Journal of Memory and Language 121 (2021) 104279

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm

Keith S. Apfelbaum [a,*], Jamie Klein-Packard [a], Bob McMurray [a,b]

[a] Dept. of Psychological and Brain Sciences, University of Iowa, United States
[b] Dept. of Communication Sciences and Disorders, Dept. of Linguistics, Dept. of Otolaryngology, University of Iowa, United States

ARTICLE INFO

Keywords:
Visual world paradigm
Visual context
Word recognition
Phonological processing

ABSTRACT

A common critique of the Visual World Paradigm (VWP) in psycholinguistic studies is that what is designed as a measure of language processes is meaningfully altered by the visual context of the task. This is crucial, particularly in studies of spoken word recognition, where the displayed images are usually seen as just a part of the measure and are not of fundamental interest. Many variants of the VWP allow participants to sample the visual scene before a trial begins. However, this could bias their interpretations of the later speech or even lead to abnormal processing strategies (e.g., comparing the input to only preactivated working memory representations). Prior work has focused only on whether preview duration changes fixation patterns. However, preview could affect a number of processes, such as visual search, that would not challenge the interpretation of the VWP. The present study uses a series of targeted manipulations of the preview period to ask if preview alters looking behavior during a trial, and why. Results show that evidence of incremental processing and phonological competition seen in the VWP are not dependent on preview, and are not enhanced by manipulations that directly encourage phonological prenaming. Moreover, some forms of preview can eliminate nuisance variance deriving from object recognition and visual search demands in order to produce a more sensitive measure of linguistic processing. These results deepen our understanding of how the visual scene interacts with language processing to drive fixations patterns in the VWP, and reinforce the value of the VWP as a tool for measuring real-time language processing. Stimuli, data and analysis scripts are available at https://osf.io/b7q65/.

Introduction

Speech is fundamentally temporal: information unfolds over time, cues are transient, and the order of sounds and words matters. Words arrive in rapid succession, forcing the listener to quickly recognize each word, and the boundaries between words are often ambiguous (e.g., *car go* vs. *cargo*). Yet listeners adeptly navigate these challenges to recognize speech.

The importance of time has led to innovative measurement techniques to capture the temporal unfolding of language processing. A widely used one is eye tracking in the Visual World Paradigm (VWP). In the VWP, participants hear spoken instructions in the presence of a visual scene containing objects that represent one or more candidate interpretations. Eye movements to the objects are used to track what listeners consider *while* they process language over time. This allows us to assess the state of the language system as processing unfolds. A longstanding concern is that the visual context may limit the ability to cleanly link eye movements in the VWP to general aspects of language processing by biasing processing or even constraining the linguistic forms that are considered. This could limit the generality of findings (see Huettig et al., 2011; Magnuson, 2019 for discussion).

The foundational work in the VWP explicitly asked if visual context affects language processing (Tanenhaus et al., 1995; see also, Altmann & Kamide, 2007; 2009; Hanna & Brennan, 2007; Sedivy et al., 1999). These studies focused on sentence processing, and here, visual context was a proxy for discourse or real-world context. Consequently, the way in which visual context was integrated with speech was critical evidence for resolving debates in sentence processing. That is, the use of visual context was the focus of this research—not a confound.

However, the VWP has also been applied to domains like spoken word recognition, where the critical questions concern auditory, phonological or semantic processes, not the sensitivity of word recognition to broader context. Here, the fixations directed to the visual display are assumed to index the degree to which words are activated in

* Corresponding author at: Dept. of Psychological and Brain Sciences, G60 PBSB, University of Iowa, Iowa City, IA 52242, United States.
E-mail address: Keith-apfelbaum@uiowa.edu (K.S. Apfelbaum).

https://doi.org/10.1016/j.jml.2021.104279
Received 28 September 2020; Received in revised form 28 June 2021; Accepted 30 June 2021
Available online 10 July 2021

the lexicon, implicitly assuming that the display does not alter lexical processing in meaningful ways. However, if the visual context alters language processing, findings in these studies could be limited to the kinds of circumstances created by the VWP, rather than reflecting language processing in general.

Measuring word recognition in the VWP

This study examines these issues in word recognition, where there are explicit models of the process, and where concerns about the role of the visual scene can be translated to experimental manipulations. We ask whether the visual display creates an artificial situation by priming phonological forms, reducing the decision space or invoking other cognitive processes epiphenomenal to word recognition. More broadly, we ask how the complex interactions of visual and lexical processing interact across time in the VWP, especially before a trial begins.

Word recognition requires rapidly identifying a word from a massive set of options (a typical adult English speaker might know over 40,000 words; Brysbaert et al., 2016). This occurs in the face of considerable temporal ambiguity from similar sounding words (e.g., *sandal* and *sandwich*), and imperfect cues to word boundaries (e.g., *car go* vs. *cargo*). This process is well understood (Dahan & Magnuson, 2006; Weber & Scharenborg, 2012). In a few hundred milliseconds, listeners activate a range of candidates that partially match the input, rule out competing words and access the correct word. Theories of word recognition rely on measures like the VWP that are sensitive to the subtle dynamics of this competition as it occurs and that can assess which words are considered over time as processing unfolds.

The VWP overcomes limitations of conventional tasks like cross-modal priming and gating with a relatively natural task that samples activation while lexical competition is ongoing. Eye movements are metabolically cheap and mostly launched without awareness. Listeners often launch multiple fixations over a trial. As a result, eye-tracking can measure where attention is directed in precise time increments with potentially close time-locking to unfolding decisions. Also, fixations to a specific image reflect attention to that word – not a general sense of processing difficulty. The VWP then harnesses this time-locking to provide a more direct measure of what items are considered as recognition unfolds.

The typical VWP design for word recognition uses (typically four) images chosen to assess activation for specific classes of words. To assess phonological competitors, a display might include a target (*sandal*), and onset competitors (*sandwich*) or rhymes (*candle*). Typically, upon hearing the *sa-* in *sandal*, listeners fixate *sandal* and *sandwich*, but not *candle*, or an unrelated item like *necklace*. As more of the word is heard, looks to onset competitors rapidly drop off; after hearing *sanda-*, the listener stops looking at the *sandwich*, whereas looks to the *sandal* continue to increase. Later, items that didn't fully match the onset receive some consideration (e.g., *candle*), suggesting that lexical activation reflects the overall phonological form of the input. The patterns of fixations across time indicate that lexical processing is incremental, parallel, and subject to competition.

The VWP has yielded considerable insight into the dynamics of lexical activation (e.g., Allopenna et al., 1998; Dahan & Gaskell, 2007; Toscano et al., 2013), and competition (e.g., Magnuson et al., 2003, 2007), and even into group-level lexical processing differences, including for people with developmental language disorders (McMurray et al., 2010), people who use cochlear implants (Farris-Trimble et al., 2014), and elderly individuals (Revill & Spieler, 2012).

Despite these findings, there remain concerns about whether the visual context meaningfully alters fixation patterns in word recognition paradigms. Critically, the specific items on the screen are rarely of individual interest. That is, the researcher doesn't (usually) care if *sandwich* is specifically accessed. Rather *sandwich* is a sample from a set of onset competitors, and the listener is assumed to activate other, non-displayed onset competitors like *sandbar* or *Santa*. This assumption is essential for treating the VWP as a measure of lexical processing.

However, in many versions of the VWP, participants preview the response options before hearing the words. This could bias fixations in at least two ways. First, the displayed pictures could prime their corresponding words, or inhibit activation of other words. This would begin the process of lexical activation before auditory input is received, which may not offer a good measure of unconstrained processing. This kind of mechanism could play out in the lexicon via priming or attentional processes (Mirman et al., 2008) by which preview activates semantic features that feed back to bias lexical activation. In this case, the fixation record is distorted by the preview—over- or under-emphasizing activation for some words—but it still fundamentally reflects lexical processing.

Second, an even more challenging possibility is that fixations in the VWP do not reflect lexical processing at all. Listeners could generate names for each object in phonological working memory ("prenaming") and recognition could play out in working memory as the incoming speech is matched to these wordforms (see discussion of this possibility in Huettig et al., 2011). Rather than viewing fixations as indicative of underlying activation dynamics, they might instead reflect performance in a memory task which is unrepresentative of processing in the 40,000-alternative lexicon. This challenges the fundamental construct validity of the VWP.

The ability of the VWP to capture lexical processing in an unbiased way thus depends on whether the visual display alters the way that words are processed, and the cognitive processes that underlie these mechanisms. Although these concerns rarely arise in published work (though see Huettig et al., 2011 for one discussion of this possibility and Andersson et al., 2011; Henderson & Ferreira, 2013, for related ideas), we and other users of the VWP frequently deal with this critique during the review process. For example, in a recent VWP paper submitted in 2020 by one of us, one reviewer critiqued the study on the basis that

> …the paper presents the VWP as providing information about language processing generally and does not address criticisms that it reflects language processing in a narrow experimental paradigm where there is a closed set of objects that could be mentioned … For example, there is evidence that typical results (e.g., cohort effects) *depend* [our emphasis] on allowing participants to preview the screen and to subvocally pre-generate the names of the objects… there is little reason to believe that the method says anything about language comprehension in the absence of a simple visual display.

The proper response to such critiques is usually to point to studies (like those summarized below) that suggest the influence of non-pictured alternatives on fixations or that show activation of words that are unlikely to be prenamed. Such studies rule the strong form of this critique. Nonetheless, beyond the methodological issues, even weaker forms this claim imply an integration of visual and language behavior that may be interesting in its own right (e.g., Huettig & Altmann, 2011; Spivey et al., 2001). Thus, this issue warrants direct empirical investigation to determine how preview of responses interacts with the complex link between lexical processing and eye movements.

The role of preview

It is helpful to frame these concerns in terms of a linking hypothesis (Magnuson, 2019; Tanenhaus et al., 2000) —the set of processes which link the thing we care about (lexical competition) to the observed behaviors. The simplest assumed linking function between the VWP and lexical activation is straightforward, but also oversimplified: a listener fixates an object to the degree that it is active. As Tanenhaus and colleagues defined it, the VWP taps "automated behavioral routines that link a name to its referent; when the referent is visually present and task relevant, then recognizing its name accesses these routines, triggering a saccadic eye movement to fixate the relevant information" (Tanenhaus

et al., 2000, p. 565).

We now know that this task is considerably more complicated than this picture (Huettig et al., 2011; Magnuson, 2019). In a typical word recognition study, several processes, both linguistic and non-linguistic, must take place to generate a meaningful fixation. The participant must hear the stimulus and activate lexical representations for various candidates (*sandal* and *sandwich*). The semantic representations of these activated lexical entries must be accessed (to match it to a picture). These processes could be reasonably described as lexical. The semantic features must be linked to visual features; and those visual features must be found and fixated in the array. These processes could be considered non-lexical, but they are critical for determining eye movements in the VWP. Interpretation of the VWP often focuses exclusively on lexical activation (typically at the level of wordforms): the degree of fixating an item reflects the degree of its activation. This ignores these other factors, treating them as noise. We need to understand how these other processes affect fixation patterns, and how all these processes interact during the course of a VWP trial, to appropriately interpret VWP data.

The preview of the visual display is designed to take care of some of the non-linguistic tasks before the word is heard. Preview lets participants activate visuo-semantic features and bind them to spatial locations. As a result, when they hear the word, fixations should be primarily driven by lexical processes. Thus, response preview might improve the specificity of the link between underlying lexical processing and fixations to the available competitors by minimizing non-lexical processes. However, preview could also alter the decision space, either by priming or inhibiting words within the lexicon, or by invoking explicit prenaming processes. That is, preview could alter the lexical parts of the process, not just the nonlinguistic parts. Thus, the validity of the VWP as measures of word recognition depends on our understanding of the effects of preview.

Preview does not block activation of non-displayed alternatives

The strongest preview critique suggests that participants treat each trial as a 4AFC closed-set task (within the lexicon or working memory). This seems unlikely based on intuitive considerations and empirical evidence.

First it is unlikely that participants could activate all possible linguistic forms from a display (Magnuson, 2019). Even a concrete object like a *wizard* could be named a *sorcerer, magician, warlock,* or *Harry*, or could indicate properties or concepts (*magic, spell, nemesis of He Who Shall Not Be Named,* etc.). Moreover, even if objects had only one name, they would likely require more than one "slot" in memory. Longer words, for example, require more resources, and each word would also need to be stored with its location in space. Moreover, work in visual memory suggests that in pseudo-naturalistic tasks participants store only what is needed right then and there, and attempt to minimize what is stored in working memory (Ballard et al., 1995), leaving information "in the world." Thus, listeners likely lack the memory resources to explicitly prename in a VWP trial, and might not employ them even if had them.

Second, both the prenaming and feedback-based closed-set arguments make empirical predictions that are not supported. For example, fixations are sensitive to the lexical frequency of the displayed items (Dahan, Magnuson, & Tanenhaus, 2001); if listeners only consider pictured items, then global lexical characteristics should not play a role (Sommers et al., 1997; though see, Clopper et al., 2006). Similarly, word recognition in the VWP shows effects of phonological density, and especially cohort density even when no neighbors are present on the screen (Magnuson et al., 2003, 2007). This implies that non-pictured neighbors are active and competing for recognition during the trial.

Evidence also supports the specific activation of non-pictured alternatives. Dahan and colleagues (Dahan, Magnuson, Tanenhaus, et al., 2001; see also, Kapnoula et al., 2015; McMurray et al., 2019) presented participants with a word (e.g. *net*) in a display that had no direct competitors. On some trials, the onset of the word (*ne-*) was spliced from another word *(neck,* not displayed); on other trials, the onset was spliced from a nonword (*nep*). When the coarticulatory cues partially activated a competing word (the *neck* case), participants showed slower fixations to the *net* than when they favored *nep*. This is evidence that they activated the competing wordform (*neck*), which inhibited the target, despite the fact that the competitor was not display. This effect also arises after training with novel wordforms with no meaning (Kapnoula et al., 2015). The visual display is not preventing even meaningless wordforms from being considered.

Two additional lines of evidence further challenge the strong form of prenaming. First, cross-linguistic competition in bilingual versions of the VWP indicates broader lexical activation. Marian, Spivey and colleagues show that bilingual speakers activate lexical items in both their languages, even when the study is conducted entirely in one language (Marian et al., 2003; Shook & Marian, 2012; Spivey & Marian, 1999), and this phenomenon has been observed even in 2nd year second language learners (Sarrett et al., submitted for publication). Working memory capacity makes it unfeasible to preactivate and maintain words in both languages for each picture.

Activation is also not limited by listeners' preferred names for objects. Pontillo, Salverda and Tanenhaus (2015) identified images that have two names (e.g. *sofa* and *couch*), but one that is likely preferred. On critical trials the target was a phonological competitor of either the dominant or non-dominant name (*soda* or *cow*). Participants showed an equal likelihood of fixating the competitor for the dominant or the non-dominant competitor. Participants appeared to map phonological wordforms to their visual/semantic representations on the fly. A second experiment altered the presentation to encourage more direct phonological encoding, by masking the visual stimuli. Here, participants showed a dominant-name effect. This suggests that prenaming may only come into play when there is visual difficulty (which is not the case during the typical long previews). Listeners don't appear to explicitly prename objects unless task demands specifically necessitate it (and when they are forced to prename, the effects are substantial).

But preview does affect fixation patterns

Although lexical access is not completely limited to pictured options, preview may still affect lexical processing in more nuanced ways. Several studies have investigated the effect of preview duration on fixations. These studies interpret changes in fixations as evidence of changes in the underlying lexical activation process; however, given that eye-movements may also support things like visual search, this is not clear.

Chen and Mirman (2015) conducted a VWP study of semantic processing (e.g., fixations to *key* after hearing *lock*) and found increased fixations to semantically related objects with longer preview. Preview duration interacted with phonological neighborhood density to reveal complex dynamics of semantic activation. Chen and Mirman hypothesized that when participants had sufficient time to preview the scene, their ability to access the semantic representations of the displayed items increased, leading to interactive boosts of these lexical entries.

Yee et al. (2011) investigated the nature of this semantic preactivation more deeply (see also, De Groot et al., 2016). They showed that with a moderate preview duration (1000 msec), items with similar conceptual shapes (e.g., *Frisbee* and *pizza*) showed competition in eye movements, whereas items with similar functions (e.g., *tape* and *glue*) did not. However, with a longer preview duration (2000 msec), function competitors showed competition. These data again suggest that visual preview boosts access of the semantic representations of the stimuli, and that this semantic pre-activation exhibits complex dynamics; it seems that the longer the participant has to preview the images, the deeper the semantic activation occurs.

These studies are consistent with the notion that preview leads to pre-activation of *something*. However, these findings show evidence of changes to *semantic* processes. This is precisely what preview hopes to accomplish in studies of word recognition, as it allows listeners to visually identify semantic features in the world and bind them to spatial

locations before lexical processing begins. Less clear is whether preview affects phonological or lexical processing.

The only study to address that is by Huettig and McQueen (2007). They manipulated preview duration and measured fixations to semantic, visual and phonological competitors. Their study included trials with no target but all three competitors, and no explicit response. When stimuli were present from the beginning of a carrier sentence (a relatively long preview before the target word), looks to phonological competitors preceded looks to other competitors. However, with a short (200 msec) preview, looks to the phonological competitor were slightly delayed relative to looks to visually similar competitors. They interpreted these findings as indicating that longer preview provides more time to activate phonological names of the displayed objects, and these names then bias looks when the target is heard. When preview is short, participants must activate the lexical items entirely from bottom-up input, and so don't receive the biasing benefit.

This study seems to indicate phonological preactivation during preview if enough time is available. However, the contrast between long and short preview does not sufficiently isolate what stimulus preview is doing. The differences between timing conditions may reflect semantic processing and location binding. For example, if participants did not know the location of the activated semantic features, they may waste a fixation or two looking for those features, or be forced to delay initial fixations until features are identified. This could result in apparent delay in competitor activation (since the earliest fixations will be at chance) even if lexical processing is unchanged. It may also explain why fixations to phonological competitors looked like fixations to visual competitors (which are largely driven by visual recognition and search processes).

Given these concerns, simply manipulating the duration of the preview is insufficient to determine its role. Rather we must systematically untangle the variety of processes that might occur during this time to understand when different processes occur and how they interact. The VWP requires a complex interplay of dynamic linguistic and non-linguistic processes; we need to understand how and when these processes occur.

The present study: Possible mechanisms of preview effects

Though there is not yet strong empirical support for prenaming, such concerns potentially challenge the VWP as a straightforward measure of lexical processing. However, an alternative (and more standard) line of thinking suggests that preview *enhances* the VWP's validity by isolating the measure of lexical processing from effects of object recognition and visual search. Without it, listeners must simultaneously process the target word, identify the semantic features of the images and locate them in space. This might require sequentially fixating each item in the display to identify the features *while lexical access is ongoing*. In this case, the earliest fixations may be noisy and not differentiated based on lexical processing. Even if the visual features are extracted using peripheral vision, semantic recognition still takes time, and may delay stimulus-relevant fixations at the start of a trial. For example, if the cohort was active, but the listener didn't know where to locate its visual features, they may be equally likely to fixate all objects. As a result, the fixations may not cleanly reflect lexical processing (e.g., be directed to the cohort and target more than the other items) until later in the trial, when processing is complete. This assumption that preview enhances the validity of the VWP also lacks empirical support. What is needed is a closer empirical look at the various cognitive processes that are potentially at play during this ecological—yet complex—language processing task.

In principle, preview could have several effects, reflecting the different processes that must operate within the VWP. First, evidence of semantic preactivation (Chen & Mirman, 2015; Yee et al., 2011) suggests that preview helps participants identify semantic features, so this need not occur during lexical processing. Second, preview may allow participants to bind these semantic representations to their locations. This would minimize visual search demands during word recognition. Both processes can make the earliest fixations more precisely reflect lexical processes, making the VWP more sensitive.

A third possible effect is that preview elicits *phonological* activation of the four items (either "prenaming" or as a form of priming or inhibition). This could alter patterns of phonological competition, artificially inflating competitor activation or, in the extreme, creating the sort of closed set task discussed earlier. Finally, these preactivated word forms may be bound to their locations which could change visual search. More than one of these processes could occur in tandem and to various degrees. Direct empirical evidence differentiating these factors is needed to disentangle these processes.

To investigate which of these effects arise from preview, we conducted a standard VWP word recognition experiment examining competition between targets and cohort competitors. The nature of the preview was manipulated between-participants with a series of conditions to capture the mechanisms outlined above.

The *Self-paced* condition used a preview in which participants saw the pictures in their correct locations until they self-initiated the stimulus (standard in the McMurray lab). This gave them as much preview as they like. However, the experimental preview conditions (described below) required a fixed preview duration (more standard in other labs). Thus, we developed a second baseline (the *Visual-Same locations* condition) in which images are present for a fixed time (1500 msec). This equates overall preview time to the other experimental conditions in which participants do not self-cue the auditory stimulus.

These typical forms of preview were contrasted against a *No preview* condition to identify any overall effects of preview as a first pass at confirming that preview matters in some way. Even without preview, listeners clearly locate visuo-semantic features in space, and could be covertly naming them as well. However, these events take time (the time, for example, between fixating an object and producing its name can be several hundred msec: Griffin, 2004); thus, these effects should be less pronounced than with full preview. Consequently, the first fixations in this condition may be relatively undifferentiated by lexical processes since there is no internal feature map to guide them.

We then considered two new conditions to isolate some of the cognitive processes during preview. First, in the *Text preview* condition the response options are shown orthographically during the preview, but in different locations than their corresponding pictures during the trial. This gave participants an explicit preview of the words, absent the visual-semantic form and the locations. Participants likely do activate the semantic representation from the text. However, our expectation is that text more directly primes a specific phonological form (e.g., *couch* not *sofa*), and that the semantic representation is not as specifically tied to the visual features that will later be on the screen (as it would be if we just showed those pictures). Thus, the text condition should show more phonological prenaming than visuo-semantic activation. Critically, since the words are not in the correct locations, this form of preview would not help with search processes.

Second, in the *Visual-New locations* condition, the pictures were shown during preview, but in different locations than during the trial. This allowed participants to preactivate semantic features, but not bind them to locations. Again, participants could still activate the phonological forms from the pictures, but this route is less direct than with text preview (Huettig & McQueen, 2007), and it may be less specific (multiple words could be active for any picture). Differences between this condition and the *Text preview* condition would reveal whether preview that is more weighted to words or visual-semantic features has different effects. Differences between this and the *Visual-Same locations* condition could reveal the contribution of binding visual features to locations in advance of lexical access, as both provide the same visual-semantic information.

These comparisons taken together can provide insight into the various processes potentially at play and reveal which seem to have particular effects during preview.

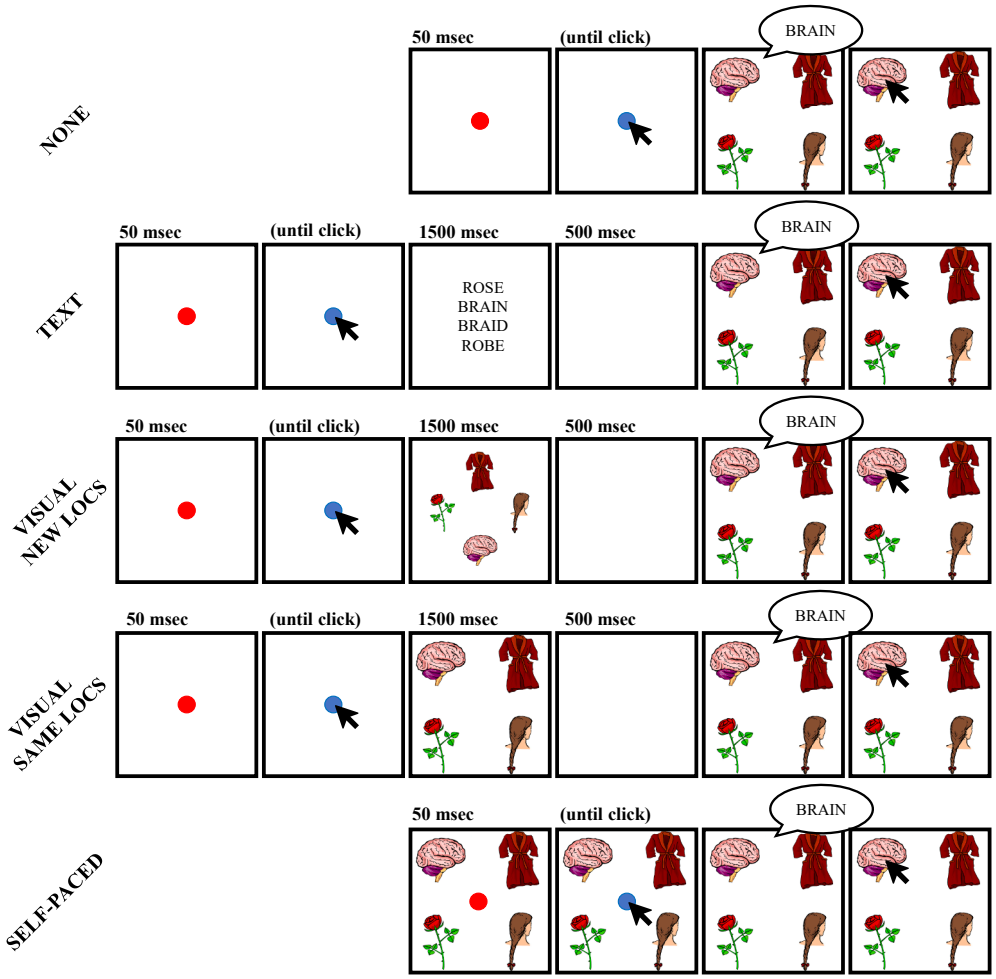

Fig. 1. Schematics of the preview conditions.

Methods

Participants

Participants were 122 monolingual English speakers with no reported vision or hearing deficits. Participants provided informed consent and were compensated with course credit or a small payment. Three were excluded: one because of technical issues, and two because of poor eye tracks. Participants were assigned to the five preview conditions approximately equally (*No-preview:* $N = 24$; *Text:* $N = 24$; *Visual-New locations:* $N = 27$; *Visual-Same locations:* $N = 22$; *Standard:* $N = 22$). Sample sizes were not determined by *a priori* power analyses, but instead were chosen to approximate typical sample sizes for similar VWP studies published around the time when data collection began (in 2013). Recruiting was intended to reach approximately 25 people per condition; variation in sign-ups, cancellations, and no-shows led to uneven final samples per condition.

The lack of *a priori* power analyses is a limitation of this study. We computed the post-hoc sensitivity of a study with this sample size for the critical analysis, comparing the proportion of looks to cohort and unrelated items by condition in a 2 (within) by 5 (between) ANOVA. This revealed sensitivity to detect an effect size of $d = .403$, or $\eta_p^2 = .039$. These constitute fairly small effect sizes, suggesting that our design has relatively high power to detect differences in the degree of competitor effects between conditions.

Participants were initially randomly assigned to the first three preview conditions (*No-preview, Text, Visual-New locations*). Upon completion of data collection for these conditions, the last two conditions (*Visual-Same locations* and *Self-paced*) were identified as crucial comparisons and new groups were recruited.

Design

Items consisted of 24 pairs of monosyllabic words that overlapped in initial consonants and vowel but differed in offset consonant (cohorts; e.g. *brain* and *braid*). A presentation set (the four items in a display) consisted of two cohort pairs with minimal phonological and semantic overlap. This design ensured that the presence of a cohort competitor could not serve as a cue to the target, as every item in the display had one cohort competitor; and it allowed the unrelated items in a trial to serve as targets or cohorts on other trials. Each item from each set was the target in four trials. This produced 12 sets × 4 items/set × 4 repetitions = 192 trials per participant. Trial order was random. Two separate pairings were developed, and these were counterbalanced between participants (Appendix A).

Participants were assigned to one of five preview conditions. These conditions differed in the sequence of events before the target auditory stimulus was presented (Fig. 1) but were otherwise identical: four pictures were displayed; an auditory stimulus was played; the participant clicked on a picture; the screen went blank; and the next trial began 300 msec later.

Initially, we planned this study as a within-participants design with preview conditions in blocks. However, during data collection (but before analysis), we realized that such a design is problematic, as the first preview conditions could alter processing for later conditions. For example, an initial block of text preview might encourage explicit

prenaming in later blocks. Or, an initial block of trials with visual preview could help preactivate the pictures from memory in later blocks. Thus, we switched to a between-participants design. For participants ($N = 75$) who had already completed the within-participant study, we only considered the first block of trials (the 192-trial design described above), which was counterbalanced to include a single condition that varied between participants. The remaining two 192-trial blocks were discarded. Later participants ($N = 47$) only completed a single block in one condition.

Stimuli

Visual stimuli were color drawings. Development of these images followed a standard procedure to ensure participants would readily recognize them as the intended word (McMurray et al., 2010). First, several potential images for the word were selected from a commercial clipart database. A focus group of four to six people selected the image that best represents the word, while adhering to a similar style as other images in the study. They also recommended changes to ensure size, brightness and complexity conformed with other study images, and to remove distracting elements and backgrounds. After these edits, a senior lab member with experience in the VWP who was uninvolved with the stimulus development approved the image.

Each target word was recorded by a male monolingual speaker of English in a sound-attenuated room at 44.1 kHz. Each word was recorded 3–4 times. An exemplar with a neutral pitch and free of artifacts was selected for use. All selected exemplars were amplitude normalized in Praat, and 100 msec of silence was appended to the beginning and end of the file. Visual and auditory stimuli are available on the OSF site for this project (https://osf.io/b7q65/).

Eye-movement recording and processing

Eye movements were recorded using an SR Research Eyelink II head-mounted eye-tracker, tracking at 250 Hz. At the beginning of the study, the standard nine-point calibration procedure was conducted, and drift correction was performed every 24 trials. Fixations were automatically parsed into saccades, fixations and blinks using the default parameters for the tracker. Saccades and fixations were combined into "looks," defined from the start of saccade onset until the end of a fixation. In assigning looks to objects, boundaries of the objects were extended by 100 pixels. This did not result in any overlap in regions of interest.

Eye movements initiated prior to the auditory stimulus were discarded from analysis, as these could not be driven by lexical information. Trials ended when a mouse-click response was registered. To deal with the fact that trials had different length, we used a form of "object padding" in which the final fixation was extended to a fixed length of 2000 msec. Consequently, late in the trial, the fixation curves reflect something akin to the asymptotic decision. However, fixations were only considered until 1300 msec post-stimulus-onset. This endpoint was chosen as the mean RT across all conditions was 1264 msec, and the slowest condition (Text) had a mean RT of 1307 msec; 1300 msec thus covers a time window in which most participants should have already responded on most trials. Note that after this time point, looks were extremely stable, and competitor effects were near zero in all conditions.

Procedure and conditions

The details of each trial differed depending on preview condition (Fig. 1).

No Preview. In the No preview condition participants saw a blank screen, with a red dot in the center. After 50 msec, the dot turned blue. When the dot was clicked, it disappeared, and the trial began. This condition thus gave no information about the possible words, the visual features of the pictures, or their locations.

Text Preview. Text preview of the response options should enhance the likelihood that participants prename the phonological forms, and eliminate ambiguity about the specific word corresponding to each response (unlike an image which could be named in multiple ways). Huettig and McQueen (2007) demonstrated that text presentation in the VWP leads to more rapid activation of phonological forms than visual presentation. With text, participants made earlier fixations to phonological competitors than semantic competitors. This accords with models of word reading (Coltheart et al., 2001; Plaut et al., 1996), which include direct links between orthography and phonology. Thus, a text preview provides the most direct manipulation of phonological prenaming; we provide participants with the words themselves, without any visual feature or location information.

In the Text preview condition, participants saw a blank screen with a red dot, which turned blue after 50 msec. Upon clicking the dot, the dot disappeared, and the names of the four pictures were presented in a column in the center of the screen. The words remained on the screen for 1500 msec and were then removed for 500 msec before the trial began.

During preview, words were presented in a random order, so they gave no indication of the location of the correct target; this differs from the text condition in Huettig and McQueen (2007), where the words were in the same locations as the responses. Preview thus provided the possible words the participant could hear, but not the visual form of the responses, nor their locations. A comparison between this condition and the No preview condition indicates the extent to which knowing the set of possible target words in advance affects looking in the VWP.

Visual-New locations. In this condition, images appeared during the preview, but in different locations than during the trial. This provides information about the visual features, and potentially allows phonological prenaming, but does not help with visual search. By comparing this to the Text preview, we can ask whether showing the pictures in advance is equivalent to providing the text. If preview effects arise because the images drive phonological preactivation, then showing images and providing text should be quite similar. Alternatively, if visual preview encourages identification of visual-semantic features, but not necessarily phonological coding (see also, Pontillo et al., 2015), this predicts different impacts of preview in these two conditions, as they emphasize different processes necessary for the VWP.

In this condition, participants saw a screen with a red dot that turned blue after 50 msec. After clicking the dot, it disappeared, and the four pictures were displayed in a diamond configuration on the screen. Critically, this configuration differed from the configuration during the trial (pictures in the four corners of the screen). Pictures remained on the screen for 1500 msec. The screen then went blank for 500 msec before the trial began. The location of the pictures in the diamond preview display was randomized so that the preview offered no information about the correct target location.

Visual-Same locations. The prior condition exposed participants to the images but did not reduce visual search demands because their ultimate locations were not the same. These search demands are a focus of arguments in support of preview before the VWP – if participants need to find the target, their fixations will be a combination of lexical activation and visual search. Thus, the Visual-Same locations condition provided the preview images in the locations in which they would appear during the trial. This condition thus provided participants with the visual-semantic features; the location information (obviating the need for visual search during the trial); and possibly the ability to prename items, if this naturally occurs during preview. Comparison of this condition with the Visual-New locations condition provides critical information about the extent to which visual search dynamics alter fixation patterns.

Self-paced. The Visual-Same locations condition mirrors the approach used in many VWP studies. However, some designs (including most used by our lab) also allow participants to cue the auditory stimulus, with the preview available until they begin the trial. This provides whatever time participants require to overcome individual differences in search speed, object recognition, and so forth. However, two differences between this condition and the other conditions in the study could have important

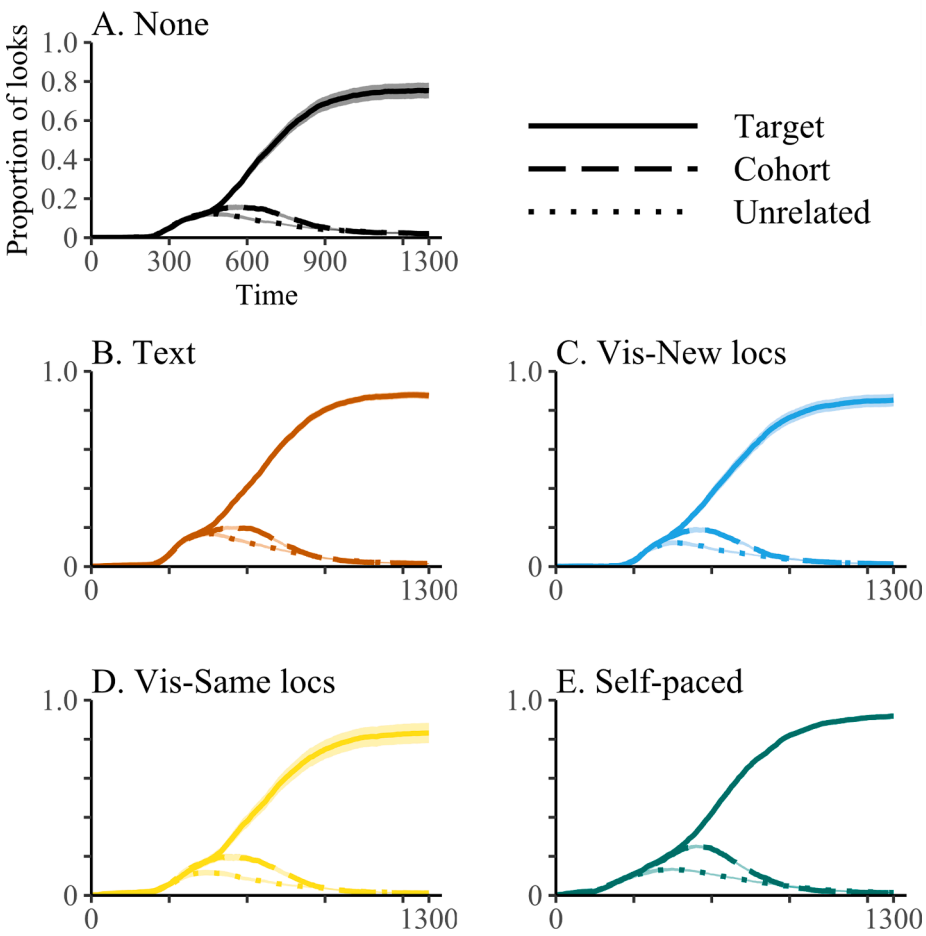

Fig. 2. Proportion of looks to the displayed objects over time in each preview condition. Time is indexed from the onset of the auditory stimulus presenting the target. Note the Unrelated lines display the mean proportion of looks to the two unrelated objects. Error ribbons signify the standard error of the mean at each time sample. A) No preview. B) Text preview. C) Visual-new locations. D) Visual-same locations. E) Self-paced.

effects.

First, in the self-paced version, listeners self-cue the auditory stimulus. This provides more certainty of when the word will arrive than in the other conditions, where the word is played after a long delay. This could particularly affect early eye-movements. Second, the self-determined preview duration could lead to different levels of semantic processing (Chen & Mirman, 2015; Yee et al., 2011), or could encourage explicit prenaming, if participants opt to wait long enough (Huettig et al., 2011). As such, the contrast between the Self-paced condition and the Visual-Same locations condition, with a constant preview, can show whether untimed preview might impact looking behavior. In the Self-paced condition, at trial onset, all four images were displayed, along with the central red dot. This dot then turned blue, and participants clicked the dot to initiate the trial. The images remained on the screen throughout this process.

Results

Approach

Our analysis consists of two major sections. First, we descriptively assess the broad patterns of fixations within each preview condition to identify high-level differences, and conduct omnibus tests across all conditions to establish overall main effects and potentially relevant differences. Second, we conduct pairwise comparisons between targeted conditions to examine how different aspects of the preview period impact fixations. These analyses examine the timing and extent of target looks, as well as the degree of competitor consideration.

This study does not use a true factorial design, but a sequence of planned contrasts to determine how specific aspects of preview impact fixation patterns. While we report omnibus tests, the focal analyses are pairwise comparisons of conditions to isolate specific hypotheses about the role of preview. We highlight five primary comparisons:

(1) *No Preview* vs. *Self-paced*. This establishes whether preview has any effect on performance, and whether competition effects emerge without preview. However, differences between these conditions could arise because of preview affecting several aspects of processing.
(2) *No Preview* vs. *Text*. This contrasts a condition in which participants have no information prior to the target word, with one in which they are provided the possible wordforms in written format. This comparison assesses how providing more access to specific phonological forms affects fixations relative to conditions with no preview. If preview impacts competitor effects because of earlier activation of phonological forms, then text-based preview should be particularly impactful.
(3) *Text* vs. *Visual-New locations*. This asks whether highlighting the phonological forms through a text preview differs from providing the visual objects, which more directly and specifically activate the visual-semantic features that will be needed later to direct fixations (note that we acknowledge text could activate semantics, and pictures could activate words, just less so than the converse). If preview helps participants by providing access to

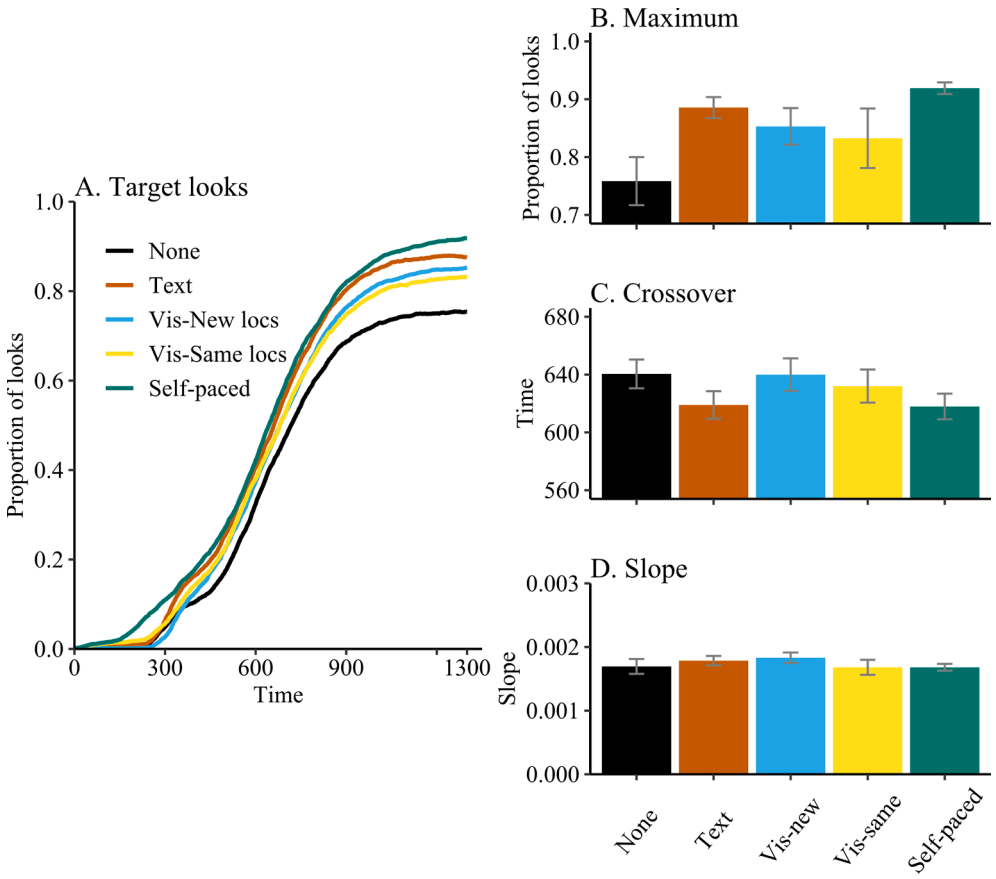

Fig. 3. Timecourse of fixations to target objects by condition, and curvefit parameters for these curves. The overall timecourse plots the raw fixation data. The individual paramaters plot the curvefit values. A) Timecourse of target looks (raw data). B) Curvefit maximum parameters. C) Curvefit crossover parameters. D) Curvefit slope parameters.

visual-semantic information, we would expect the Visual-New locations condition to elicit more robust lexically driven competition than the Text condition. If preview effects are more due to phonological preactivation, we would expect the Text condition, which provides more direct access to phonological forms, to lead to greater competitor effects.

(4) *Visual-New locations* vs. *Visual-Same locations*. This comparison contrasts a case where visual search must be completed during the trial with one in which spatial location information is available before search begins (but visual-semantic features are available in both). Differences in fixation patterns between these conditions would index how the ability to locate the semantic features in space prior to lexical access affects looking behavior.

(5) *Visual-Same locations* vs. *Self-paced*. This comparison includes cases where the preview provides the images in their correct locations. However, the Self-paced condition allows participant control of presentation timing. This comparison thus assesses how preview time and expectations about stimulus timing affect looking behavior.

Descriptive results

Data are available on the OSF page for this project (https://osf.io/b7q65/). Analyses of fixations considered only trials when the correct referent was selected. Given the ease of the task, accuracy was extremely high (mean = 99.6%). All conditions showed mean accuracy over 99%, and no participant performed worse than 96.3% correct (7 incorrect trials out of 192). For these analyses, unrelated looks are presented as the mean proportion of looks to the two unrelated items, as there is only a single target and cohort on each trial.

Fig. 2 shows mean proportion of looks to each item type over time for the five preview conditions. Figs. 3 and 4 compare fixations across conditions for the target (Fig. 3) and competitors (Fig. 4), Several aspects of these curves suggest complex effects of preview. We break down these differences descriptively first, and then proceed to statistical analyses.

First, all conditions show rapid separation of target looks from other objects, and all show greater cohort fixations than unrelateds. Despite the preview differences, participants fixated objects consistent with the phonological form of the word, and showed incremental processing as cohort fixations returned to baseline (e.g., the unrelated object) over about one second.

Second, in the No preview condition, the asymptotic level of target fixations at the end of the trial is substantially lower (Fig. 2A, 3) than in the other conditions, and the cohort and unrelated objects continue to receive looks even late in the trial (Fig. 4A). Note this occurs despite participants choosing the correct object and overall accuracy over 99%. A lack of preview noticeably alters fixations.

Third, both the No Preview (Fig. 2A) and the Text (2B) conditions show delayed cohort fixations. In these conditions, all three object types are fixated a similar amount during the first 500 msec, after which *both* cohorts and unrelateds show little increase and fall off. This differs substantially from the other preview conditions, in which the target and cohort separate from the unrelateds, and the unrelated looks appear to drop off earlier (typically at 250–350 msec). These latter patterns more closely match typical theories of incremental processing during lexical access (McClelland & Elman, 1986; Norris, 1994); listeners are expected to activate targets and cohorts over unrelated words initially, and then suppress cohorts once disambiguating information arrives. The No Preview and Text conditions may result in inflated fixations to unrelated

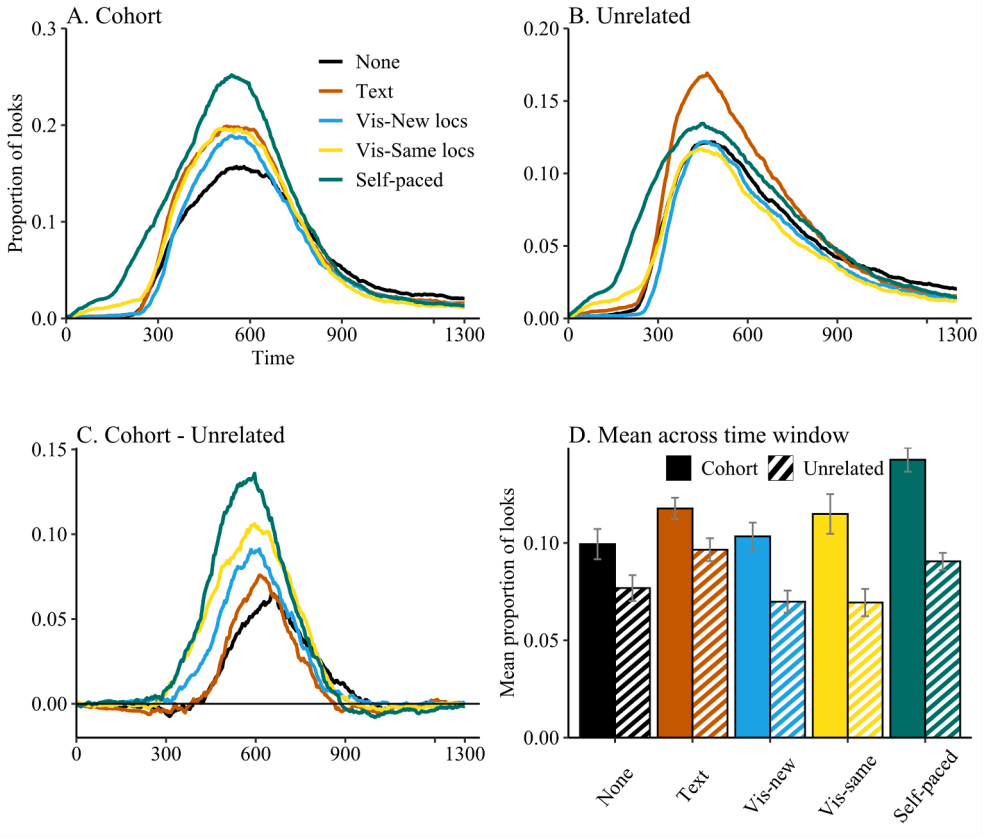

Fig. 4. Timecourse of fixations to non-target objects. A) Fixations to cohort objects. B) Fixations to unrelated objects. Plots the mean of the two unrelated objects. C) The difference between cohort and the mean of the unrelated objects. This panel represents the degree of cohort fixation over and above looks to unrelated objects. D) The mean proportion of looks across time to cohort and unrelated items over the time-window 250–1000 msec.

objects, making it difficult to observe incremental activation.

Finally, the Self-paced condition shows overall greater looking to all objects early in the trial. This difference is interesting given that it typically takes 200 msec to plan and launch an eye-movement. Thus, the early fixations to all objects likely arise before auditory information is available. Participant control of trial onset may result in a greater likelihood of launching a non-specific fixation initially, rather than waiting for lexical access to start biasing fixations.

Omnibus analyses

Targets

To investigate the effects of the preview on fixations to the target, we first fit a four-parameter logistic to the target fixations for each participant (Farris-Trimble & McMurray, 2013). This provides meaningful estimates of specific properties of the fixation curves: the *minimum* measures the initial baseline level of looking; the *crossover* measures the midpoint of the rise, and the *slope* measures the speed of this rise, providing two indices of the timing of target identification; and the *maximum* measures the peak (asymptotic) looking. These parameters have proven meaningful: across development, the slope and crossover points become faster (Rigler et al., 2015), and people with language disorders show changes in maxima (McMurray et al., 2010). Our analysis approach ignores eye movements before the auditory stimulus, so minima are by definition quite close to zero; we thus don't consider them further.

Fits were performed using a constrained gradient descent technique that minimized the least squared error between the estimated function and the data (McMurray, 2020). Each fit was manually checked against the data and refit using hand-selected starting parameters if necessary. The fitted curves matched the underlying data quite strongly, with a mean fit of $r = .998$ (SD = .0017); all curves had a fit of at least $r = .990$. No fit was discarded for a poor fit.

Fig. 3 shows the target looks (Panel A) and mean curvefit parameters (Panels B-D) by condition (for a version with error bands in Panel A, see https://osf.io/b7q65/). We ran separate ANOVAs for each parameter, with preview condition as the IV. The crossover showed no effect of condition by participant ($F_1(4,114) = 1.13$, $p = .34$, $\eta^2_{ges} = .038$, though it was significant by items $F_2(4,188) = 4.81$, $p = .001$, $\eta^2_{ges} = .011$); slope was also not affected by condition ($F_1(4,114) = .60$, $p = .67$, $\eta^2_{ges} = .020$; $F_2(4,188) = .67$, $p = .62$, $\eta^2_{ges} = .004$), suggesting similar speed and timing of target fixations between conditions. However, there was a significant effect of condition on the maximum ($F_1(4,114) = 3.15$, $p = .017$, $\eta^2_{ges} = .10$; $F_2(4,188) = 154.26$, $p < .00001$, $\eta^2_{ges} = .69$). This effect indicates that the different conditions yielded different peaks, despite similar timing.

Competitors

The degree of competitor fixations was assessed by comparing the proportion of fixations to the cohort to the mean proportion of fixations to the two unrelated objects (Fig. 4; for a version with error bands in Panel A, see https://osf.io/b7q65/). This relative measure accounts for the potential that some participants or conditions may show greater overall looking independent of object identity, and thus isolates the contribution of phonological similarity. Fig. 4C visualizes this in terms of the cohort minus unrelated difference over time.

For analysis, we averaged the proportion of fixations to the cohort and the two unrelated objects for each participant from 250 to 1000 msec post-stimulus. This window includes the times when a competitor effect is seen in all conditions in Fig. 4C, and thus is appropriate to capture the full degree of competitor-driven looking behavior. These proportions were entered into a 5 (condition, between participants) × 2

Table 1
t-tests comparing proportion of looks to cohort objects and to the average of the two unrelated objects from 250 to 1000 msec in each condition. Displayed as By participants / By items. All $p < .005$.

Condition	df_1 / df_2	t_1 / t_2
None	23 / 47	7.30 / 3.21
Text	23 / 47	5.73 / 3.44
Vis-New locs	26 / 47	9.51 / 5.79
Vis-Same locs	21 / 47	9.32 / 7.55
Self-paced	21 / 47	13.17 / 6.74

(object type, within participants) ANOVA.

There was a significant main effect of condition ($F_1(4,114) = 4.053$, $p = .004$, $\eta^2_{ges} = .12$; $F_2(4,188) = 71.48$, $p < .00001$, $\eta^2_{ges} = .096$), signifying overall differences in fixations to the non-target objects between groups. There was also a main effect of object type ($F_1(1,114) = 415.12$, $p < .00001$, $\eta^2_{ges} = .23$; $F_2(1,47) = 35.18$, $p < .00001$, $\eta^2_{ges} = .19$), signifying greater cohort than unrelated fixations. There was a significant interaction ($F_1(4,114) = 12.36$, $p < .00001$, $\eta^2_{ges} = .034$; $F_2(4,188) = 17.35$, $p < .00001$, $\eta^2_{ges} = .028$), indicating differences in the degree of cohort relative to unrelated fixations between conditions.

Follow-up analyses within each condition showed a significant difference between cohort and unrelated fixations in every condition (Table 1). Changes in preview did not eliminate cohort effects, though there were changes in the extent of these effects. Critically, even when no preview was provided, cohort looks exceeded unrelated looks, arguing against the strongest claim that previous VWP competitor effects only arise from prenaming. To further characterize these differences and determine what aspects of looking behavior are impacted by preview, we next turn to planned pairwise comparisons between contrasts of particular interest.

Pairwise comparisons for testing hypotheses

Pairwise comparisons took the form of simple-effects comparisons focused on the planned contrasts described above. These comparisons can reveal whether fixation patterns to targets or degree of competitor fixation differed depending on the preview.

No preview vs. Self-paced

We first asked whether a lack of preview alters fixations relative to Self-paced preview. This comparison conflates all possible preview effects (phonological prenaming, semantic feature identification, visual search and control of stimulus timing) to establish whether more nuanced comparisons are needed.

First, we compared parameters of the target. We found no significant effects for the crossover ($t_1(44) = 1.68$, $p = .10$; though it was by items $t_2(47) = 3.73$, $p = .00052$), nor for slope ($t_1(44) = .10$, $p = .92$; $t_2(47) = .307$, $p = .76$), in line with the lack of an omnibus effect of condition for these variables. However, there was a significant effect for maximum ($t_1(44) = -3.61$, $p = .00077$; $t_2(47) = -19.50$, $p < .00001$), as the Self-paced condition showed a higher maximum ($M = .92$) than No preview ($M = .76$). The lack of preview led to lower overall target fixations, despite high accuracy and the inclusion of only correct trials.

Next, we considered competitor looking. As in the omnibus analysis, we used a 2 × 2 ANOVA with timing condition and object type (cohort vs. unrelated) as factors. The DV was the proportion of looks between 250 msec and 1000 msec. This revealed main effects of condition (Self-paced > No preview; $F_1(1,44) = 10.58$, $p = .002$, $\eta^2_{ges} = .18$; $F_2(1,47) = 163.14$, $p < .00001$, $\eta^2_{ges} = .12$) and object type (cohort > unrelated; $F_1(1,44) = 225.34$, $p < .00001$, $\eta^2_{ges} = .28$; $F_2(1,47) = 27.76$, $p < .00001$, $\eta^2_{ges} = .19$). There was also a significant interaction ($F_1(1,44) = 35.37$, $p < .00001$, $\eta^2_{ges} = .057$; $F_2(1,47) = 50.22$, $p < .00001$, $\eta^2_{ges} = .036$) due to significantly greater competitor effects (a larger difference between cohorts and unrelateds) in the Self-paced condition than in the No preview condition (Fig. 4, black vs. green lines).

These analyses establish substantial differences in the amount and timing of looking to different competitors as a function of preview. Critically, with preview, both targets and phonological competitors receive more looks, and the differentiation of cohort from unrelated objects is enhanced. Including a stimulus preview is clearly doing *something*. However, these differences could be driven by a variety of factors including phonological prenaming, visual feature identification, and visual search which are examined in the next comparisons.

No preview vs. Text

The contrast between No preview and Text preview asks if there is an effect of a preview which provides more efficient and unambiguous access to the wordforms of the responses, without directly previewing visual-semantic information, and providing no information about location. This comparison isolates an effect of prenaming in the absence of other information – if participants are told the possible wordforms, they should easily access the phonological forms of the words and this condition should elicit particularly strong effects.

A series of t-tests comparing the target curves showed no effect for crossover ($t_1(46) = 1.56$, $p = .13$; though it was significant by items $t_2(47) = 4.42$, $p = .000057$) nor slope ($t_1(46) = -.66$, $p = .51$; $t_2(47) = -.70$, $p = .49$), mirroring the omnibus analysis. However, there was an effect for maximum ($t_1(46) = -2.81$, $p = .0073$; $t_2(47) = -16.9$, $p < .00001$), as the Text condition ($M = .89$) reached a higher peak than the No preview condition ($M = .76$). Providing the wordforms before the trial led to heightened target fixations.

Competitor effects were examined with a 2 (item type: Cohort vs. Unrelated) × 2 (condition: No preview vs. Text) ANOVA, using the proportion of looking in the 250–1000 msec window. This analysis showed main effects of item type (cohort > unrelated; $F_1(1,46) = 82.55$, $p < .00001$, $\eta^2_{ges} = .11$; $F_2(1,47) = 12.36$, $p = .00098$, $\eta^2_{ges} = .087$) and condition (Text > No preview; $F_1(1,46) = 4.57$, $p = .038$, $\eta^2_{ges} = .085$; $F_2(1,47) = 139.89$, $p < .00001$, $\eta^2_{ges} = .066$). However, there was no interaction ($F_1(1,46) = .085$, $p = .77$, $\eta^2_{ges} = .00013$; $F_2(1,47) = .12$, $p = .74$, $\eta^2_{ges} = .00011$), signifying similar degrees of competitor effects between the two conditions (Fig. 4). This lack of interaction indicates that explicitly providing the wordforms to the participants before the trial did not lead to increased competitor effects relative to providing no preview at all. It did increase fixations more generally, but this was the case for cohorts *and* unrelateds (as well as targets).

These results suggest that preview of the words (but not the images or locations) yielded more looks to *all* objects (starting early and lasting well into the auditory stimulus). Despite knowing what words are possible targets, in the text preview condition, participants look more even to unrelated items than if they did not know the possible words (Fig. 4B. orange vs. black lines). More damning for the prenaming account is the limited difference in competitor effects between this condition and the No preview condition. The overall analysis showed no difference in competitor effects (the cohort – unrelated, Fig. 4C) between these conditions. That is, competitor effects do not increase when the text of the words is directly provided during preview.

Text preview provides the most direct, unambiguous access to phonological forms; participants need not activate names via the images, and are told exactly what the names of the objects are. Nevertheless, the Text condition did not yield consistently larger competitor effects. If competitor effects are partially driven by preactivation of phonological forms, these conditions should show differences. Instead, competitor effects proved similar whether or not participants were told the wordforms, and even unrelated fixations were affected, suggesting that providing wordforms without their visual realizations may have raised visual search demands.

Text vs. Visual-New locations

The small differences in competitor effects between the No preview and Text conditions suggest that preview might do something other than

elicit prenaming. Next, we asked how adding visual-semantic information affects fixation patterns in the Visual-New locations condition, which previews response images, but in different locations. Whereas the Text condition more directly links to phonological representations and explicitly provides the wordforms, this condition links more directly to the visual-semantic information about the responses before the trial begins. If preview effects arise because typical preview cues participants to the possible phonological items, then these conditions might look quite similar. Alternatively, the preview of visual forms could reduce the need to identify the objects at each location during the trial. Location information is not available in either condition, so search demands should be similar in the two conditions. This contrast thus examines whether earlier access to visual information does something other than highlight what word might be upcoming.

Comparisons of target looks revealed no significant effects (maximum: $t_1(49) = .86$, $p = .39$; though it was by items $t_2(47) = 5.47$, $p < .00001$; crossover: $t_1(49) = -1.41$, $p = .16$; though it was by items $t_2(47) = -2.49$, $p = .016$; slope: $t_1(49) = -.42$, $p = .68$; $t_2(47) = -.49$, $p = .63$). These preview conditions led to extremely similar patterns of target fixations.

In contrast, an ANOVA of mean fixations to the cohort and unrelated objects in the 250–1000 msec time window revealed significant main effects of item type (cohort > unrelated; $F_1(1,49) = 114.60$, $p < .00001$, $\eta^2_{ges} = .17$; $F_2(1,47) = 21.75$, $p = .00003$, $\eta^2_{ges} = .15$), and condition (Text > Visual-New locations; $F_1(1,49) = 6.13$, $p = .017$, $\eta^2_{ges} = .10$; though not by item: $F_2(1,47) = .57$, $p = .46$, $\eta^2_{ges} = .00050$). Importantly, there was a significant interaction ($F_1(1,49) = 5.86$, $p = .019$, $\eta^2_{ges} = .010$; $F_2(1,47) = 5.47$, $p = .024$, $\eta^2_{ges} = .0060$), as the competitor effect (cohort-unrelated) was larger in the Visual-New locations condition than in the Text condition (Fig. 4C, teal vs. orange lines). Adding visual-semantic information in the Visual-New locations condition led to stronger competitor effects, despite more direct access to phonological forms in the Text condition.

These findings again argue against strong forms of phonological prenaming, and suggest that phonological pre-activation is not the sole (or even primary) effect of preview on subsequent stimulus-driven fixations. When participants are given the wordforms via text, they show *smaller* competitor effects than when shown the images. Although the visual preview could elicit pre-naming, it should do so less effectively than text; the images could be named in various ways, whereas the text names are unambiguous. Nonetheless, phonological competition is heightened for the visual preview, suggesting preview of the visual-semantic information before the trial (rather than the names) may allow participants to identify the available semantic features before the trial begins. This may remove variance from looking behavior based on needs to identify images, increasing sensitivity to effects of phonological processing.

Visual-New locations vs. Visual-Same locations

The preceding analysis suggests that reducing the need to identify visual-semantic features during word recognition allows a more direct measure of phonological competition. However, the Visual-New locations condition still requires visual search (to find those features) during the trial. The need to search for the semantic features may add unwanted variance. The Visual-Same locations condition counteracted this by presenting the images during preview in the locations where they would appear during the trial.

Analysis of target showed no significant differences (maximum: $t_1(47) = .35$, $p = .72$; though it was by items $t_2(47) = 2.98$, $p = .0046$; crossover: $t_1(47) = .49$, $p = .62$; $t_2(47) = .62$, $p = .54$; slope: $t_1(47) = 1.08$, $p = .28$; $t_2(47) = .68$, $p = .50$).

Competitor looks were analyzed in the same time window as previous analyses (250–1000 msec after auditory stimulus onset). There was a significant effect of item type (cohort > unrelated; $F_1(1,49) = 180.30$, $p < .00001$, $\eta^2_{ges} = .23$; $F_2(1,47) = 49.91$, $p < .00001$, $\eta^2_{ges} = .25$), but not condition ($F_1(1,49) = .29$, $p = .59$, $\eta^2_{ges} = .006$; though it was significant by item $F_2(1,47) = 6.02$, $p = .018$, $\eta^2_{ges} = .006$). However, the interaction was significant ($F_1(1,47) = 4.12$, $p = .048$, $\eta^2_{ges} = .007$; $F_2(1,47) = 9.54$, $p = .0030$, $\eta^2_{ges} = .008$), as the Visual-Same locations condition had an overall larger competitor effect than the Visual-New locations condition (Fig. 4C, teal vs. yellow lines). Reducing in-trial search demands (without any additional phonological information) enhanced the observed competitor effect.

These findings suggest that providing visual locations during preview further enhances competitor effects; importantly, these differences arise despite identical opportunity to prename, as the same visual images were presented (just in different orientations). The addition of location information during the preview increases sensitivity to phonological competition. This suggests that eliminating search demands during the trial allows fixations to more directly reflect phonological activation, but may also introduce some pre-stimulus noise in looking patterns.

Visual-Same locations vs. Self-paced

The final comparison examined the effect of changes in the triggering of the stimulus and the timing of the preview. When participants can trigger the auditory stimulus, they can process the images for as long as they like, and control of when the word is heard. This could impact the depth of semantic processing (Yee et al., 2011), or perhaps encourage phonological prenaming (Huettig et al., 2011). We thus compared the Visual-Same locations condition (with a fixed preview duration) to a Self-paced condition.

We first assessed the preview duration in the Self-paced condition. The mean duration[1] was 986 msec (SD = 135 msec). This was substantially *faster* than the Visual-Same locations preview condition (fixed 1500 msec preview), and all participants averaged faster than this condition (range: 830–1348 msec). Participants in the Self-paced condition thus received less preview time than in the Visual-Same locations condition.

Target parameters for these conditions revealed no significant differences (maximum: $t_1(42) = -1.65$, $p = .11$; though it was significant by items $t_2(47) = -14.2$, $p < .00001$; crossover: $t_1(42) = .97$, $p = .34$; $t_2(47) = 1.48$, $p = .15$; slope: $t_1(42) = -.001$, $p = .999$; $t_2(47) = .84$, $p < .40$).

The ANOVA for competitor effects revealed significant main effects of item type (cohort > unrelated; $F_1(1,42) = 241.44$, $p < .00001$, $\eta^2_{ges} = .35$; $F_2(1,47) = 55.18$, $p < .00001$, $\eta^2_{ges} = .29$), and of condition (Self-paced > Visual-Same locations; $F_1(1,42) = 6.31$, $p = .016$, $\eta^2_{ges} = .12$; $F_2(1,47) = 114.64$, $p < .00001$, $\eta^2_{ges} = .093$). The interaction was not significant ($F_1(1,42) = 1.14$, $p = .29$, $\eta^2_{ges} = .003$; $F_2(1,47) = 2.51$, $p = .12$, $\eta^2_{ges} = .002$), indicating a similar size of competitor effect for the two conditions.

Self-triggered trial onsets and a shorter preview duration in the Self-paced condition did not lead to overall changes in Target fixations or competitor effects. Some differences did arise – there were increased looks to both competitors *and* unrelateds in the Self-paced condition, and visual inspection of the timecourse curves suggests widespread increased looks early in trials (though these did not reach the level of significance in the timing parameters for the target curvefits). Allowing participants to initiate trial onset thus does not substantially impact competitor effects, but it may introduce noise to the fixation patterns early in a trial—participants are more likely to begin launching fixations to everything in the display before auditory input is heard. However, sensitivity to phonological competition is approximately unchanged.

[1] The first trial for many participants proved substantially longer than other trials (M=30.3 sec!) – some participants may have missed the instructions explaining how to self-initiate trials. As a result, the duration analyses ignored the first trial.

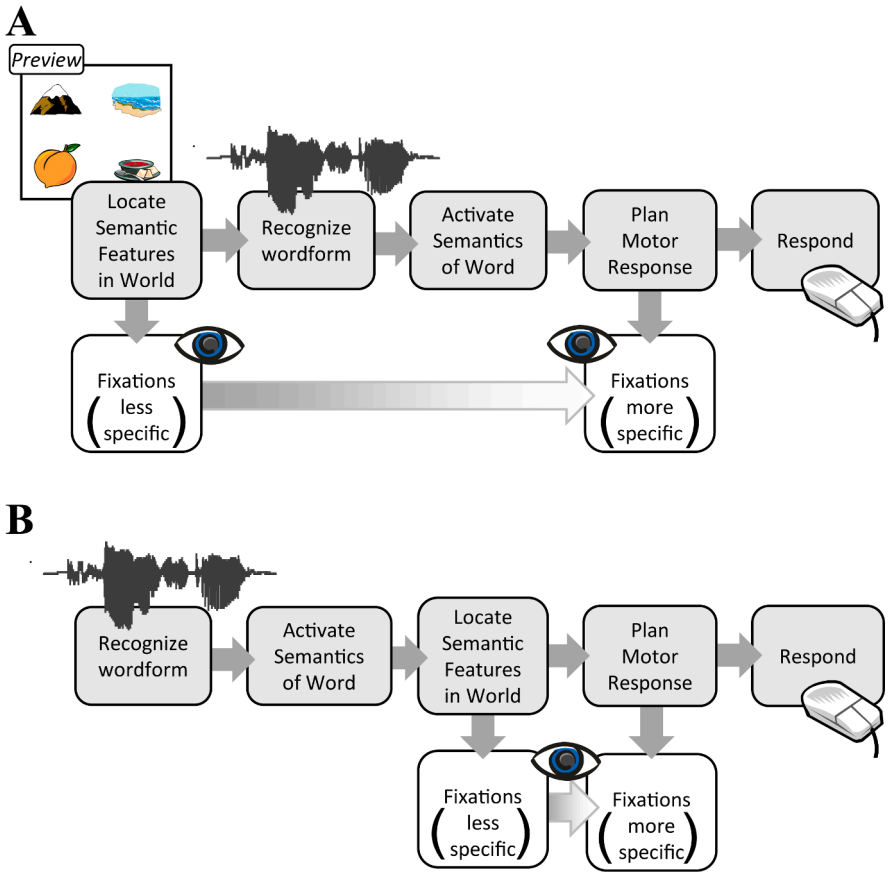

Fig. 5. Schematic of linking functions for the VWP with and without preview. A) With preview. B) Without preview.

Discussion

Typical VWP studies preview response alternatives prior to the onset of the word. Preview is intended to reduce the influence of object recognition and visual search on fixations during word recognition to offer a more sensitive measure of lexical processing. However, criticisms of the VWP often emphasize this preview period as a potential threat to the validity of the measure. Specifically, concerns about phonological prenaming or preactivation of the presented response options could limit whether VWP findings generalize to more unconstrained language processing contexts.

As described in the introduction, the most extreme forms of this critique have not held up to empirical evidence. However, prior studies have shown effects of preview duration suggesting a complex role for preview. We attempted to unpack a variety of processes that could occur during preview using a series of preview conditions that provide different types of information. This is intended to help elucidate the complex processes that give rise to eye movement behavior in the VWP. These analyses addressed several questions.

Do phonological competitor effects depend on preview?

The strong form of the argument against preview in the VWP is that competitor effects might arise solely because participants are cued to the identity of the possible words they might encounter, or even prename these words in verbal working memory and complete the task on these representations. Our results strongly reject this for several reasons. First, competitor effects were present regardless of preview condition–even in the No preview condition. Stimulus preview is not a prerequisite for observing competitor effects in the VWP.

Second, competitor effects were quite similar when we provided the written form of the possible response options during preview and when there was no preview, suggesting that specifically highlighting the names of the responses does not substantially alter competitor effects. Text preview offers easy and unambiguous access to the wordforms, whereas image preview (if it leads to prenaming) requires participants to identify the images correctly and access the intended phonological forms. This has greater ambiguity, since multiple words could be used for any image (Pontillo et al., 2015) and is slower (Huettig & McQueen, 2007). If competitor effects partially depend on prenaming, then Text preview is suited to boost this effect. Despite this, no such enhancement was found. The increased competitor effects seen when preview emerges do not seem to be based on access to the set of wordforms prior to the task.

These findings—along with the evidence for influences of non-displayed competitors (Dahan, Magnuson, Tanenhaus, et al., 2001; Kapnoula & McMurray, 2016; Magnuson et al., 2003)—strongly argue against concerns that the VWP only reflects activation within a closed-set, as well as the more subtle arguments that measured competition might be enhanced via priming or inhibition because words are being prenamed. Participants showed early sensitivity to phonological competitors from early points in time even when they had no idea what the responses would be until just as the auditory stimulus was presented, and they showed comparable consideration of competitors when they knew exactly what wordforms were possible. This fits with aforementioned work showing strong effects of cross-language competitors (Sarrett et al., submitted for publication; Spivey & Marian, 1999)—even though people are most likely naming the objects in one language, competition from the other language emerges. Although stimulus preview does affect patterns of looking, it is not the cause of competitor effects, nor is prenaming likely to be the most important aspect of it.

How does preview impact looking behavior?

Providing object names during preview did not enhance competition effects, but there were clear differences between No preview and Self-paced preview. The Self-paced condition showed greater target looks, greater early looks to the other item types, and a larger competitor effect. The lack of an effect in the Text condition suggests these are likely not a result of prenaming. So, what causes these changes?

The VWP relies on a complex interaction of both linguistic and nonlinguistic processes to link eye movements to displayed items with underlying lexical activation (Magnuson, 2019). Preview in the VWP is intended to allow some of the non-linguistic aspects of visual processing required in this task to complete before the trial (Fig. 5A). Object identification and visual search require cognitive resources, and can introduce substantial variance in looking time between participants and trials. Participants may need to make fixations to the objects in order to identify them and bind their features to locations in space. If participants are doing these things while simultaneously recognizing the word, looks cannot entirely reflect the lexical processing (Fig. 5B). As a result, when these visuo-cognitive processes must be put off until lexical processing is underway (e.g., with no preview), early fixations could be more uniformly distributed to all four objects and thus cannot cleanly reflect lexical processing.

Previous studies that examined stimulus preview contrasted trials with preview to those without, or identical previews of different duration (Chen & Mirman, 2015; Yee et al., 2011). These designs confound prenaming with locating and identifying visuo-semantic features. Although Huettig and McQueen (2007) provide a different preview scenario (with text instead of images), they used the same information in preview and responses, and in the same locations, making it impossible to separate these factors from prenaming. The present study parses some of the various processes at play to determine what components of processing occur at what points.

The Visual-New locations condition provided the opportunity to perform object recognition before the trial begins, but without providing locations. The Text condition, meanwhile, provided the wordforms, but not the visual realization of these wordforms nor their locations. This comparison showed clearer competitor effects in the Visual-New locations condition. Critically, in both conditions, participants have access to the response options in some form. However, the lack of visual-semantic information in the Text condition *reduces* the competitor effect and delays it. As a result, both cohort and unrelated items receive heightened looks, and differentiation occurs substantially later than predicted by theories of incremental processing. The Text condition seems to have greater visual processing demands to identify what is in the display during the period when lexical processing is ongoing.

This contradicts arguments that competitor effects arise *because* of prenaming – the more explicit prenaming condition (Text preview) leads to *less* observed competition. The Visual-New locations condition did not provide any additional phonological cuing over the Text condition. While participants could have still activated phonological forms in the Visual conditions, this (pictorial) entry point to the wordforms is more distal than when providing the text of each word in part because pictures may cue multiple words, while an orthographic string cues only one. The increase in competition for the Visual conditions thus likely indicates greater sensitivity to ongoing phonological competition. When participants do not need to complete object recognition after the word is heard, fixations can more directly reflect phonological processing.

In addition to recognizing semantic features, participants must also bind objects in space, to know where to look to find the visual features they have activated. The Visual-Same locations condition added location information to the visual object information in the Visual-New locations condition, while keeping other information consistent– participants did not have any phonological cuing, nor increased time to prename the objects. Nonetheless, the Visual-Same locations condition showed a further increase in competitor effects. When participants were cued to the object locations prior to the trial, competitor looks (cohort minus unrelated) increased. There is no reason this condition should particularly boost phonological activation, as the same images are shown in both conditions. Instead, this points to an increased sensitivity to measure ongoing phonological competition as a result of this preview manipulation.

Thus, it appears that preview interacts with several aspects of object recognition and processing in the VWP in ways that can mask or reveal standard lexical competition effects. However, it does not appear to substantially impact phonological processing.

What form should preview take?

On the whole, the results argue that stimulus preview does not play a causal role for competitor effects in the VWP. Rather, competition is observed whether or not preview occurs; it is not enhanced when wordforms are directly provided; and it is enhanced by manipulations that provide non-phonological information. That is, conditions like the Self-paced and the Visual-Same locations conditions, which target non-linguistic aspects of the task, offer the most precise characterization of competition (e.g., the cohort-unrelated looking).When preview fails to include visual properties (as these conditions do), this disrupts the measurement of competitor effects: competitor effects are reduced, and differentiation of fixations is delayed, limiting the ability to time-lock analysis to the ongoing incremental processing of stimuli.

However, these conditions raise a smaller issue that is worth considering. Both conditions (Visual-Same locations and Self-paced) showed small increases in the early looks to all objects, irrespective of their fit to the phonological information, and well before auditory information could drive looks. These early eye movements might be driven more by visual salience of objects, strategies (e.g., always fixating the top-left object), or attempts to guess what object might be the target. Whatever their provenance, these early looks could add variance to the measures of phonologically relevant looking that occur once the auditory stimulus begins.

This heightened tendency for early fixations was strongest in conditions that provided object locations during preview. These conditions might draw attention to these locations before the trial starts – the participants know exactly where the objects can occur, so they know where they should direct their eyes. This knowledge reduces search demands – the competitor effects ultimately show greater sensitivity – but it also introduces a small amount of noise early in the trial. In contrast, conditions without location information, such as the Visual-New locations condition, do not directly cue the response locations directly.

This cueing effect is strongest in the Self-paced condition. Three factors might contribute to this. First, objects remain on the display between preview and stimulus onset. Participants have a constant view of the objects in their locations, making fixating them extremely easy. Second, the participants dictate when the auditory stimulus is presented, so they know exactly when they can begin making eye movements. The other conditions had a fixed delay between the onset of preview and its offset, plus an additional fixed delay before response options appear and the stimulus is presented. Participants would have to estimate these times to accurately predict when the trial will begin. Inaccuracies in these estimates might reduce early predictive looks. Finally, preview time tended to be shorter in this condition than in the other conditions. One might imagine that the likelihood of fixating the objects decreases over preview time (as listeners have extracted the information they need); in this case, the particular fixed-duration preview used here simply provided more time for that reduction to occur.

This conflict of the benefits of reducing search demands while increasing location cues raises the question of whether preview could be further improved. The self-triggering used in many VWP studies (including our own) could be somewhat problematic, as this condition produced the most early fixations (though numerically there were still

few).

A second degree of freedom might be removing location information (e.g., as in the Visual-New locations condition). Eliminating location information entirely would likely weaken sensitivity to competitor effects. Thus, ideally, location information should be delivered without drawing looks to the specific response locations. For example, during preview, the images could be displayed in their correct orientation, but not in the identical locations used during the study—the preview objects could be shown in a small rectangle nearer the center of the screen. This would help with visual search on a gross level – the participant can know that the *brain* image is in the top-left – while not drawing attention to the exact screen locations where responses will appear. Alternatively, the pictures could be shown in the correct location, followed by a blank screen just before the stimulus is heard (as in the Visual-Same locations condition). It may also be helpful to introduce a small variable delay between the participants' clicking the dot and the auditory stimulus onset. Future research should investigate whether these kinds of approaches maintain the benefits of sensitivity to competitor effects while reducing the tendency for early looks. Despite the potential value of such work to further strengthen the structure of preview, the present study clearly indicates that previewing both the objects and their locations is not problematic, and improves sensitivity.

Toward a better linking function

Skeptics of the VWP have correctly argued that we need a linking hypothesis that more effectively captures the varied processes needed to complete the VWP task. Such a hypothesis can do more than help us refine an important method in psycholinguistics as it may offer broader theoretical insight into how language processing interacts a rich and potentially dynamic visual environment (Altmann & Mirković, 2009; Magnuson, 2019; Spivey, 2007).

One proposed process was that people name the items during preview and competition plays out in working memory. This is clearly wrong. As we have described, there is considerable empirical evidence against this. More importantly, our data directly rule this out by showing that the situations most conducive to naming have some of the weakest competition effects.

However, the broader need for a more sophisticated linking hypothesis remains. The simplest linking function, that fixations to objects are a read-out of their activation level, is also clearly insufficient. Mere phonological read-out ignores how phonology is mapped to the visual-semantic representations, which in turn are bound to locations and subject to visual search. More critically, the results of this study show that patterns of fixations depend on factors both within and outside phonological processing, such as identifying the visual objects and locating them in space. This may suggest something closer to an interactive visual search process. Specifically, we suggest that the VWP has a linking function that includes several distinct processes, and that the timing of these processes depends on the nature of the trials.

We present a schematic of possible linking functions and how they might interact with preview in Fig. 5. Note that while we visualize these as sequential boxes and talk about them as stages, these clearly operate in a continuous cascade—as is consistently shown in psycholinguistics (Apfelbaum et al., 2011; Marslen-Wilson & Zwitserlood, 1989; McClelland & Elman, 1986; Sarrett et al., 2020). In the case of a typical preview of responses (Fig. 5A), participants first activate spatially localized visual and/or semantic features in the visual world. At this point, participants perform some aspects of object recognition – for example, they identify the colors of objects and their component parts. However, these objects are not likely to be explicitly prenamed at this time – as discussed above, there are myriad reasons why such prenaming is unlikely, and our data indicates that it plays a minimal role. Although images can elicit phonological naming, either participants seem not to do so without compelling need, or they do so, but this exerts minimal effects on later fixations. At this time, participants can also bind these features to locations – they recognize that a red item is in the top left quadrant, for example, or that the floppy ears are in the bottom right. Once the participant hears an auditory stimulus, they then begin activating words that match auditory input. As they activate the words, they also activate their semantic representations. As they do this, they map the activated semantic representations onto the identified semantic features in the display. This stage is what drives the patterns of eye movements – as a word becomes activated, fixations are directed toward semantic features that match those of the word (c.f., Spivey, 2007, chapter 7 for simulations). These stages may interact, but as a normal part of interpreting language in a visual environment, not via some dedicated epiphenomenal task-specific process. Critically, we've left off a role for activating names from the pictures (either during preview or the trial) – at this point, there appears little evidence that this plays a role in the VWP used here. However, listeners clearly can do this, and we see multiple avenues where it could be integrated into this simplified model down the road.

According to this linking function, when these processes begin depends on the timing of the displays. If there is a preview, participants can complete the first two steps before hearing the auditory stimulus, while visually processing the scene. Then when the stimulus is presented, they only need to map activated semantic features onto the already identified and located visual representations. However, if preview does not occur (Fig. 5B), these processes must occur simultaneously, adding noise and delaying what appear as phonologically-driven fixations – they "waste" fixations on visual processing and search, leading early trial information to be less informative. This is just as we found in the No preview and Text conditions here, and as (Huettig & McQueen, 2007, Experiment 2) found with extremely short previews.

A further piece of evidence for this linking hypothesis comes from work showing fixations to objects with colors that match the target in the VWP (Huettig & Altmann, 2011). This study found that participants direct eye movements toward objects that match the color of the target (when the target is *frog*, participants make looks to *spinach*), but only when the objects are displayed in their typical colors. These looks to the color-match even occurred for objects without typical colors (e.g., a green blouse). This pattern is exactly what is predicted if participants activate visual features (like color) and bind them in space, and then direct fixations to features in the display that match the semantic features of the words that are activated. When the participant hears *frog*, she begins to direct eye movements toward objects that share features with frogs – in this case, objects that are green. But if objects are presented in black and white, no color features are initially activated, so these objects do not draw looks.

This linking hypothesis is akin to the "Just-in-Time Deep Interaction" linking hypothesis detailed by Magnuson (2019). Under this linking hypothesis, levels of processing interact throughout processing, but these interactions depend on the task at hand. Our major addition here is to work out the specifics of preview. This hypothesis suggests considerable flexibility. For example, when a person can assume the visual display is stable, they have no need to internally code that display – they can refer back to it as a form of memory offloading (c.f., Ballard et al., 1995). Object names are only accessed when they are needed – "just in time" – if they are needed at all. When context demands more immediate encoding (e.g. when stimuli were visually masked in Pontillo et al., 2015), people are more likely to name the items during preview, and resort to use of working memory to accomplish the task since visual-semantic features may not be available later. In most cases, the VWP operates like the former; images are provided during preview, and processing of these images can begin, but the participant need not (and indeed, typically likely cannot) maintain all possible names of all objects in working memory.

This model is not new. Versions of it are seen in Spivey (2007, chapter 7), and it is consistent with Chen and Mirman (2012; 2015), and with models of sentence processing that stress the continuous interaction of language processing with real-world knowledge, and non-linguistic events (Altmann & Mirković, 2009). In each of these cases,

linguistic and visual domains exhibit interactive crosstalk, but the nature of this cross-talk is highly dependent on context, and the context can alter the dynamics of these interactions. Yee and colleagues elegantly showed how sensitive these interactions are to temporal manipulations (Yee et al., 2011). The current study demonstrates other factors relevant to these dynamic processes and argues for more thorough consideration of the factors at play in the VWP. Crucially, we stress that visual-semantic and search processes should not be ignored when developing VWP paradigms, and that careful consideration of these processes can help clarify the linking function.

Limitations

The present study dissects the possible effects of stimulus preview to identify how prenaming, object identification and visual search impact fixation patterns during VWP trials. Critically, these manipulations showed that phonological competitor effects are not caused by preview, and seem insensitive to manipulations that most encourage phonological prenaming. However, past work on preview suggests that effects are sensitive to time manipulations as well (see especially, Huettig & McQueen, 2007). It is possible that the current manipulations of the form that preview takes might also be sensitive to time manipulations. For example, providing visual information for longer periods might eventually lead to stronger evidence of reduced sets of consideration. Chen and Mirman (2012; 2015) argued that increased processing time during preview can lead to greater phonological-semantic cross-activation, so particularly long preview durations might eventually lead to some form of lexical activation bias for the previewed items. This seems unlikely to be a problem for most versions of the VWP, as the current design used a fairly long preview without incurring this issue (1500 msec), and this was longer than participants used when allowed to self-cue. Still, this could help more fully understand the interaction between the visual scene and phonological processing.

Moreover, even in our Text condition, we do not know whether participants actually prenamed the objects, and we cannot say that no prenaming occurred in the visual conditions. In fact, according to the linking hypothesis developed here they may not. But relative to typical (picture) previews in the VWP, the Text preview should have provided far easier access to the correct names (e.g., there's no chance of misnaming the *couch* as a *sofa*), and it should have encouraged this strategy more than other conditions. The fact that few differences are observed suggests that either participants did not prename in this condition (and therefore they most likely did not prename in the picture conditions) or that they did prename and it had minimal effect.

The present study focused on phonological competition during word recognition to investigate how preview impacts the VWP. The results provide strong evidence that preview does not weaken the construct validity of the VWP for measuring phonological competition, and in fact likely enhances its validity. However, it is possible that preview might have different effects for other linguistic constructs. For example, semantic processing has been shown to be highly sensitive to preview duration (Chen & Mirman, 2015; Yee et al., 2011). Perhaps preview is more requisite for semantic processing. Additionally, higher level language research often uses the visual scene in the VWP as a principle tool for investigating context effects on linguistic processing (e.g., Hanna & Brennan, 2007; Sedivy et al., 1999; Tanenhaus et al., 1995). For these domains, the use of visual preview might require deeper study.

Conclusions

Persistent critiques of the VWP highlight the use of pre-trial stimulus preview as a potential cause of research findings in this paradigm. These critiques draw from studies that contrast a standard preview at different durations, or preview and lack of preview. The current study identified a range of processes that may be carried out during preview: object recognition, locating features in space, and prenaming the responses, to understand how these processes interact throughout VWP trials. Ultimately however, phonological competitor effects proved not to rely on stimulus preview – they were apparent even when no preview was provided – strongly countering doubts about what the VWP is measuring. Instead, some aspects of preview that reduced variance of visual-semantic factors during the trial appeared critical for sensitivity to competitor effects by reducing other sources of noise in the measurement. The results strongly support stimulus preview as beneficial for the VWP and demonstrate the continued value of this technique for measuring the real-time dynamics of spoken word recognition.

Table A1
Stimulus sets used in the study.

SET 1				SET 2			
Pair 1 Word 1	Pair 1 Word 2	Pair 2 Word 1	Pair 2 Word 2	Pair 1 Word 1	Pair 1 Word 2	Pair 2 Word 1	Pair 2 Word 2
bark	barn	jug	judge	bark	barn	crown	crowd
bat	bag	plane	plate	bat	bag	race	rake
beak	beach	moon	moose	beak	beach	horn	horse
bell	bed	cat	cab	bell	bed	cake	cave
pit	pig	snake	snail	pit	pig	brain	braid
bug	bus	goat	goal	bug	bus	corn	cork
brain	braid	robe	rose	jug	judge	snake	snail
cake	cave	hole	hose	plane	plate	robe	rose
race	rake	well	web	moon	moose	plum	plug
corn	cork	peach	peace	cat	cab	peach	peace
crown	crowd	peak	peas	goat	goal	peak	peas
horn	horse	plum	plug	hole	hose	well	web

CRediT authorship contribution statement

Keith S. Apfelbaum: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Jamie Klein-Packard:** Conceptualization, Data curation, Project administration. **Bob McMurray:** Conceptualization, Methodology, Visualization, Supervision, Funding acquisition, Writing - review & editing.

Acknowledgements

This work was funded by DC008089 awarded to BM.

Appendix A

See Table A1.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2021.104279.

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*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558.

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*(4), 502–518. https://doi.org/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. https://doi.org/10.1016/j.cognition.2008.12.005.

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x.

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), 208–216. https://doi.org/10.1016/j.actpsy.2011.01.007.

Apfelbaum, K. S., Blumstein, S. E., & McMurray, B. (2011). Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. *Psychonomic Bulletin & Review, 18*(1), 141–149. https://doi.org/10.3758/s13423-010-0039-8.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology, 7*, 1–11. https://doi.org/10.3389/fpsyg.2016.01116.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review, 119*(2), 417–430. https://doi.org/10.1037/a0027175.

Chen, Q.i., & Mirman, D. (2015). Interaction between phonological and semantic representations: Time matters. *Cognitive Science, 39*(3), 538–558. https://doi.org/10.1111/cogs.2015.39.issue-310.1111/cogs.12156.

Clopper, C. G., Pisoni, D. B., & Tierney, A. T. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology, 17*(5), 331–349. https://doi.org/10.3766/jaaa.17.5.4.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J., Andrews, S., … Ki-, S. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. https://doi.org/I0.1037//0033-295X.108.1.204.

Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language, 57*(4), 483–501. https://doi.org/10.1016/j.jml.2007.01.001.

Dahan, D., & Magnuson, J. S. (2006). Spoken Word Recognition. In Handbook of Psycholinguistics (pp. 249–283). Elsevier. https://doi.org/10.1016/B978-012369374-7/50009-2.

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*(4), 317–367. https://doi.org/10.1006/cogp.2001.0750.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5–6), 507–534. https://doi.org/10.1080/01690960143000074.

De Groot, F., Huettig, F., & Olivers, C. N. L. (2016). When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 180–196. https://doi.org/10.1037/xhp0000102.

Farris-Trimble, A., & McMurray, B. (2013). Test – retest reliability of eye tracking in the Visual World Paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research, 56*(August), 1328–1346. https://doi.org/10.1044/1092-4388.

Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 308–327. https://doi.org/10.1037/a0034353.

Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. M. Henderson, & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World2* (pp. 213–247). Psychology Press.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*(4), 596–615. https://doi.org/10.1016/j.jml.2007.01.008.

Henderson, J. M., & Ferreira, F. (2013). The Interface of Language, Vision, and Action. In *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*. Psychology Press. https://doi.org/10.4324/9780203488430.

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology, 64*(1), 122–145. https://doi.org/https://doi.org/10.1080/17470218.2010.481474.

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. https://doi.org/10.1016/j.jml.2007.02.001.

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. https://doi.org/10.1016/j.actpsy.2010.11.003.

Kapnoula, E. C., & McMurray, B. (2016). Training alters the resolution of lexical interference: Evidence for plasticity of competition and inhibition. *Journal of Experimental Psychology: General, 145*(1), 8–30. https://doi.org/10.1037/xge0000123.

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition, 134*, 85–99. https://doi.org/10.1016/j.cognition.2014.09.007.

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science, 3*(2), 113–139. https://doi.org/10.1007/s41809-019-00035-3.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*(1), 133–156. https://doi.org/10.1080/03640210709336987.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202.

Marian, V., Spivey, M. J., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86*(1), 70–82. https://doi.org/10.1016/S0093-934X(02)00535-7.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 576–585. https://doi.org/10.1037/0096-1523.15.3.576.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0.

McMurray, B., Klein-Packard, J., & Tomblin, J. B. (2019). A real-time mechanism underlying lexical deficits in developmental language disorder: Between-word inhibition. Cognition, 191(April 2018), 104000. https://doi.org/10.1016/j.cognition.2019.06.012.

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003.

Mirman, D., McClelland, J., Holt, L. L., & Magnuson, J. S. (2008). Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms. *Cognitive Science, 32*(2), 398–417. https://doi.org/10.1080/03640210701864063.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56.

Pontillo, D. F., Salverda, A. P., & Tanenhaus, M. K. (2015). Flexible use of phonological and visual memory in language-mediated visual search. *Proceedings of the Cognitive Science Society*, 1895–1900.

Revill, K. P., & Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychology and Aging, 27*(1), 80–87. https://doi.org/10.1037/a0024113.

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology, 51*(12), 1690–1703. https://doi.org/10.1037/dev0000044.

Sarrett, M. E., McMurray, B., & Kapnoula, E. C. (2020). Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations. *Brain and Language, 211*(January), Article 104875. https://doi.org/10.1016/j.bandl.2020.104875.

Sarrett, M. E., Shea, C., & McMurray, B. (submitted for publication). Within- and between-language competition in adult second language learners: Implications for language proficiency. https://psyarxiv.com/bwv7c/.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6.

Shook, A., & Marian, V. (2012). Bimodal bilinguals co-activate both languages during spoken comprehension. *Cognition, 124*(3), 314–324. https://doi.org/10.1016/j.cognition.2012.05.014.

Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). Some Considerations in Evaluating Spoken Word Recognition by Normal-Hearing, Noise-Masked Normal-Hearing, and Cochlear Implant Listeners. I: The Effects of Response Format. *Ear and Hearing, 18*(2), 89–99. https://doi.org/10.1097/00003446-199704000-00001.

Spivey, M. J. (2007). *The Continuity of Mind*. Oxford University Press.

Spivey, M. J., & Marian, V. (1999). Cross Talk Between Native and Second Languages: Partial Activation of an Irrelevant Lexicon. *Psychological Science, 10*(3), 281–284. https://doi.org/10.1111/1467-9280.00151.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically Mediated Visual Search. *Psychological Science, 12*(4), 282–286. https://doi.org/10.1111/1467-9280.00352.

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. https://doi.org/10.1023/A:1026464108329.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863.

Toscano, J. C., Anderson, N. D., & McMurray, B. (2013). Reconsidering the role of temporal order in spoken word recognition. *Psychonomic Bulletin & Review, 20*(5), 981–987. https://doi.org/10.3758/s13423-013-0417-0.

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(3), 387–401. https://doi.org/10.1002/wcs.1178.

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), 348–363. https://doi.org/10.1037/a0022840.