too much: Effects of array size on the link between language comprehension and eye movements. *Visual Cognition*, *15*(1984), 112–115. doi:10.1080/13506280600975486

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(2), 248–261. doi:10.1037/0096-1523.31.2.248

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, *12*(August), 342–348. doi:10.1016/j.tics.2008.05.007

Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(3), 730–737. doi:10.1037/0096-1523.33.3.730

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235–241. doi:10.1007/s004260100059

Sprague, N., Ballard, D., & Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Transactions on Applied Perception*, *4*, 11–es. doi:10.1145/1265957.1265960

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580. doi:10.1023/A:1026464108329

Tanenhaus, M. K., Spivey-knowlton, M. J., Eberhard, K. M., Sedivy, J. C., Series, N., & Jun, N. (1995). Comprehension Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, *268*(5217), 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. *Handbook of Psycholinguistics*. Retrieved from http://www.international.ucla.edu/languages/nhlrc/2010summer/readings/tanenhaus_trueswell.pdf

Tobin, S., & Cho, P. W. (2010). Effects of anticipatory coarticulation on lexical access. *... Mtgs. Cognitive Sci*, 2200–2205. Retrieved from http://palm.mindmodeling.org/cogsci2010/papers/0517/paper0517.pdf

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–207. doi:10.3758/s13428-012-0314-x

Weast, R. a., & Neiman, N. G. (2010). The Effect of Cognitive Load and Meaning on Selective Attention. *COGSCI 2010*, 1477–1482. Retrieved from https://mindmodeling.org/cogsci2010/papers/0395/paper0395.pdf

Wickham, H. & Francois, R. (2015). dplyr: A Grammar of Data Manipulation. R package version 0.4.2. http://CRAN.R-project.org/package=dplyr

Wickham, H. (2014). tidyr: Easily Tidy Data with spread() and gather() Functions.. R package version 0.2.0. http://CRAN.R-project.org/package=tidyr

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. doi:10.3758/BF03200774

Woodman, G. F., & Luck, S. J. (2007a). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology. Human Perception and Performance*, *33*(2), 363–377. doi:10.1037/0096-1523.33.2.363

Woodman, G. F., & Luck, S. J. (2007b). Doe the contents of working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 363–377. doi:10.1037/0096-1523.33.2.363

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–9. doi:10.3758/PBR.15.5.971

Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2011). Function follows form: activation of shape and function features during object identification. *Journal of Experimental Psychology. General*, *140*(3), 348–63. doi:10.1037/a0022840

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), 1–14. doi:10.1037/0278-7393.32.1.1

Xu, R.H. (2003) Measuring explained variation in linear mixed effects models. Statistics in Medicine, 22, 3527–354

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: an effect of object name length on eye movements. *Psychological Science : A Journal of the American Psychological Society / APS*, *11*(2), 125–131. doi:10.1111/1467-9280.00227

Appendix A

		4-word displays in Exp. 1-4				Additional distractors for 6-word displays in Exp. 5	
	Trial Type	Target	Competitor	Distractor	Distractor	Distractor	Distractor
1	Experimental	buckle	bucket	diesel	kicker	mantle	office
2	Experimental	alarm	alone	olive	diner	merry	honey
3	Experimental	cloth	cloud	lunch	stair	broom	drill
4	Experimental	token	topaz	bunny	vomit	label	mango
5	Experimental	postal	poster	bomber	absorb	tissue	monkey
6	Experimental	hammer	hangar	tavern	mobile	faucet	ginger
7	Experimental	gospel	golfer	tribal	accent	mirror	income
8	Experimental	barrel	banner	ferret	liquid	coffee	kitten
9	Experimental	creek	cream	plane	stain	flare	thorn
10	Experimental	friend	fridge	basket	string	groove	throat
11	Experimental	breast	breath	shield	drawer	fierce	pledge
12	Experimental	amazed	amused	healer	sultan	pillow	gravel
13	Experimental	milk	mint	gold	farm	crab	hand
14	Experimental	band	bank	mist	salt	lard	pact
15	Experimental	sauna	saucy	asset	genie	motel	nylon
16	Experimental	slang	slash	prize	dream	tract	waist
17	Experimental	twig	twin	golf	lark	punk	wind
18	Experimental	button	buzzer	cement	oyster	gutter	sailor
19	Experimental	police	patrol	impact	battle	spinal	talent
20	Experimental	stamp	stand	grade	world	blank	cliff
21	Experimental	walnut	walrus	kettle	ballet	tycoon	person
22	Experimental	poker	polar	comet	apple	towel	vinyl
23	Experimental	liquor	lizard	autumn	bamboo	collar	almond
24	Experimental	bleach	blazer	parrot	phrase	shrimp	cruise
25	Experimental	grain	grape	wrist	score	bounty	candle
26	Experimental	tickle	ticket	insect	window	morale	decree
27	Experimental	lunar	lucid	belly	pedal	award	tonic
28	Experimental	tangle	tanker	forest	coupon	reward	paddle
29	Experimental	carrot	castle	advice	pastry	dosage	errand
30	Experimental	dealer	deacon	shovel	farmer	marine	butter
31	Experimental	medal	melon	villa	hotel	salad	racer
32	Experimental	bread	break	steam	guest	hound	plank
33	Experimental	dollar	doctor	waiter	funnel	jockey	legion
34	Experimental	slime	slice	coast	bloom	porch	troll
35	Experimental	muscle	muzzle	result	sermon	flower	cuddle
36	Experimental	puffy	puppy	lemon	haven	sugar	zebra

Appendix A continued

		4-word displays in Exp. 1-4				Additional distractors for 6-word display in Exp. 5	
	Trial Type	Target	Competitor	Competitor	Distractor	Distractor	Distractor
1	Filler Overlap	bald	scam	scan	cube	epic	flag
2	Filler Overlap	bunk	cold	colt	drip	gasp	kind
3	Filler Overlap	flat	swab	swan	dorm	harp	lift
4	Filler Overlap	coma	army	arty	guru	body	duty
5	Filler Overlap	beard	grill	grimm	cross	dwarf	floor
6	Filler Overlap	drake	stock	stomp	greed	bland	carve
7	Filler Overlap	ghost	fruit	froze	drink	blind	clown
8	Filler Overlap	heart	fraud	frost	plump	crown	gland
9	Filler Overlap	lance	stork	storm	press	green	mount
10	Filler Overlap	scout	crime	crane	toast	guard	place
11	Filler Overlap	apron	combo	condo	bingo	decor	expel
12	Filler Overlap	bosom	femur	fetal	atlas	cargo	depot
13	Filler Overlap	choir	bagel	basin	forum	dowry	final
14	Filler Overlap	cynic	armor	arson	fatal	baron	dozen
15	Filler Overlap	disco	valid	valor	basil	cobra	forty
16	Filler Overlap	dummy	legal	liter	amber	curry	glory
17	Filler Overlap	fever	hippo	hindu	daisy	elbow	jelly
18	Filler Overlap	foggy	metal	metro	crazy	equal	jumbo
19	Filler Overlap	gravy	candy	carol	boxer	liver	major
20	Filler Overlap	hobby	bacon	baker	cedar	meter	sever
21	Filler Overlap	flight	bridge	bricks	lounge	coarse	adult
22	Filler Overlap	bazaar	trophy	trojan	keeper	mother	spider
23	Filler Overlap	beauty	vulgar	vulcan	locker	pinkie	ladder
24	Filler Overlap	bishop	muffin	mullen	layer	cookie	stupid
25	Filler Overlap	bottle	harbor	harlem	magnet	sleepy	afraid
26	Filler Overlap	cactus	barber	barbie	madman	diaper	silent
27	Filler Overlap	buffet	saline	saloon	meadow	people	asleep
28	Filler Overlap	curfew	piglet	picnic	napkin	bubble	rattle
29	Filler Overlap	dental	nature	napalm	orange	toilet	auntie
30	Filler Overlap	dragon	tanner	tavern	pencil	hungry	unsafe
31	Filler Overlap	ethnic	making	mating	potion	sister	forget
32	Filler Overlap	finish	bandit	batman	raffle	pretty	gentle
33	Filler Overlap	garage	neural	neuter	roster	little	marker
34	Filler Overlap	guitar	turner	turtle	sector	wiggle	driver
35	Filler Overlap	inmate	hunter	hunger	silver	dinner	beaver
36	Filler Overlap	hockey	armpit	arctic	tattoo	sticky	follow

Appendix A continued

		4-word displays in Exp. 1-4				Additional distractors for 6-word displays	
	Trial Type	Target	Distractor	Distractor	Distractor	Distractor	Distractor
1	Filler No Overlap	atom	blob	harm	lava	jest	copy
2	Filler No Overlap	bump	frog	hoax	menu	omen	gram
3	Filler No Overlap	clap	fuel	lion	hero	hint	item
4	Filler No Overlap	crop	gala	pond	prom	lazy	fury
5	Filler No Overlap	agree	blood	cider	store	plate	bring
6	Filler No Overlap	album	cadet	blink	stump	fluff	messy
7	Filler No Overlap	cigar	abuse	leper	swish	great	drive
8	Filler No Overlap	dress	cable	magic	tango	plant	spill
9	Filler No Overlap	eagle	cello	moral	topic	stick	first
10	Filler No Overlap	flame	clock	navel	trend	story	table
11	Filler No Overlap	flood	crate	paste	valet	smell	dizzy
12	Filler No Overlap	grief	crave	party	wagon	angel	scare
13	Filler No Overlap	image	cruel	pulse	window	spank	quick
14	Filler No Overlap	lever	donor	ranch	woman	sweep	child
15	Filler No Overlap	loser	farce	rural	vodka	bench	clean
16	Filler No Overlap	movie	flair	scalp	treat	juicy	crawl
17	Filler No Overlap	onion	fluid	shark	tenor	panda	color
18	Filler No Overlap	quota	giant	slave	level	boots	crumb
19	Filler No Overlap	regal	gorge	snail	linen	ankle	brake
20	Filler No Overlap	round	haste	speed	loyal	bunch	climb
21	Filler No Overlap	robot	jeans	squid	lotus	fuzzy	nasty
22	Filler No Overlap	arcade	fluent	notice	sample	winter	reader
23	Filler No Overlap	attack	gamble	outlaw	septic	ladder	secret
24	Filler No Overlap	report	boiler	golden	pardon	stupid	listen
25	Filler No Overlap	rudder	bullet	gopher	pastor	wallet	answer
26	Filler No Overlap	skater	wizard	candid	hazard	travel	giggle
27	Filler No Overlap	sniper	yellow	carbon	honest	beaver	cloudy
28	Filler No Overlap	infect	stress	zombie	cavern	peanut	normal
29	Filler No Overlap	insult	suitor	warden	chunky	cowboy	blonde
30	Filler No Overlap	cognac	kidney	tablet	zipper	bouncy	gobble
31	Filler No Overlap	cousin	letter	ticker	weapon	second	poison
32	Filler No Overlap	piston	danger	maiden	trader	middle	return
33	Filler No Overlap	polite	desert	master	twelve	bloody	unfair
34	Filler No Overlap	stable	priest	donkey	mister	lotion	bounce
35	Filler No Overlap	static	purple	fabric	murder	safety	clover
36	Filler No Overlap	vermin	weaken	reason	famine	become	bright

Note: Each row represents the words displayed during one trial. In Experiments 1-4, the Target column represent the auditory target that subjects will hear. The final two columns represent additional distractors used in 6-word displays for Experiment 5.

Supplementary Section
Experiment 1

Target vs. Competitor RF Output

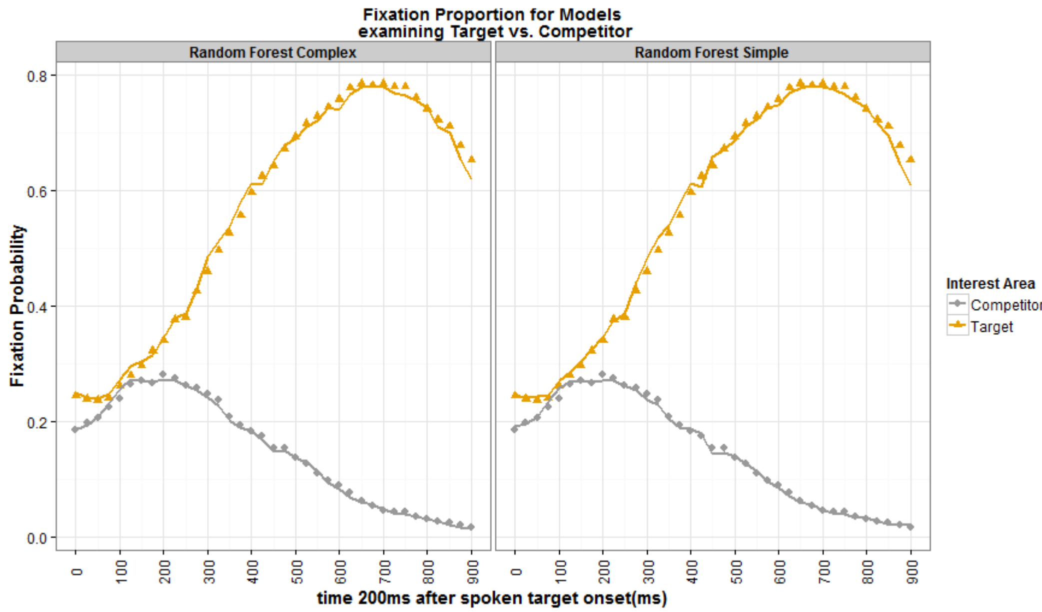
The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the root mean square error.

Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.974
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.984

Random Forest (Complex)	
Predictor	**Node purity increase**
T1:Condition	47.76
Condition (Target vs. Competitor)	19.45
T2:Condition	11.65
T1	11.62
T3:Condition	7.17
T2	2.41
T4:Condition	2.02
T3	1.45
T4	.91
Random Forest (Simple)	
Predictor	**Node purity increase**
T1:Condition	82.21
X1	18.99
Condition (Target vs. Competitor)	3.98

Table 1: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between target and competitor

Unlike the GCA model reported in the text of the dissertation, the RF complex model puts more weight on the interaction between quadratic polynomial and condition rather than cubic polynomial and condition. Moreover, RF complex indicates that a quartic polynomial is likely redundant. RF simple indicates that cross validation performance is greater if one simply considers a model that has only the linear polynomial.



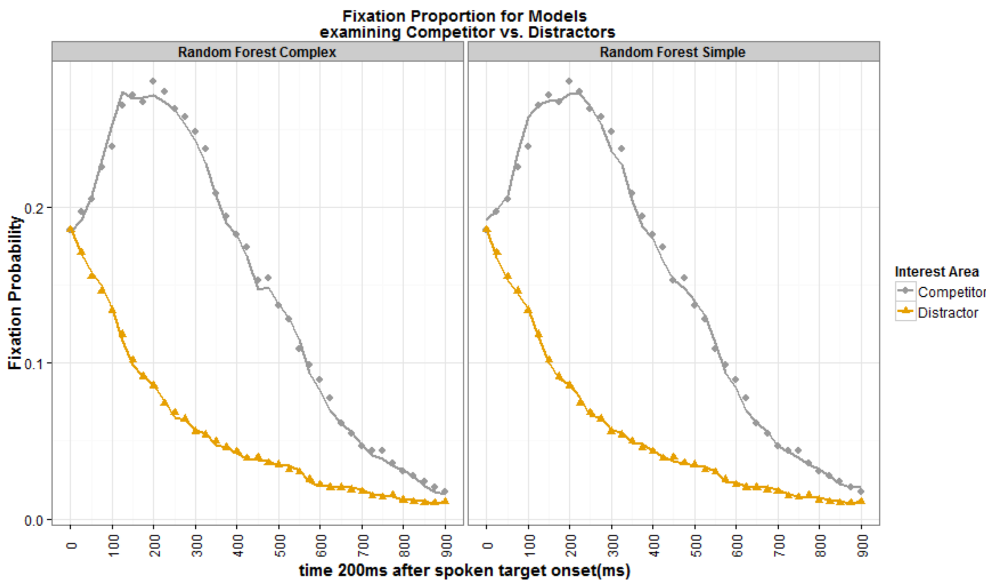
Competitor vs. Distractors RF Output

Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.941
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.955

Random Forest (Complex)	
Predictor	**Node purity increase**
T1	6.08
T3:Condition	2.85
Condition (Competitor vs. Distractor)	0.86
T2	0.69
T1:Condition	0.51
T3	0.48
T4	0.30
T2:Condition	0.28
T4:Condition	0.10
Random Forest (Simple)	
Predictor	**Node purity increase**
T1	7.47
T1:Condition	2.76
Condition (Competitor vs. Distractor)	1.72

Table 2: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between competitor and distractors.

The complex RF indicates that the reduction in MSE is driven by the interaction term T3:Condition and the main effect of Condition, with other interaction terms playing a minor role. A RF with a simpler structure indicates that a simpler model is sufficient to achieve good prediction accuracy.



Experiment 2

Target vs. Competitor RF Output

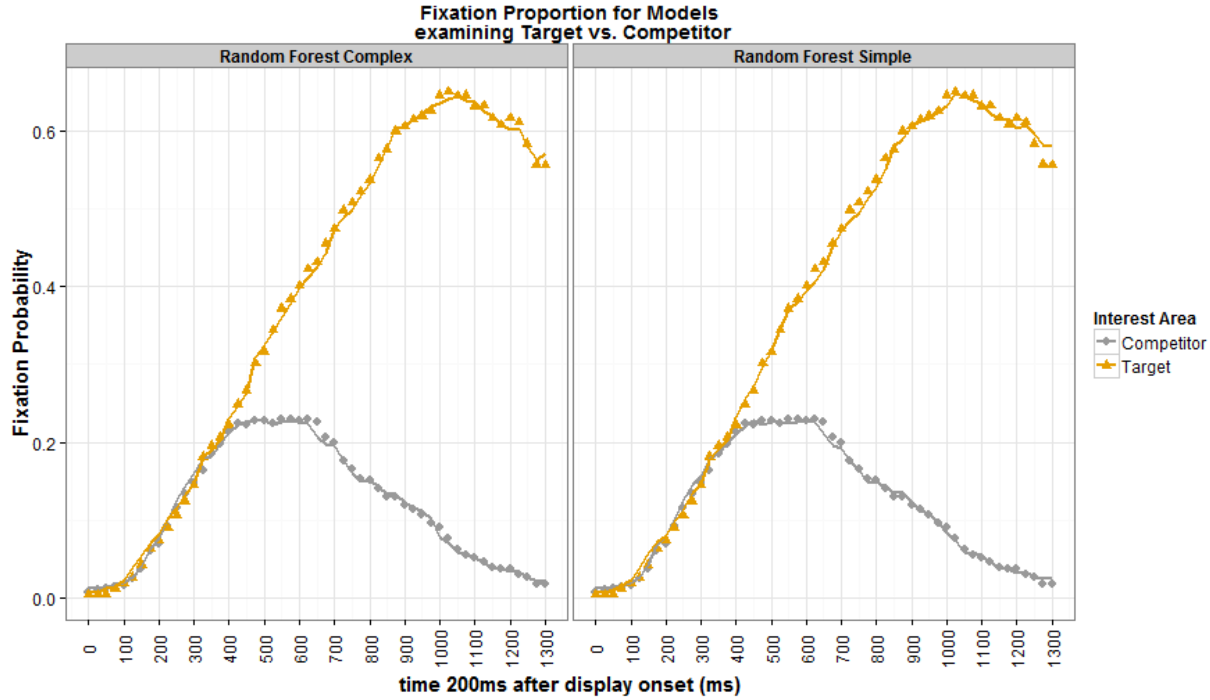
Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.977
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.981

The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing MSE.

Random Forest (Complex)	
Predictor	**Node Purity**
T1:Condition	81.29

T2	14.34
T4:Condition	3.26
T2:Condition	2.62
T3	1.69
T1	1.10
T4	0.96
T3:Condition	0.47
Condition (Target vs. Competitor)	0.10
Random Forest (Simple)	
Predictor	**Node Purity**
T1:Condition	83.71
X1	20.02
Condition (Target vs. Competitor)	2.08

Table 3: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between target and competitor.



The output of the RF complex is also similar to what was observed in Experiment 1, except that the interaction between quadratic polynomial and condition (T2:Condition) does not play as big of a role in reducing MSE across regression tress. RF simple again indicates that the linear polynomial is sufficient for the model to have good cross-validated performance.

Competitor vs. Distractors RF Output

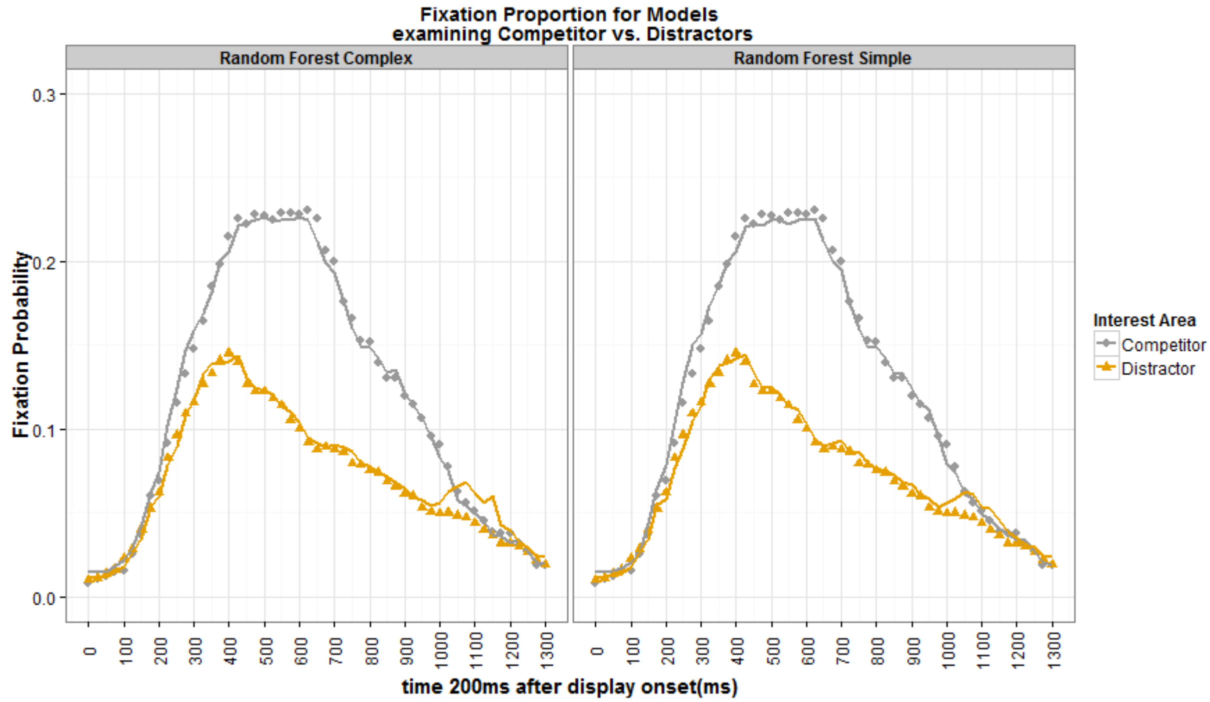
Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.899
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.918

The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the mean square error for out-of-bag data.

Random Forest (Complex)	
Predictor	**Node Purity**
T2	5.48
Condition (Competitor vs. Distractor)	1.69
T3	1.63
T4	0.59
T3:Condition	0.46
T1	0.39
T2:Condition	0.26
T1:Condition	0.09
T4:Condition	0.08
Random Forest (Simple)	
Predictor	**Node Purity**
T1	8.22
Condition (Competitor vs. Distractor)	1.99
T1:Condition	0.36

Table 4: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between competitor and distractors.

The complex RF model indicates that the main driver of reduction in MSE is Condition, followed by the interaction between quadratic polynomial and condition, and that the least important role in error reduction is due to T4:Condition. Again, this indicates that it is very likely that the interaction between T4:Condition in GCA is due to overfitting the data. The simple RF model again is superior to the complex model.



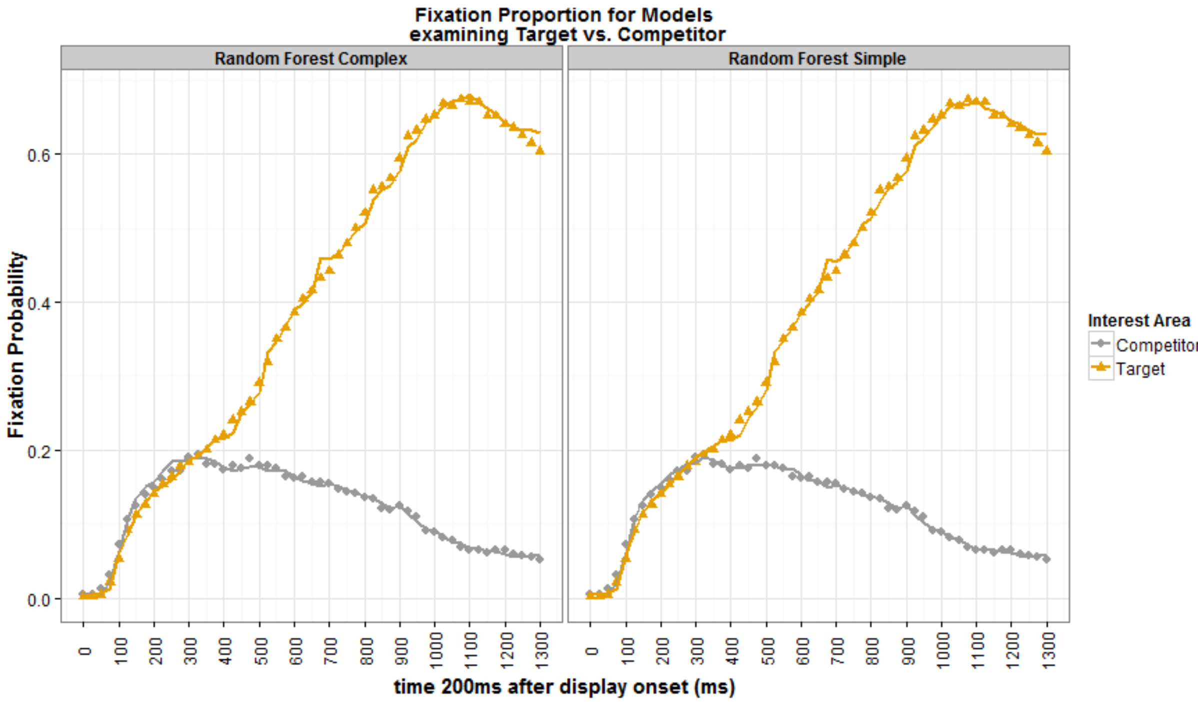
Experiment 3

Target vs. Competitor RF Output

Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.973
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.981

Random Forest (Complex)	
Predictor	**Node Purity**
T1:Condition	87.45
T2:Condition	9.32
T2	4.54
T1	2.51
T3	1.29
T4	1.19
T4:Condition	0.74
T3:Condition	0.46
Condition (Target vs. Competitor)	0.10
Random Forest (Simple)	
Predictor	**Node Purity**
T1:Condition	90.58
X1	12.38
Condition (Target vs. Competitor)	3.67

Table 5: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between target and competitor.



RF complex indicate that the main driver of reduction in MSE is the interaction between linear polynomial and condition and quadratic polynomial and condition. All other interactions have much smaller importance in reducing the error across bagged tress. RF simple again outperforms the RF complex.

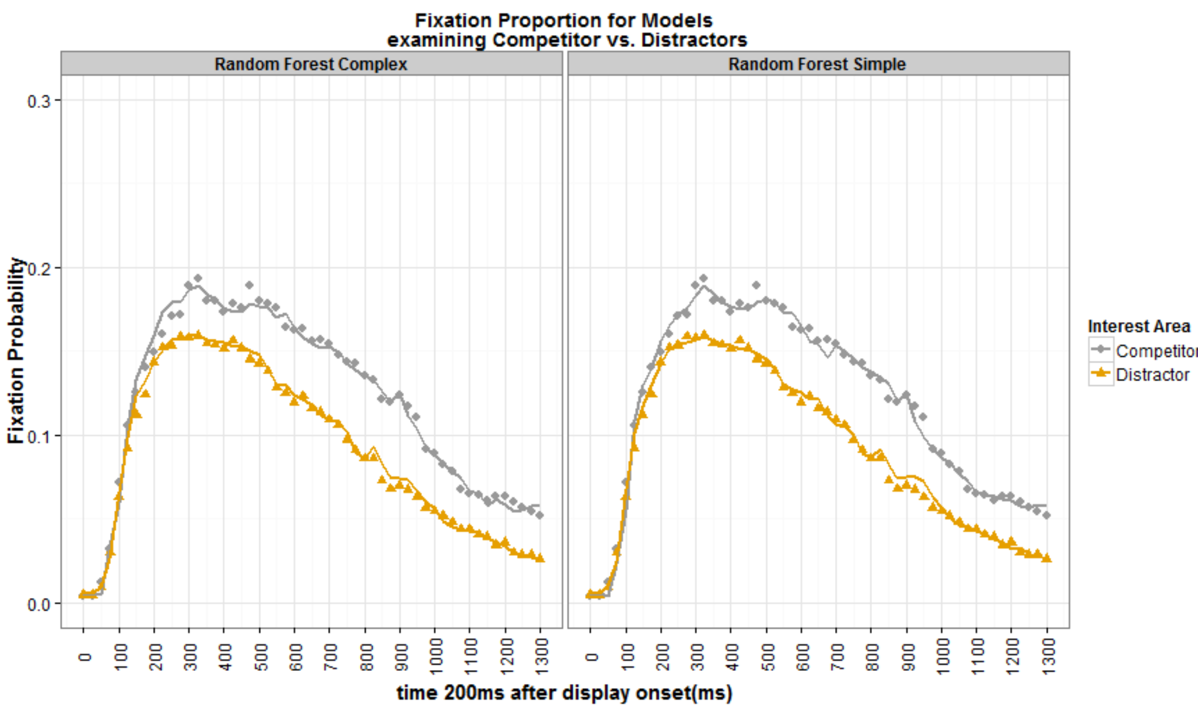
Competitor vs. Distractors RF Output

Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.891
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.912

The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the root mean square error for out-of-sample data.

Random Forest (Complex)	
Predictor	**Node Purity**
T2	3.02
T1	2.40
T3	0.89
T4	0.61
T1:Condition	0.53
T2:Condition	0.36
T3:Condition	0.30
Condition (Competitor vs. Distractor)	0.19
T4:Condition	0.17
Random Forest (Simple)	
Predictor	**Node Purity**
T1	6.70
T1:Condition	1.29
Condition (Competitor vs. Distractor)	0.42

Table 6: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between competitor and distractors.



The complex RF models points to the same result: the interaction terms have minimal contribution in reducing MSE across bagged trees. A simple RF again provided a better fit.

Experiment 4

Target (congruous) vs. Target (incongruous comp.) vs. Target (incongruous distr.)

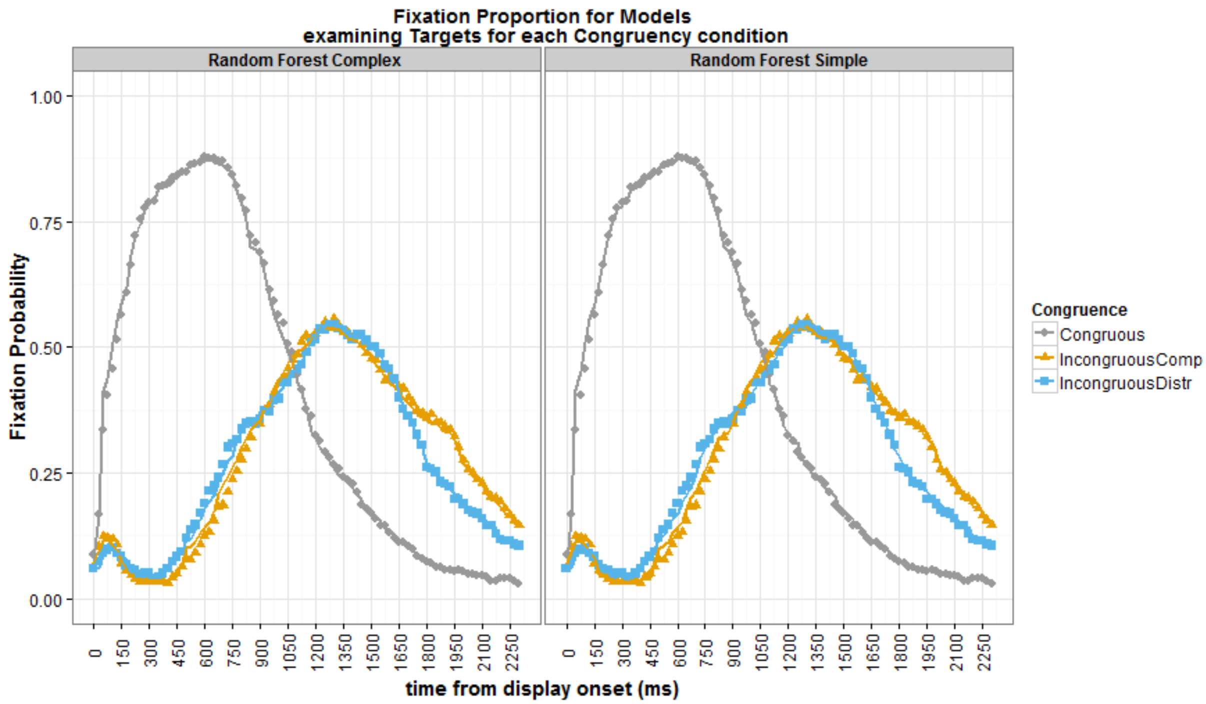
Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.967
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.976

The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the root mean square error for out-of-sample data.

Random Forest (Complex)	
Predictor	**Node Purity**
T1:(Congruous vs. Incon. Comp.)	161.21
T2	102.98
T3:(Congruous vs. Incon. Comp.)	47.92
T1	43.60
T4	24.58
T2:(Congruous vs. Incon. Comp.)	21.70
T4:(Congruous vs. Incon. Comp.)	20.10
T3	19.27
T2:(Incon. Distr. vs. Incon. Comp.)	16.93
T1:(Incon. Distr. vs. Incon. Comp.)	7.58
Congruous vs. Incon. Comp.	6.61
T3:(Incon. Distr. vs. Incon. Comp.)	4.88
T4:(Incon. Distr. vs. Incon. Comp.)	4.17
Incon. Distr. vs. Incon. Comp.	3.63
Random Forest (Simple)	
Predictor	**Node Purity**
T1:(Congruous vs. Incon. Comp.)	261.41
T1	183.80
T1:(Incon. Distr. vs. Incon. Comp.)	17.08
Incon. Distr. vs. Incon. Comp.	11.02
Congruous vs. Incon. Comp.	8.73

Table 7: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between target and competitor.

The complex RF model support the same conclusion as the GCA model reported in the body of the dissertation. The RF model indicated that targets in the congruous condition have substantially different fixation curves from the targets in the other two conditions.



Competitor (congruous) vs. Competitor (incongruous comp.) vs. Competitor (incongruous distr.)

Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.963
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.966

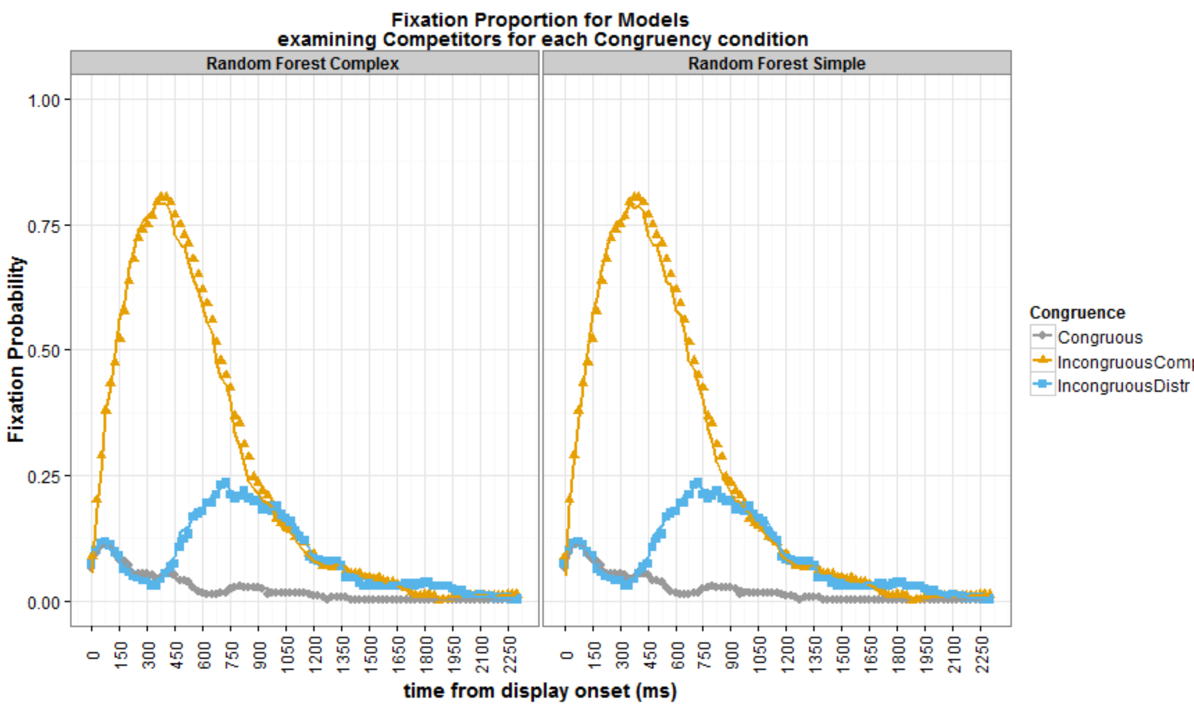
The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the root mean square error for out-of-sample data.

Random Forest (Complex)	
Predictor	**Node Purity**
T1:(Incon. Distr. vs. Incon. Comp.)	206.98
T1	14.94
T4:(Incon. Distr. vs. Incon. Comp.)	12.27
T4	11.83
T2	6.60
Congruous vs. Incon. Comp.	6.31
T3	5.43
T2:(Incon. Distr. vs. Incon. Comp.)	4.18
T3:(Incon. Distr. vs. Incon. Comp.)	2.49

T3:(Congruous vs. Incon. Comp.)	1.55
Incon. Distr. vs. Incon. Comp.	0.67
T1:(Congruous vs. Incon. Comp.)	0.55
T2:(Congruous vs. Incon. Comp.)	0.38
T4:(Congruous vs. Incon. Comp.)	0.29
Random Forest (Simple)	
Predictor	**Node Purity**
T1:(Incon. Distr. vs. Incon. Comp.)	224.96
T1	38.75
Congruous vs. Incon. Comp.	7.93
Incon. Distr. vs. Incon. Comp.	2.93
T1:(Congruous vs. Incon. Comp.)	1.33

Table 9: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between competitor and distractors.

RFs indicate that the main driver of differences in fixation curves between competitor in the incongruous distr. condition and competitor in the incongruous comp. condition is the interaction with the linear polynomial.



Experiment 5

Target vs. Competitor Group 1

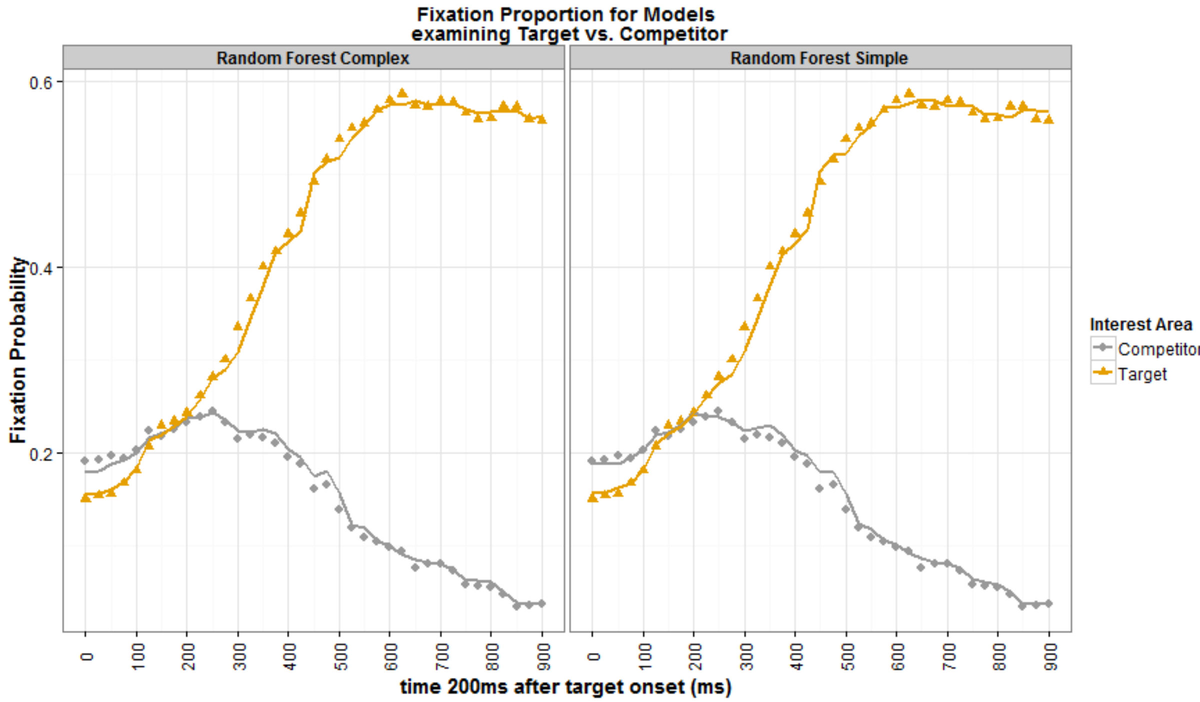
Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.967
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.974

The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the root mean square error for out-of-sample data.

Random Forest (Complex)	
Predictor	**Node Purity**
T1:Condition	40.91
T2:Condition	6.07
T1	5.52
T3	0.73
T2	0.64
T4	0.39
T4:Condition	0.28
T3:Condition	0.26
Condition (Target vs. Competitor)	0.04
Random Forest (Simple)	
Predictor	**Node Purity**
T1:Condition	44.07
X1	10.26
Condition (Target vs. Competitor)	0.44

Table 10: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between target and competitor.

Both GCA and RF models show the same pattern of results that have been observed in Experiments 1-3. The linear interaction seems to be the main driver between fixation curves for target vs. competitor.



Target vs. Competitor Group 2

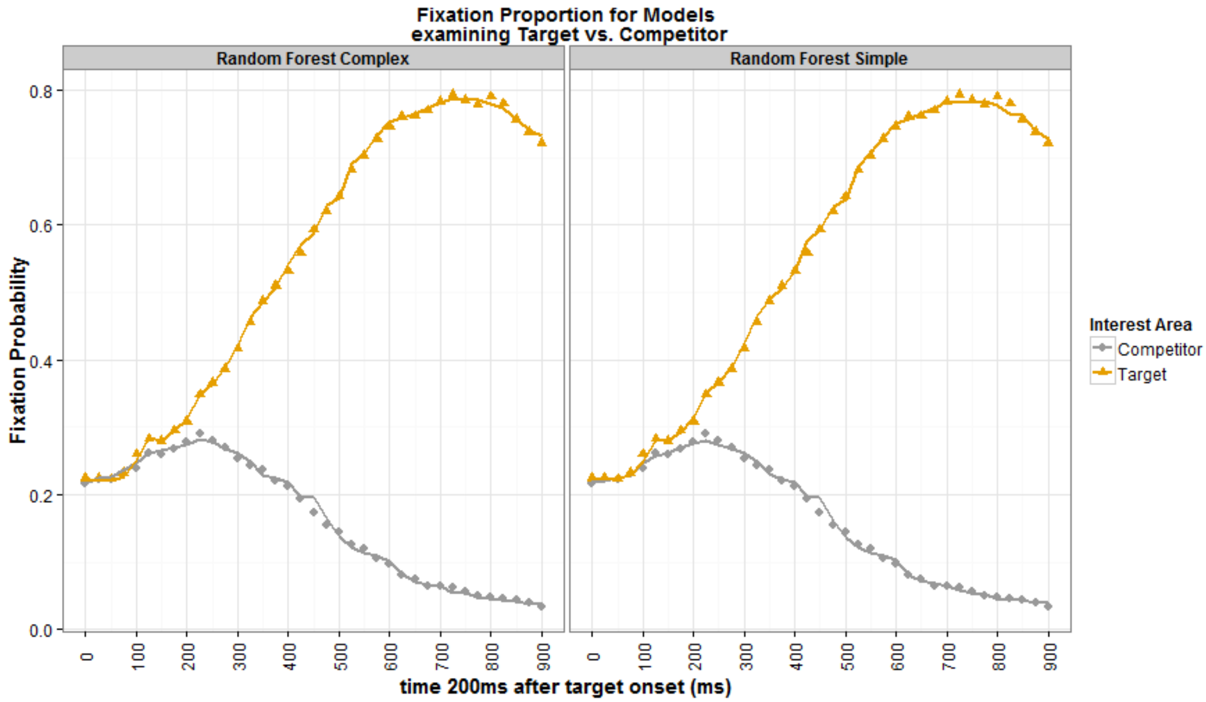
Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.983
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.987

The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the root mean square error for out-of-sample data.

Random Forest (Complex)	
Predictor	**Node Purity**
T1:Condition	58.66
T1	13.31
Condition (Target vs. Competitor)	12.60
T2:Condition	5.90
T3:Condition	5.33
T2	2.50
T4:Condition	1.87
T3	1.37
T4	0.72
Random Forest (Simple)	
Predictor	**Node Purity**

T1:Condition	93.64
X1	15.99
Condition (Target vs. Competitor)	3.01

Table 11: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between target and competitor.



Similar results obtained in Group 1 have been obtained in Group 2.

Competitor vs. Distractors Group 1

Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.923
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.940

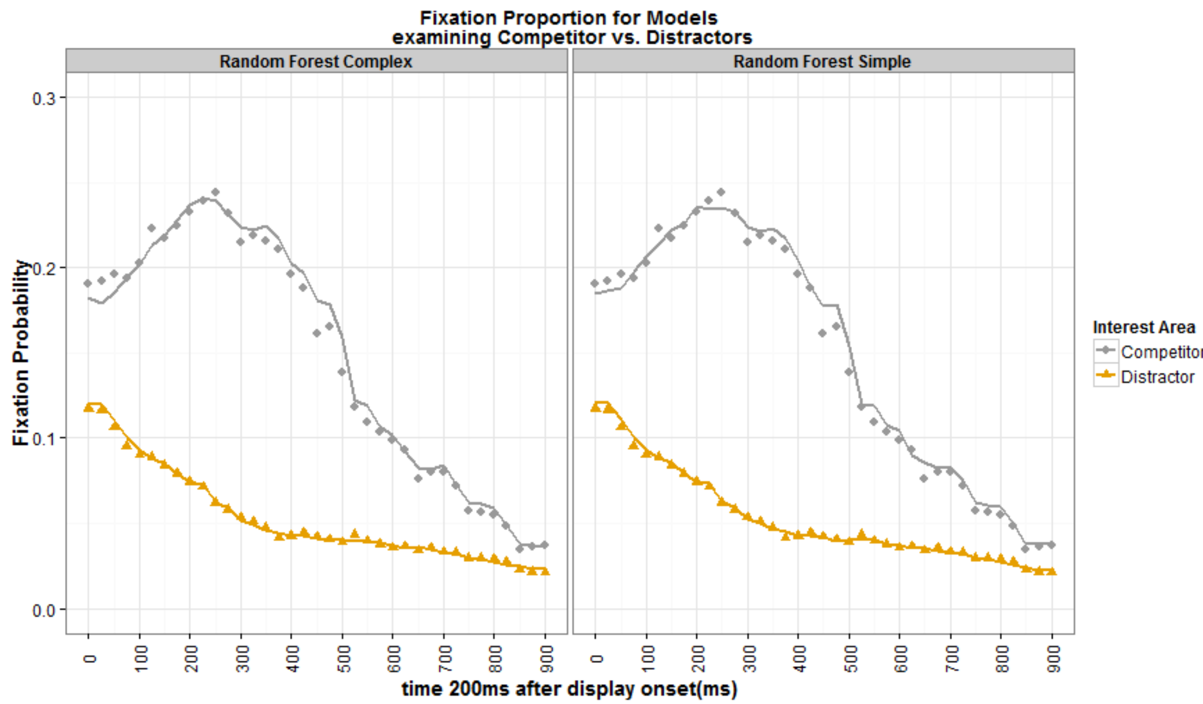
The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the root mean square error for out-of-sample data.

Random Forest (Complex)	
Predictor	**Node Purity**
T1	3.90
Condition (Target vs. Competitor)	3.47
T2	0.32
T3	0.27

T1:Condition	0.27
T4	0.18
T3:Condition	0.02
T2:Condition	0.01
T4:Condition	0.01
Random Forest (Simple)	
Predictor	**Node Purity**
T1	4.66
Condition (Competitor vs. Distractor)	3.44
T1:Condition	0.33

Table 12: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between competitor and distractors.

Both GCA and RF indicate to the presence of the competitor effect.



Competitor vs. Distractors Group 2

Model	Model Structure	R^2
RF (complex)	Fixation Proportion ~ Condition*(T1+T2+T3+T4)+SubjectID	.960
RF (simple)	Fixation Proportion ~ Condition*(T1)+SubjectID	.966

The table below presents the variable importance output for the complex and simple random forests models. Variables with higher values for node purity had greater importance in reducing the root mean square error for out-of-sample data.

Random Forest (Complex)	
Predictor	**Node Purity**
T1	6.91
Condition (Competitor vs. Distractor)	2.23
T2:Condition	1.92
T3:Condition	1.12
T2	0.84
T1:Condition	0.74
T3	0.43
T4	0.21
T4:Condition	0.05
Random Forest (Simple)	
Predictor	**Node Purity**
T1	8.54
T1:Condition	4.51
Condition (Competitor vs. Distractor)	1.32

Table 13: Node purity importance for the main variables used in the complex and simple random forest models comparing fixations between competitor and distractors.

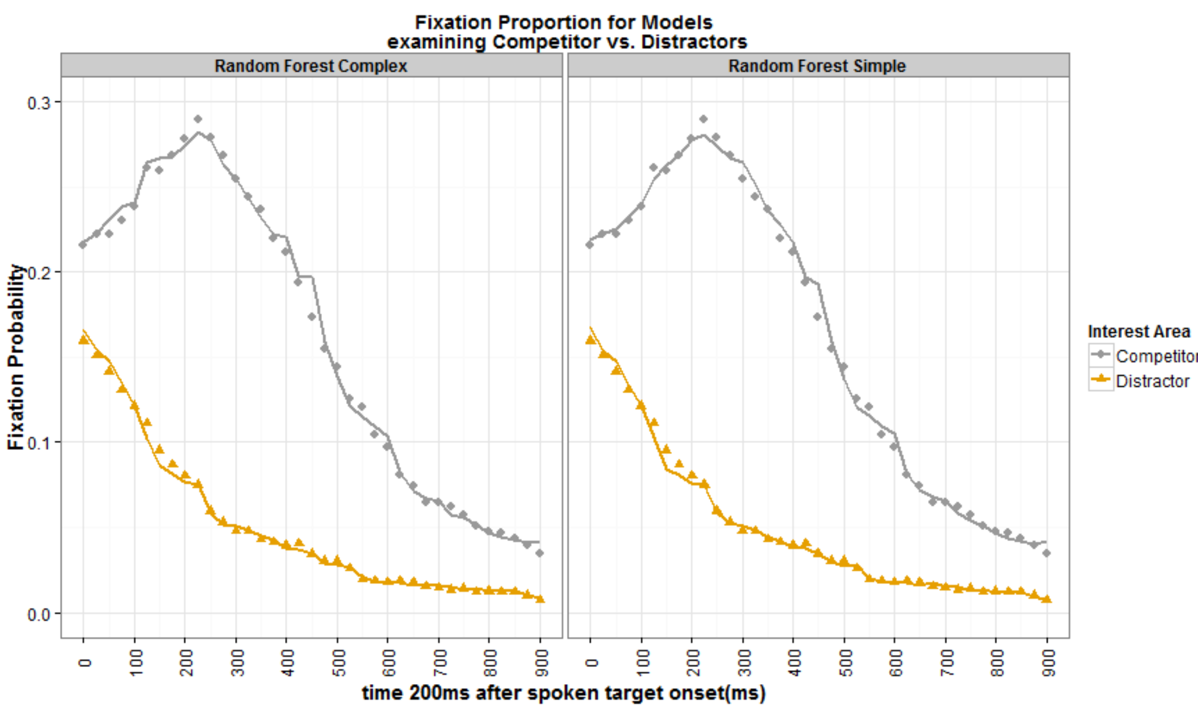


Figure 43

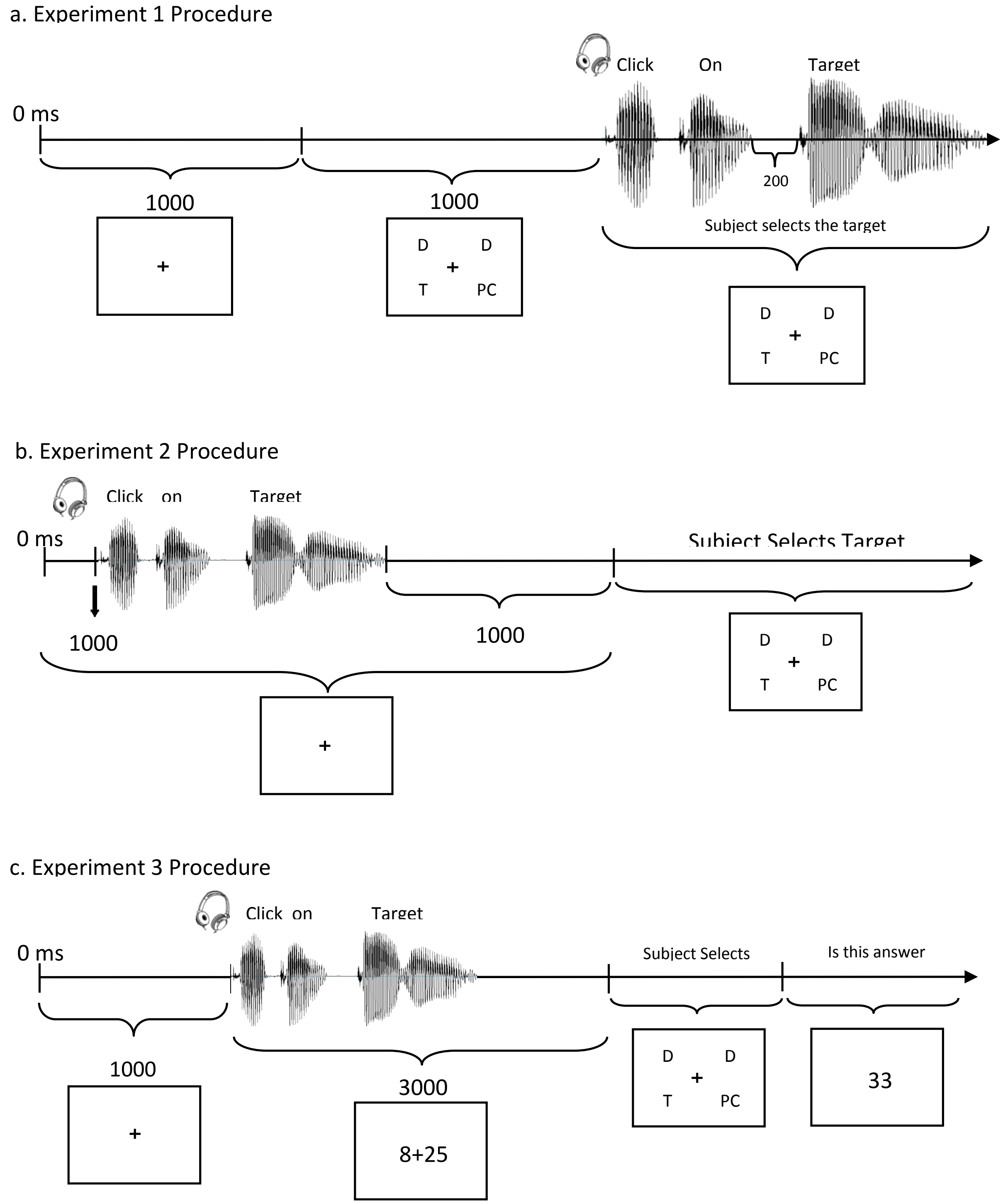
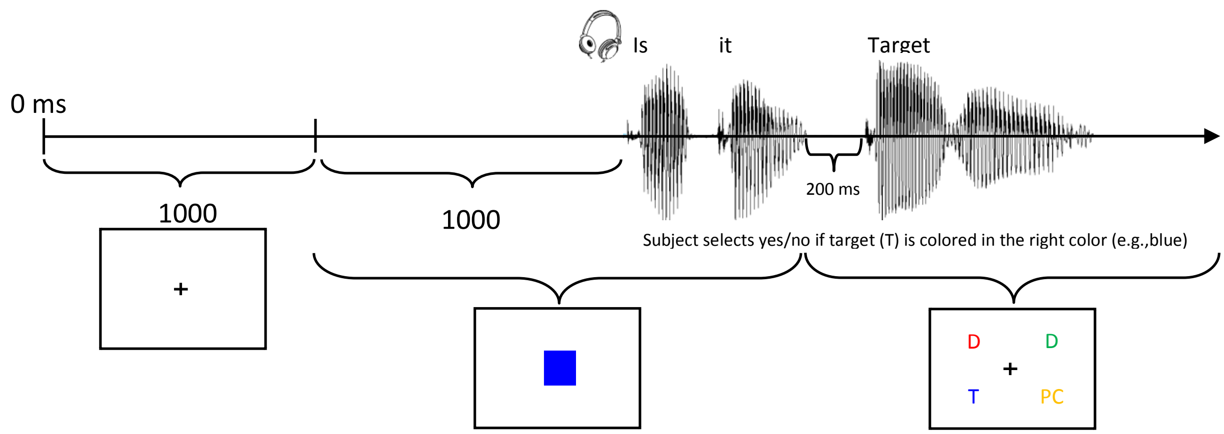

a. Experiment 1 Procedure
Click
On
Target
0 ms
1000
1000
200
Subject selects the target
+
D
D
+
T
PC
D
D
+
T
PC
b. Experiment 2 Procedure
Click
on
Target
0 ms
Subject Selects Target
1000
1000
D
D
+
T
PC
+
c. Experiment 3 Procedure
Click
on
Target
0 ms
Subject Selects
Is this answer
1000
3000
+
8+25
D
D
+
T
PC
33

Figure 7 continued

d. Experiment 4 Procedure



e. Experiment 5 Procedure