Cognition 222 (2022) 104994

Contents lists available at ScienceDirect

Cognition

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

Language is activated by visual input regardless of memory demands or capacity

Sarah Chabal, Sayuri Hayakawa *, Viorica Marian

Northwestern University, United States

ARTICLE INFO

Keywords:
Phonological competition
Language activation
Visual search
Working memory
Cognitive load
Visual world paradigm

ABSTRACT

In the present study, we provide compelling evidence that viewing objects automatically activates linguistic labels and that this activation is not due to task-specific memory demands. In two experiments, eye-movements of English speakers were tracked while they identified a visual target among an array of four images, including a phonological competitor (e.g., *flower-flag*). Experiment 1 manipulated the capacity to subvocally rehearse the target label by imposing linguistic, spatial, or no working memory load. Experiment 2 manipulated the need to encode target objects by presenting target images either before or concurrently with the search display. While the timing and magnitude of competitor activation varied across conditions, we observed consistent evidence of language activation regardless of the capacity or need to maintain object labels in memory. We propose that language activation is automatic and not contingent upon working memory capacity or demands, and conclude that objects' labels influence visual search.

Analogies for human perception abound – the eye is like a camera, or the ear is like a microphone. Unlike their mechanical counterparts, however, human sensory organs are connected to a broader cognitive network. Perception, for example, is highly interconnected with language (e.g., Bles & Jansma, 2008; Görges, Oppermann, Jescheniak, & Schriefers, 2013; Lupyan, Rahman, Boroditsky, & Clark, 2020; Meyer, Belke, Telling, & Humphreys, 2007; Noizet & Pynte, 1976; Zelinsky & Murphy, 2000), and the linguistic forms of visually-presented objects can impact performance even on non-linguistic visual-processing tasks (Chabal & Marian, 2015; Görges et al., 2013; Meyer et al., 2007; Walenchok, Hout, & Goldinger, 2016). However, the circumstances under which language influences perception and the way in which linguistic forms are bound to visual and spatial representations have not been fully defined. In addition to open questions regarding the mechanisms underlying linguistically-mediated visual processing, ambiguity surrounding the automatic vs. goal-driven nature of language-vision interactions can cast doubt on the degree to which laboratory-based findings inform our knowledge of real-world cognition (see Magnuson, 2019 for discussion). The current study advances our understanding of both the mechanisms and extent of language-vision interactivity by assessing whether language activation during visual search is contingent upon the availability of working memory resources and the need to remember an object for further processing.

When participants are instructed to search for a visually-presented target picture (e.g., a *bat* [baseball]), their eye movements are drawn to pictures of objects with the same name as the target (e.g., *bat* [animal]; Meyer et al., 2007). Similar linguistic effects have been demonstrated with only partial name overlap (e.g., *clock-clouds*; Chabal & Marian, 2015; *cat-hat*; Görges et al., 2013). Because objects in these studies do not share semantic or shape similarity, fixations to phonological competitors are taken as evidence that language features of visual objects become automatically activated even in non-linguistic tasks. However, while such search paradigms are elegant because they can be constructed non-linguistically (i.e., with images as target cues), they do introduce a working memory component because participants must remember the target in advance of search. Specifically, working memory is believed to play an important role in the maintenance of “target templates” so that a visual or linguistic target can be identified from within a subsequently presented search array (Desimone & Duncan, 1995; Duncan & Humphreys, 1989). This may encourage participants to adopt a task-specific strategy of intentionally generating linguistic labels to maintain the target in working memory. If so, internal repetition of the target word might activate linguistic information that would increase competitor fixations. Language activation may then follow a

* Corresponding author at: Department of Communication Sciences and Disorders, Northwestern University, 2240 North Campus Drive, Evanston, IL 60208, United States.

E-mail address: sayuri.hayakawa@northwestern.edu (S. Hayakawa).

https://doi.org/10.1016/j.cognition.2021.104994
Received 8 February 2021; Received in revised form 9 December 2021; Accepted 16 December 2021
Available online 10 January 2022

bottom-up process similar to competition in traditional visual world paradigm studies, in which participants are presented with an auditory word (e.g. "beaker") that must then be located from within a visual display (Allopenna, Magnuson, & Tanenhaus, 1998; Marian & Spivey, 2003a; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The use of language-based strategies additionally invites the possibility that, rather than reflecting natural language processes, phonological competition during visual search and visual world tasks could plausibly be an artifact of experimental demands (see Magnuson, 2019 for discussion).

On the other hand, it could be argued that even if subvocal rehearsal is deliberately engaged under some conditions, this does not necessarily negate the contributions of more implicit processes where the visual forms of familiar objects automatically spread activation to associated labels. Indeed, although subvocal rehearsal and articulation has been shown to facilitate search when a target is presented linguistically (e.g., seeing or hearing the word "clock" and then identifying the *clock* from an array of other visual objects; Lupyan & Swingley, 2012), or when multiple objects or digits must be held in memory (Conrad, 1964; Logie, Gilhooly, & Wynn, 2016; Salamé & Baddeley, 1982), it is less clear that rehearsal would improve upon a non-linguistic strategy when a single target and the display are presented visually (e.g., seeing an image of a *clock* and then identifying the *clock* from an array of other visual objects). Given evidence that visual cues are often more efficient than verbal cues (Vickery, King, & Jiang, 2005; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004), participants' initial viewing of the target should be a sufficient search cue in a simple visual discrimination task. An efficient search strategy, then, could be to form a mental image of the target, *not* to rehearse its label.

Furthermore, intentional encoding and rehearsal of target names is not sufficient to explain all effects of language that have been observed during visual scene processing, for example, those that are found in the bilingual literature on co-activation across languages. When bilingual speakers are presented with a visual search task, linguistic information from both of their languages is accessed. Specifically, upon viewing an image of a target object (e.g., a *clock*), bilingual English-Spanish speakers tested in English fixated images whose names overlapped phonologically with the target in both English (e.g., *clock-cloud*) and Spanish (e.g., *reloj-regalo* [clock-gift]), even when no language was used during the task (Chabal & Marian, 2015). If this effect were due to intentional encoding and rehearsal of the target, competition in both languages would be unlikely, as bilinguals of two spoken languages would not have the need or capacity to *simultaneously* rehearse labels in both languages. The deliberate strategic rehearsal of multiple labels would instead require sequential switching between the two languages, which would not only be inefficient (given that rehearsal in a single language should be sufficient to remember the target), but also implausible given the typically short duration of time between the target preview and onset of the search display (e.g., 1000 ms in Chabal & Marian, 2015). In other words, if rehearsal of the label were responsible for phonological competition, then activation would only be seen for the label that bilinguals were rehearsing in one language, as opposed to the simultaneous co-activation of the labels in both the first and the second languages (as has been observed with no verbal input in any language in Chabal & Marian, 2015; with verbal input in one language in Blumenfeld & Marian, 2005, Marian & Spivey, 2003a, 2003b, Spivey & Marian, 1999, and many others). Such evidence of phonological competition during simple visual discrimination lends support to the notion that deliberate subvocal rehearsal is not necessary to elicit language activation during visual processing.

Here, we assess the automaticity of language activation during non-linguistic visual processing by directly manipulating the *capacity* to strategically maintain a linguistic target template in working memory (Experiment 1), as well as the *need* to remember the object before initiating visual search (Experiment 2).

1. Experiment 1: The influence of memory *load* on language activation

According to Shiffrin and Schneider's (1977) dual-mode model of information processing, the content of short-term memory constitutes nodes which can be retrieved from long-term memory via resource-demanding controlled processes (e.g., rehearsal, long-term memory search), as well as through relatively resource-independent, stimulus-driven cascading activation. The latter is often considered to be a characteristic of automatic processing, such that activation is not contingent on the availability of working memory resources or goal-driven strategies (Klauer & Teige-Mocigemba, 2007; Moors & De Houwer, 2006). Within this framework, if linguistically-mediated visual processing results from the automatic, cascaded activation of object labels from visual inputs, we should find evidence of language activation even when verbal working memory resources are depleted by a concurrent linguistic memory task.

The extent to which language activation relies on verbal working memory and the subvocal maintenance of a target label can be tested using a dual-task paradigm. In the dual-task paradigm, participants are asked to perform two tasks simultaneously; if performance on one of the tasks is affected by the need to perform the other, the two tasks are said to compete for similar cognitive processing resources (for a review see Kahneman, 1973). If participants engage in strategic subvocal rehearsal of the target label, introducing verbal working memory load should disrupt this process and no phonological competition should be observed (see Winawer et al., 2007 for a similar approach to assessing the role of language in color perception). If phonological competition is observed even when verbal working memory load is imposed, this would suggest that language can be automatically activated when viewing visual objects without the intentional maintenance of the target label in memory.

In Experiment 1, participants were asked to complete a simple visual search task either under normal viewing conditions, in the presence of a linguistic memory load, or in the presence of a spatial memory load. During the search task, participants saw a target object as a search cue (e.g., a picture of a *flower*) and were asked to click on an identical image from a search display of four images. Present in the display was an item whose name overlapped phonologically with the name of the target (e.g., *flower-flag*). Past research has demonstrated that phonologically-similar items compete for selection and attract eye movements, presumably because linguistic information about the visually-presented objects is automatically activated (e.g., Chabal & Marian, 2015; Meyer et al., 2007; see Huettig and McQueen's (2007) *Cascaded Activation Model* of visual–linguistic interactions). By imposing a dual task, we can reduce the capacity to subvocally rehearse the target label prior to initiating the search and assess the automaticity of phonological competition between visual objects.

The inclusion of a spatial load condition allows us to further explore the mechanisms through which language is activated during visual scene processing. Because the spatial task does not block direct access to language information, it may not be expected to impact the maintenance of a linguistic target template or the emergence of phonological competition (see Winawer et al., 2007). However, as proposed in Huettig, Olivers, and Hartsuiker's (2011) model of language-vision interactions, working memory may play a key role beyond the maintenance of a target template. Specifically, Huettig et al. (2011) propose that while viewing an object may be sufficient to automatically activate associated features stored in long-term memory (e.g., visual, semantic, and linguistic representations), working memory may be critical for binding the representations activated by the target and display objects, as well as the spatial location of the display object (subsequently enabling the initiation of visual fixations to particular items in specific places). Indeed, there is evidence that imposing a concurrent spatial working memory load can disrupt the efficiency of visual search (Woodman & Luck, 2004).

By directly comparing fixations to targets and phonological competitors under linguistic, spatial, and no-load conditions, we can obtain more fine-grained insight into the roles of different components of working memory, as well as the automaticity of language activation during visual search. Specifically, to the extent that spatial working memory is utilized to initiate fixations to the locations of likely candidates, we would expect that looks to both the target and potential competitors may be delayed and/or reduced when completing a concurrent spatial memory task. Furthermore, if we are correct that subvocal rehearsal of the target label does not meaningfully facilitate target identification beyond the use of a non-linguistic strategy, we should not expect to find an effect of linguistic load on looks to the target. Lastly, and most central to our investigation, if language activation is not contingent on subvocal rehearsal, we should observe that phonological competition emerges despite linguistic load, providing compelling support for the robustness and automaticity of language activation during visual processing.

2. Method

2.1. Participants

Twenty-six native English speakers (9 males) participated in Study 1. The number of participants was determined based on an a priori power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) for linear multiple regression random models at an assumed power of 0.8, an alpha level of 0.05, and a Cohen's f^2 effect size of 0.37 (based on pilot data), which yielded a necessary sample size of 24. This sample size is comparable to prior research using similar paradigms (e.g., Chabal & Marian, 2015, $N = 20$; Walenchok et al., 2016, Ns = 20 (Exp 1a), 22 (Exp 1b), 23 (Exp 2a), 22 (Exp 2b), 23 (Exp 3a), 23 (Exp 3b)). Participants ranged in age from 18 to 29 years (mean age = 21.27, $SD = 2.92$) and reported normal or corrected-normal vision and no history of hearing impairments. Native English status was confirmed by self-report measures on the *Language Experience and Proficiency Questionnaire* (Marian, Blumenfeld, & Kaushanskaya, 2007). See supplementary materials for participant demographics and cognitive information. Informed consent was obtained from all participants and the study protocol was approved by an Institutional Review Board.

2.2. Design and materials

Study 1 was conceptualized as a 2 × 3 repeated-measures design, with item type (Competitor, Control) and load condition (No-Load, Linguistic-Load, Spatial-Load) as within-subject variables.

Thirty critical stimuli sets were constructed following the protocol of Chabal and Marian (2015), each consisting of a target object (e.g., a *flower*), a phonological competitor adjacent to the target whose name in English overlapped phonologically with the English name of the target (e.g., a *flag*), a control object adjacent to the target (e.g. a *knife*), and a filler object diagonal to the target (e.g., a *cat*). Controls and fillers did not share any initial phonological overlap with any other item in the set. An additional 45 stimuli sets were constructed for filler trials designed to mask the phonological manipulation. Each filler trial included a target and three filler items that did not share initial phonological overlap with any other item in the set. Stimuli were depicted by black and white line drawings selected from the *International Picture Naming Project* (*IPNP*) database (Bates et al., 2000); images unavailable from IPNP were independently normed. Stimuli images were controlled to ensure that no items in critical displays were visually or semantically similar. Items were matched on lexical characteristics; see supplementary materials for comparisons between item types and a full stimuli list.

Each participant completed three blocks of search trials corresponding to the three load conditions. All search trials began with a fixation cross, which was displayed for 500 ms, followed by a preview screen containing only the target picture for 1000 ms. Participants were then shown a fixation cross for 1000 ms, which was replaced by a search display containing four images (i.e., target, competitor, control, and filler). Participants were instructed to click on the target as quickly as possible, and the display remained on screen until a response was made. The experiment contained 90 critical trials and 135 filler trials. Each of the 30 critical and 45 filler stimuli sets was repeated under the three load conditions, with the position of items in the display randomized across blocks. Trials were arranged in a pseudo-randomized order that was fixed between participants (but that varied between blocks). Half of the participants received the stimuli within each block in the reverse order.

One block was completed in the absence of any secondary task (*No-Load* condition), one block while performing a linguistic interference task (*Linguistic-Load* condition), and one block while performing a spatial interference task (*Spatial-Load* condition). Block order was counterbalanced across participants. Load conditions were modeled after those used by Winawer et al. (2007). In the *No-Load* block, participants completed the visual search task as described above. In the *Linguistic-Load* block, participants were shown an eight-digit number for three seconds and were instructed to silently rehearse that number. Subjects rehearsed the number series while completing three search trials, then their recall was tested using a two-alternative forced-choice memory probe. Participants were asked to choose between the original number and a foil that differed by one digit to ensure that the numbers were held in verbal working memory during the visual search trials.[1] In the *Spatial-Load* block, participants had three seconds to view a 4 × 4 square grid containing four randomly-shaded squares; they were instructed to "create a mental snapshot" of the grid pattern and maintain that picture in their mind until testing. Participants then completed three search trials, and recall was tested with a two-choice test in which the incorrect grid differed in the location of one shaded square. See Fig. 1 for the structure of each experimental block.

2.3. Apparatus

The experiment was controlled by a 3.1 GHz Intel Core i5 running MATLAB 2010. Stimuli were displayed on a 27 in. monitor, with a screen resolution of 2560 × 1440. Eye movements were recorded using a desk-mounted eye-tracker (EyeLink 1000 Version 1.5.2, SR Research Ltd.) at a sampling rate of 1000 Hz. Accuracy and response time were recorded.

2.4. Procedure

After providing consent, participants were familiarized with the eye-tracker and read the instructions for the eye-tracking procedure. Eye-tracker calibration was obtained twice (at the beginning of the study and half-way through) using a standard 9-point calibration and validation procedure with drift correction. The eye-tracking portion of the experiment lasted approximately 40 min. Participants were instructed to click on a central fixation cross to begin each trial, and to click on the target item as quickly and accurately as possible.

Following the eye-tracking procedure, participants were asked to provide names for each of the target and competitor items seen throughout the experiment. Images that were named incorrectly or were unnamed were discarded individually for each participant on a trial-by-trial basis before further analyses (9.74% of trials discarded).

[1] Though rehearsal of the number string could have instead been confirmed by having participants rehearse them out loud during the visual search trials, overt production may have an effect on performance above and beyond the contents of working memory (e.g., speech motor planning and articulation). As it was not possible to control for independent effects of overt production in the Linguistic-Load block by also requiring production on the No-Load and Spatial-Load blocks (without introducing linguistic load in the latter two), numbers were rehearsed silently and then confirmed via a memory test.

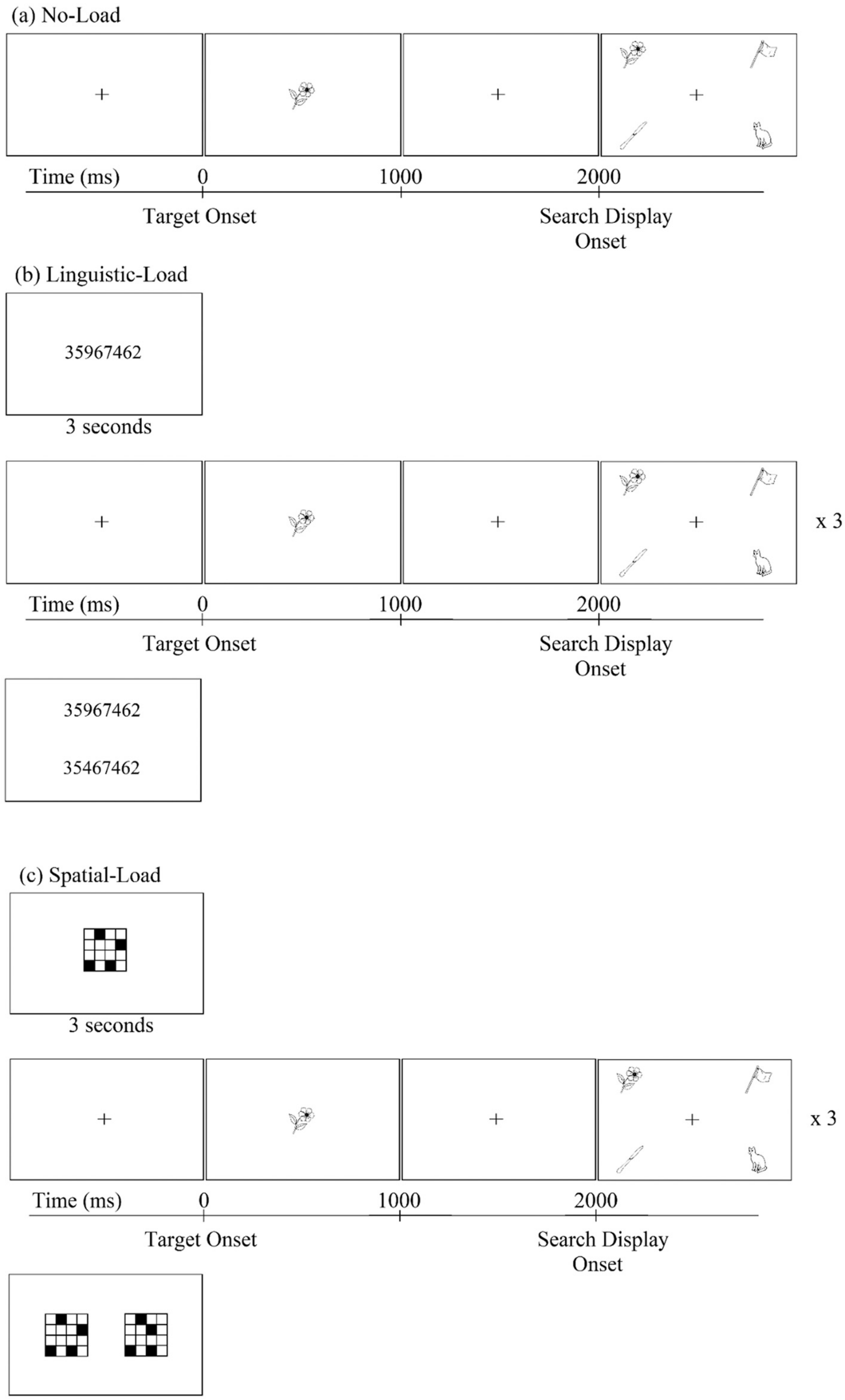

Fig. 1. Sample trial structure for the No-Load (a), Linguistic-Load (b), and Spatial-Load (c) conditions. The target (e.g., *flower*) was present in the search display along with a phonological competitor (e.g., *flag*) and control and filler items (e.g., *knife, cat*) which did not overlap phonologically. Participants were instructed to click on the target object as quickly as possible.

2.5. Data analysis

Phonological competition and the effects of memory load were assessed by examining the proportion of visual fixations that were made to competitor, control, and target pictures at each time point between 0 ms and 2000 ms following the presentation of the search display. After excluding fixations that were less than 70 ms in duration, the number of fixations to each object were first summed across trials at each time point, and then divided by the total number of trials. The time and amplitude of peak activation (i.e., the point at which each object was fixated the most over the timecourse) was then identified for complementary analyses of peak shape and latency. Significance of fixed effects was assessed using Chi-square tests on -2LogLikelihood change in model fit. Parameter-specific *p*-values were estimated by using a normal approximation, treating the t-value from the model as a z-value (See Barr, Levy, Scheepers, & Tily, 2013).

2.5.1. Peak shape

The rise and fall of fixations around the peak of the waveform were analyzed using growth curve analysis with fourth-order orthogonal polynomials (Mirman, Dixon, & Magnuson, 2008; Mirman, Magnuson, Graf Estes, & Dixon, 2008), a form of multilevel regression optimized for assessing change over time. Though a priori predictions were not made regarding the precise way in which phonological competition would influence visual fixations, evidence of phonological competition could be expected to emerge as an overall increase in the proportion of competitor fixations compared to controls, resulting in an effect of competition on the intercept term, which captures the vertical shift in the curve. Phonological similarities between target and competitor objects could also increase the rate at which fixations to competitor objects approach the peak amplitude (i.e., a steeper rise for competitors compared to controls) and/or decrease the rate at which competitor objects are discounted and fixations return to baseline (i.e., a more gradual fall for competitors compared to controls). The linear (or slope) term captures the overall angle or steepness of the curve, while the quadratic, cubic, and quartic time terms each capture the steepness of curve around inflection points: the quadratic term captures the angle of the curve around the central inflection point in the middle of the window (i.e., at peak activation), while the cubic and quartic terms reflect the steepness of the curve around inflections at the tail ends of the waveform. As the time of peak activation varied across participants, objects, and conditions (see *Peak Latency*), individual participants' timecourses for each condition were re-centered so that 0 ms corresponded to the time of peak fixation. Visual fixations to competitors and controls were analyzed in a window from −200 to 200 ms around the peak using binomial generalized linear mixed effects regression with the glmmTMB package (Brooks et al., 2017) in R (version 4.1 see supplementary materials for additional details).

2.5.2. Peak latency

In addition to effects of competition on the magnitude and pattern of visual fixations, phonological similarities between target and competitor objects could result in *earlier* peaks for competitors compared to controls. Peak latency was identified following procedures outlined by Kiesel, Miller, Jolicoeur, and Brisson (2008). In order to account for multiple peaks and the overall distribution of visual fixations, peak latency was defined as the time point corresponding to half of the area under the fixation curve (i.e., the fractional area technique). As the areas under the curves were substantially greater for target fixations, peak latency for targets was additionally determined based on the averaged times corresponding to fixations within 5% of the maximum amplitude (i.e., the peak amplitude technique), which allowed for more fine-grained distinctions (see supplementary materials for additional details). Because of the low signal to noise ratio in eye-tracking fixations, peak latencies were analyzed using a jackknife approach, where the contribution of each participant is assessed by how the average timecourse changes when that participant is removed (Kiesel et al., 2008; Ulrich & Miller, 2001; see supplementary materials for additional details). Peak latencies identified from each of the timecourses were analyzed with generalized linear mixed effects regression using the lme4 package (Bates, Machler, Bolker, & Walker, 2014).

2.5.3. Exclusions

Visual search trials that were responded to incorrectly (0.3% of trials; mean accuracy = 99.7%, SD = 1.30), or trials in which the response time was two standard deviations above or below the mean (4.7% of trials; mean RT = 1136.22 ms, SD = 115.50) were excluded from the analysis. Accuracy on the linguistic (M = 93.4%, SD = 7.03) and spatial (M = 96.2%, SD = 5.08) memory tests was high and did not differ from each other ($p > .05$), confirming that participants attended to the secondary load tasks. Visual search trials corresponding to inaccurate responses on the load tasks were excluded from the eye movement analyses (3.0% of all trials; 4.8% of trials in the linguistic and spatial load blocks). Tukey-adjusted pairwise tests confirmed that visual search accuracy on No-Load trials (M = 99.7%, SD = 1.6) did not differ from Linguistic- (M = 99.7%, SD = 1.2; p = .998) or Spatial-Load trials (M = 99.7%, SD = 1.0; p = .996), and the two load trials did not differ from each other (p = .987). Likewise, response times on No-Load trials (M = 1136.5 ms, SD = 136.7) did not differ from Linguistic- (M = 1139.5 ms, SD = 106.6; p = .973) or Spatial-Load trials (M = 1132.6 ms, SD = 104.9; p = .999), which did not differ from each other (p = .962). All reported p-values were generated with double-sided tests.

3. Results

3.1. Competitor and control fixations

3.1.1. Peak shape

The effects of load and phonological competition were examined with a growth-curve analysis with fixed effects of Competition (Control −0.5 vs. Competitor 0.5), Load, and their interactions on all time terms. The effect of Load was modeled as a single factor with two dummy coded comparisons (using R's default treatment contrasts): No-Load (0) vs. Spatial-Load (1) and No-Load (0) vs. Linguistic-Load (1). The model additionally included a random intercept for subject and by-subject random slopes for all time terms. Random effects of item were not included as the proportion of fixations to each object (competitor, control) were determined by aggregating across trials at each time point. Model comparisons between full and depleted models (dropping each fixed effect) revealed significant effects of Competition (Likelihood Ratio Test, $\chi^2(1) = 209.93$, $p < .001$) and Load ($\chi^2(2) = 477.89$, $p < .001$), indicating that the patterns of fixations differed between competitors and controls, as well as across load conditions. There were also significant interactions between Competition and Load on the intercept ($\chi^2(2) = 127.97$, $p < .001$), and on the linear ($\chi^2(2) = 139.63$, $p < .001$), quadratic ($\chi^2(2) = 125.38$, $p < .001$), cubic ($\chi^2(2) = 41.46$, $p < .001$), and quartic ($\chi^2(2) = 68.62$, $p < .001$) time terms, indicating that the effects of competition on fixation curves differed across the load conditions. Parameter estimates of the full model (Table 1) reveal that the effect of competition on the intercept and each time term significantly differed between the No-Load and Spatial-Load conditions, as well as between the No-Load and Linguistic-Load conditions. A significant simple effect of competition on the intercept indicates that competitors were fixated more often than controls in the No-Load (reference level) condition. There were additionally significant simple effects of Competition on the linear and cubic time terms, which together captured the steeper curvature of fixations to competitors relative to controls and the increase in competitor (vs. control) fixations at the tail end of the window under No-Load.

Follow-up analyses examined the simple effects of phonological competition on the rate and magnitude of visual fixations around the peak for the Linguistic- and Spatial-Load conditions. The effects of competition on the overall proportion of fixations (i.e. the intercept) and on the effects of each time term were assessed with pairwise comparisons of the estimated marginal means in each load condition. The effect of competition on the intercept was tested using the *emmeans* function,

Table 1
Growth curve analysis of fixation peak shapes to competitor and control pictures in varying load conditions.

Fixed Effect	Estimate	SE	95% CI	z value	
Intercept	−4.26	0.22	[−4.26, −3.83]	−19.38	***
Linear	3.63	1.32	[3.63, 6.22]	2.74	**
Quadratic	−5.07	0.66	[−5.07, −3.78]	−7.73	***
Cubic	1.88	0.36	[1.88, 2.59]	5.16	***
Quartic	−0.26	0.25	[−0.26, 0.22]	−1.07	
Intercept:Competition	0.29	0.02	[0.29, 0.33]	14.40	***
Linear:Competition	−0.57	0.16	[−0.57, −0.26]	−3.63	**
Quadratic:Competition	0.00	0.15	[0, 0.3]	0.01	
Cubic:Competition	0.54	0.14	[0.54, 0.82]	3.92	**
Quartic:Competition	0.34	0.13	[0.34, 0.6]	2.51	
Intercept:Spatial	0.23	0.01	[0.23, 0.25]	17.15	***
Linear:Spatial	−1.30	0.10	[−1.3, −1.1]	−13.05	***
Quadratic:Spatial	1.55	0.10	[1.55, 1.75]	15.88	***
Cubic:Spatial	−0.80	0.09	[−0.8, −0.62]	−8.80	***
Quartic:Spatial	−0.12	0.09	[−0.12, 0.05]	−1.40	
Intercept:Comp:Spatial	−0.19	0.03	[−0.19, −0.13]	−7.06	***
Linear:Comp:Spatial	−0.39	0.20	[−0.39, 0]	−1.97	*
Quadratic:Comp:Spatial	1.32	0.20	[1.32, 1.71]	6.78	***
Cubic:Comp:Spatial	−1.10	0.18	[−1.1, −0.75]	−6.10	***
Quartic:Comp:Spatial	−0.72	0.18	[−0.72, −0.38]	−4.13	***
Intercept:Linguistic	−0.04	0.01	[−0.04, −0.01]	−2.64	**
Linear:Linguistic	0.10	0.11	[0.1, 0.32]	0.91	
Quadratic:Linguistic	0.01	0.11	[0.01, 0.21]	0.05	
Cubic:Linguistic	−0.26	0.10	[−0.26, −0.07]	−2.71	**
Quartic:Linguistic	1.04	0.09	[1.04, 1.22]	11.01	***
Intercept:Comp: Linguistic	−0.32	0.03	[−0.32, −0.27]	−11.19	***
Linear:Comp:Linguistic	1.91	0.22	[1.91, 2.35]	8.55	***
Quadratic:Comp: Linguistic	−0.74	0.21	[−0.74, −0.32]	−3.45	***
Cubic:Comp:Linguistic	−1.00	0.20	[−1, −0.62]	−5.12	***
Quartic:Comp:Linguistic	0.73	0.19	[0.73, 1.1]	3.85	***

Note: Model estimates are given on the logit scale. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 2
Growth curve analysis of fixation peak shapes to competitor and control pictures within Linguistic- and Spatial-Load conditions.

Fixed Effect	Estimate	SE	95% CI	t value
A) Linguistic Load				
Intercept:Competition	−0.04	0.02	[−0.09, 0.02]	−1.66
Linear:Competition	1.34	0.16	[0.89, 1.80]	8.42***
Quadratic:Competition	−0.73	0.15	[−1.15, −0.31]	−4.96***
Cubic:Competition	−0.45	0.14	[−0.84, −0.06]	−3.32*
Quartic:Competition	1.06	0.13	[0.68, 1.44]	7.96***
B) Spatial Load				
Intercept:Competition	0.1	0.02	[0.05, 0.15]	5.98***
Linear:Competition	−0.96	0.12	[−1.31, −0.62]	−7.91***
Quadratic:Competition	1.32	0.12	[0.98, 1.67]	11.01***
Cubic:Competition	−0.56	0.12	[−0.89, −0.23]	−4.83***
Quartic:Competition	−0.39	0.11	[−0.71, −0.06]	−3.41**

Note: Model estimates are given on the logit scale. *P*-values were corrected for multiple comparisons using the Tukey method. *** $p < .001$, ** $p < .01$, * $p < .05$.

while the effect of competition on the slope of each time term was assessed using the *emtrends* function (Lenth, 2021). Family-wise error rates were controlled with Tukey-adjusted *p*-values within each set of comparisons (i.e., effects of competition on the intercept, linear, quadratic, cubic, and quartic time terms). Model estimates for the Linguistic- and Spatial-Load conditions are presented in Table 2 and the proportions of competitor and control fixations in each of the three Load conditions are depicted in Fig. 2.

In the Linguistic-Load condition, there were significant effects of Competition on the linear, quadratic, cubic, and quartic time terms, which like in the No-Load condition, reflected a steeper rise and fall of fixations around the peak, as well as an increase in fixations at the tail end of the window for competitors compared to controls.[2] The relative increase in competitor fixations at the tail was greater under Linguistic-Load than No-Load condition (captured by the significant Competition x Linguistic-Load effect on the cubic and quartic time terms; see Table 1), and unlike the No-Load condition, there was no significant effect of Competition on the intercept, indicating that linguistic load reduced the overall magnitude of competition (reflected in the significant Competition x Linguistic-Load effect on the intercept; see Table 1).

In the Spatial-Load condition, there was a significant effect of Competition on the intercept, indicating that competitors were fixated more than controls under Spatial-Load. Relative to the No-Load condition, the overall difference in the proportion of competitor vs. control fixations was reduced under Spatial-Load (reflected in the significant Competition x Spatial-Load effect on the intercept; see Table 1). There were additionally significant effects of Competition on each of the time terms under Spatial-Load, which together, capture the more gradual and sustained fixations to competitors relative to controls. The more gradual rise and fall of competitor (vs. control) fixations under Spatial-Load differed from the relatively steeper curvature for competitors observed under No-Load (reflected in the significant Competition x Spatial-Load effects on the linear, quadratic, cubic, and quartic time terms; see Table 1). The pattern and/or magnitude of competitor fixations therefore differed from controls within all three load conditions, but also varied as a function of load.

3.1.2. Peak latency

Because peak latencies were non-normally distributed, the effects of Load condition and competition were analyzed using a generalized linear mixed-effects model with an inverse gaussian and identity link (Lo & Andrews, 2015). The model included fixed effects of Competition (Controls −0.5 vs. Competitors +0.5), Load (two contrasts with No-Load 0 vs. Linguistic +1 and Spatial Load +1), and their interactions, plus a random intercept for subject. As in the analyses of peak shape, no random effects of item were included as data were aggregated across items. Model comparisons revealed a significant interaction between Competition and Load ($\chi^2(2) = 21.56$, $p < .001$) and no main effects of Competition or Load (both $p > .05$). Parameter estimates revealed that the effect of Competition differed between the No-Load and Spatial-Load conditions (*Estimate* = 124.01, *SE* = 49.41, $z = 2.51$, $p = .012$, 95% CI [27.18, 220.84]), while the effect of Competition did not differ between the No-Load and Linguistic-Load conditions ($p = .994$). Tukey-adjusted pairwise comparisons of the estimated marginal means revealed that in the Spatial-Load condition, peak latencies were earlier for control pictures than competitors by 97.05 ms ($z = 2.62$, $p = .008$; Fig. 3), likely as a result of more sustained activation of competitors (see Peak Shape, above). In contrast, peak latencies did not differ between competitor and control pictures in the No-Load condition (competitor − control = −26.89 ms, $z = −0.82$, $p = .411$) and peak latencies for competitors were significantly later in the Spatial-Load condition (679.4 ms) compared to

[2] All effects remained significant when block order was included in the model as a continuous covariate. There was additionally a significant interaction between block order and load ($\chi2(2) = 81.42$, $p < .001$). Pairwise comparisons indicate that there were significantly fewer looks to competitors and controls in later blocks compared to earlier blocks in all three conditions (all $p < .001$), but that the effect of block order was greater under Linguistic-Load relative to No-Load (*Estimate* = −0.15, *SE* = 0.02, $t = 6.46$, $p < .001$), as well as under Spatial-Load relative to No-Load (*Estimate* = −0.19, *SE* = 0.02, $t = −8.71$, $p < .001$).

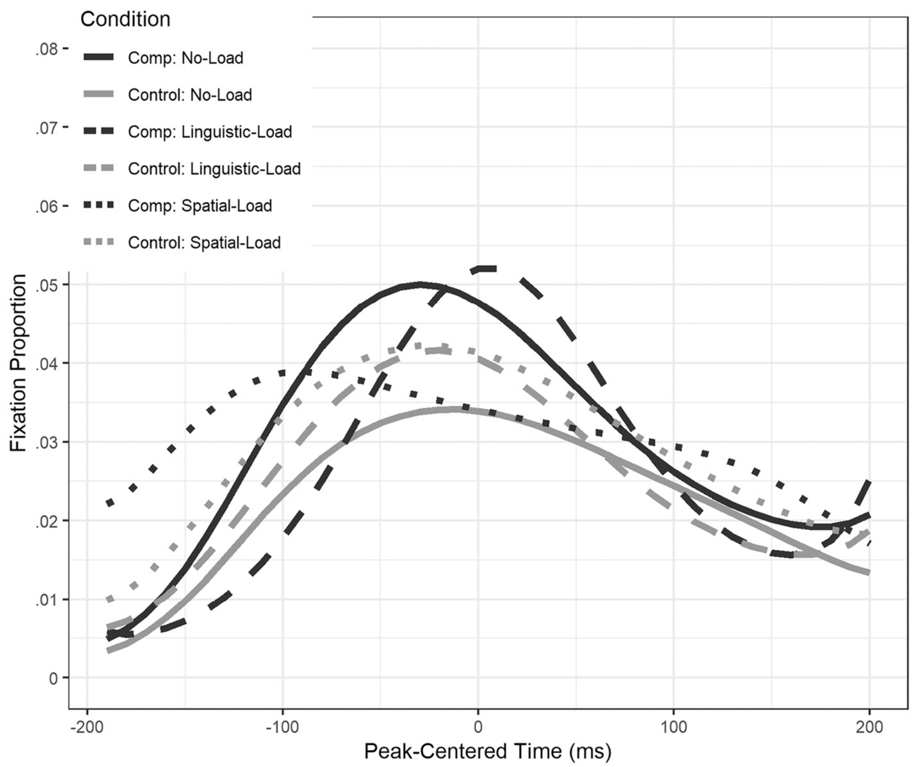

Fig. 2. Visual fixation peak shapes. Significant effects of Competition were observed on the linear and cubic terms in all three conditions, the combination of which reflects differences in the steepness of the curves for competitors (black) vs. controls (gray). In particular, the rise and fall of fixations around the peak were steeper for competitors than controls in the No-Load (solid lines) and Linguistic-Load conditions (dashed lines). In contrast, fixations to competitors in the Spatial-Load condition (dotted lines) were more gradual and sustained relative to controls. There were additionally significant effects of Competition on the intercept for the No-Load (solid lines) and Spatial-Load (dotted lines) conditions, indicating that overall, competitors were fixated more than controls. Lines represent best fit logistic regression model estimates. Timecourses were re-centered for each subject with 0 ms corresponding to peak activation for each condition.

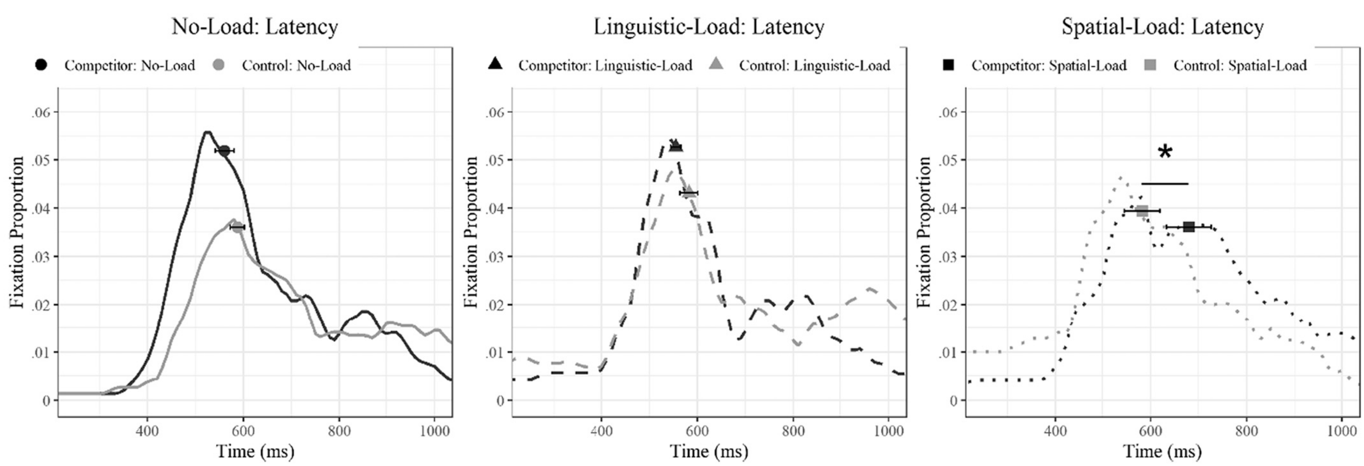

Fig. 3. Eye-tracking latency across load conditions. Peak latencies revealed comparable rates of peak activation for phonological competitors (black) and controls (gray) in No-Load and Linguistic-Load conditions. In the Spatial-Load condition, competitors peaked later than controls due to more sustained activation of the competitor. (Note that peak shapes depicted in Fig. 2 were recentered around the peak time, while Fig. 3 plots fixations across the unadjusted timecourse. Under Spatial-Load, the competitor peak identified using the half-area method was later than that of the control – as a result, the relative timing of competitor and control fixations appeared to reverse in Fig. 2 when the plots were recentered at their peaks). Curves represent observed data, dots represent peak latency, and horizontal lines represent standard error.

the No-Load condition (570.9 ms, $z = 2.97$, $p = .003$). No effects of Load condition were found for control items (all $p > .05$).[3]

Taken together with the analyses of peak shape (see Tables 1 and 2), these findings demonstrate that in all three Load conditions, the timing and/or magnitude of fixations to competitors differed from that of controls, suggesting that linguistic information was activated even when imposing concurrent linguistic and spatial memory loads. The ways in which competitors differed from controls, however, varied as a function of memory load. Peak shape analyses revealed that relative to the No-Load condition, Linguistic-Load attenuated the difference in magnitude between competitor and control activation (resulting in a significant Competition x Linguistic-Load effect on the intercept) and elicited a steeper increase in competitor (vs. control) fixations at the tail end of the window (Competition x Linguistic-Load effects on the cubic and quartic

[3] The same pattern of results was observed when block order was included in the model as a covariate. There was a significant effect of the Competition x Spatial-Load contrast ($p = .010$), with follow-up tests revealing that competitors peaked later than controls in the Spatial-Load ($p = .007$), but not the No-Load condition ($p = .401$), and that competitors ($p = .002$), but not controls ($p = .636$) peaked later in the Spatial-Load than No-Load condition. There were no effects of block order on peak latencies ($ps > 0.05$).

time terms; see Table 1). Analyses of peak latency, however, indicated that the timecourses of activation were similar under No-Load and Linguistic-Load. Shape analyses revealed that Spatial-Load also attenuated the magnitude of competitor activation relative to No-Load (Competition x Spatial-Load effect on the intercept; see Table 1), but also resulted in more gradual and sustained competitor (vs. control) activation (Competition x Spatial-Load effects on the linear, quadratic, cubic, and quartic time terms; see Table 1). Latency analyses also revealed that peak competitor activation under Spatial-Load was delayed relative to No-Load.

3.2. *Target fixations*

3.2.1. *Peak shape*

A growth curve analysis was conducted to examine the effect of load on the shape of target fixations. The model included a fixed effect of Load (No-Load 0 vs. Spatial-Load 1 and No-Load 0 vs. Linguistic-Load 1) plus interactions on all time terms, as well as random effects of subject on all time terms (Fig. 4, *left*). Model comparisons revealed a significant effect of Load on the intercept ($\chi^2(2) = 591.75$, $p < .001$), as well as on the linear ($\chi^2(2) = 488.46$, $p < .001$), cubic ($\chi^2(2) = 45.33$, $p < .001$), and quartic ($\chi^2(2) = 65.12$, $p < .001$) time terms. Parameter estimates (Table 3) show significant effects of both the Spatial-Load and Linguistic-Load contrasts on the intercept. The overall proportion of fixations to the target were greater under Linguistic-Load than under No-Load, which were in turn greater than under Spatial-Load. There were additionally significant effects of Spatial- and Linguistic-Load on the linear and cubic time terms and of Linguistic-Load on the quartic time term, the combination of which reflects differences in the steepness of the curve between the No-Load conditions and the two load conditions. Specifically, target activation under No-Load was relatively steeper than under the Linguistic- and Spatial-Load conditions preceding peak activation and more sustained following peak activation.[4]

3.2.2. *Peak latency*

The generalized linear mixed-effects model for target peak latency included a fixed effect of Load (No-Load 0 vs. Spatial-Load 1 and No-Load 0 vs. Linguistic-Load 1) and a random intercept for subject. Unsurprisingly, the areas under the curves for target fixations were substantially greater than those of competitors and controls, and the overall timecourse of target fixations was relatively comparable across load conditions. Peak latencies identified using the half-area method did not yield any effects of load on peak latencies of the target ($ps > 0.05$). In order to allow for more fine-grained distinctions between load conditions, peak latencies for target fixations were identified by averaging the times corresponding to fixation proportions within 5% of the maximum amplitude (see supplementary materials for additional details). Model comparisons revealed a significant effect of Load ($\chi^2(2) = 13.48$, $p = .002$). Parameter estimates from the model summary showed that fixations to targets peaked earlier in the No-Load condition compared to the Spatial-Load condition (*Estimate* = 74.54, *SE* = 22.45, $z = 3.32$, $p < .001$, 95% CI [30.53, 118.55]), and to a lesser extent compared to the Linguistic-Load condition (*Estimate* = 47.52, *SE* = 21.94, $z = 2.17$, $p = .030$, 95% CI [4.51, 90.53]; Fig. 4, *right*).[5] In sum, results from the target peak shape and latency analyses suggest that Spatial-Load resulted in relatively fewer target fixations compared to the No-Load and Linguistic-Load conditions, as well as more gradual target activation leading up to the peak and less sustained activation following the peak. The magnitude of target fixations under Linguistic-Load was slightly greater than that of the No-Load condition, but were more gradual preceding peak activation and less sustained following peak activation.

4. Discussion

The results from Experiment 1 demonstrate that phonological competition emerges even under linguistic and spatial working memory load, confirming the robustness of language activation during visual processing. Importantly, the phonological competition observed in the dual-task conditions cannot be attributed to participants ignoring instructions to mentally rehearse the digit strings or grid patterns, as all search trials associated with incorrect responses on the secondary tasks were excluded from the analyses. Though there is reason to expect that the effects of linguistic load on phonological competition would be greater had participants been instructed to vocalize the digit strings out loud, the present study reveals that subvocal rehearsal of task-irrelevant verbal stimuli does not prevent the activation of linguistic labels during visual search.

Nevertheless, both linguistic and spatial memory loads attenuated the magnitude and rate of language-based interference, suggesting that working memory does play a role in linguistically-mediated visual orienting. Compared to No-Load, the addition of linguistic load primarily reduced the overall difference in competitor vs. control peak amplitudes, while having a relatively smaller impact on the timing and shape of the fixation curves. It may therefore be the case that while initial activation of object labels is relatively automatic, memory load impacts the extent and duration of competitor activation. One unexpected pattern observed under Linguistic-Load, however, is that the smaller magnitude of competition relative to the No-Load condition appeared to be driven primarily by increased looks to the control object rather than a reduction in competitor fixations. Though follow-up pairwise comparisons confirmed that the overall reduction in competitor fixations (*Estimate* = −0.20, *SE* = 0.02, $z = -9.93$, $p < .001$; effect of Linguistic-Load on the intercept) was, in fact, greater than the corresponding increase in control fixations (*Estimate* = 0.12, *SE* = 0.02, $z = 5.95$, $p < .001$), the effect of load on peak amplitude was greater for controls than competitors. This apparent discrepancy for the overall proportion vs. peak amplitude of competitor and control fixations can be explained by the relatively steeper and narrower curvature of linguistic competitor fixations around the peak, suggesting that competitor activation was robust, but less sustained compared to the No-Load condition. This rapid rise and fall of competitor fixations under Linguistic-Load may also help explain the overall increase in peak amplitudes for control items, as the quicker disengagement from the competitor may have enhanced attentional processing of other objects in the display.

Like Linguistic-Load, Spatial-Load attenuated the difference in competitor vs. control peak amplitudes (and did so to a greater extent), while also affecting the relative timing and shape of fixations compared to No-Load. Specifically, phonological competition emerged in the form of more gradual and sustained competitor (vs. control) fixations under Spatial- than No-Load, which also contributed to the relatively later peak latencies for competitors compared to controls. In other words, while competitor and control fixations rose at similar rates, the time at

[4] All effects remained significant when block order was included in the model. There was also a significant interaction between block order and load ($\chi^2(2) = 1542.1$, $p < .001$). Pairwise comparisons indicate that there were more looks to the target in later compared to earlier blocks under No-Load (*Estimate* = 0.07, *SE* = 0.01, $t = 9.59$, $p < .001$), while there were fewer looks to the target in later compared to earlier blocks under Linguistic- (*Estimate* = −0.05, *SE* = 0.01, $t = -6.07$, $p < .001$) and Spatial-Load (*Estimate* = −0.36, *SE* = 0.01, $t = -49.67$, $p < .001$).

[5] The same pattern was observed when block order was included as a covariate, with target fixations peaking earlier under No-Load compared to Spatial- ($p < .001$) and Linguistic-Load ($p = .027$). There were no effects of block order ($ps > 0.05$).

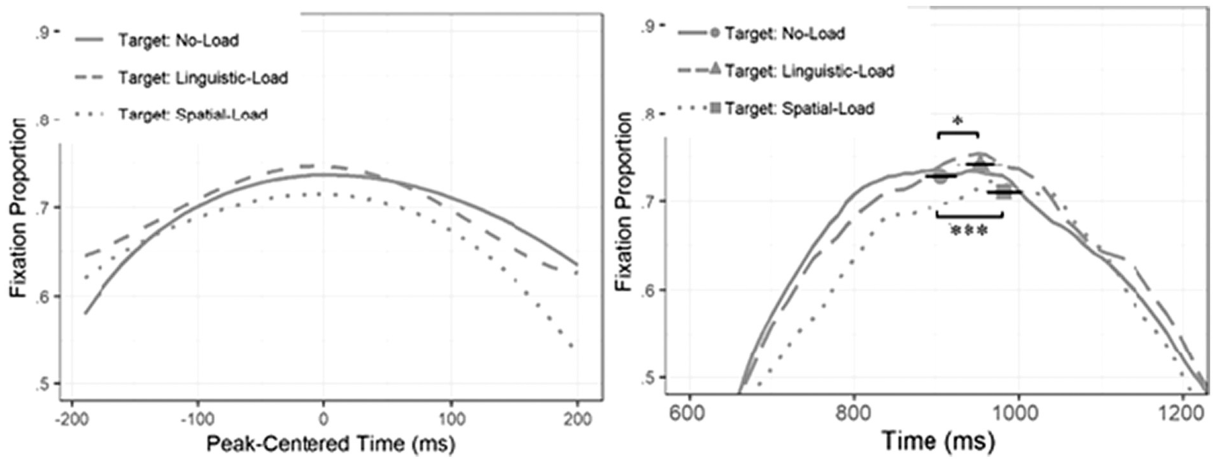

Fig. 4. *Left:* Peak shapes revealed reduced target activation in the Spatial-Load condition (dotted line) relative to No-Load (solid line), particularly at and following peak onset (0 ms). Overall fixations to the target were greater under Linguistic-Load (dashed line) than No-Load (solid line). Lines represent best fit logistic regression models. Timecourses were re-centered for each subject with 0 ms corresponding to peak activation for each condition. *Right:* Peak latencies revealed that target fixations peaked later under Spatial-Load (dotted lines, square), and to a lesser extent under Linguistic-Load (dashed lines, triangle) relative to No-Load (solid lines, circle). Curves represent observed data, dots represent peak latency, and horizontal lines represent standard error.

Table 3
Growth curve analyses for visual fixations to target pictures in varying load conditions.

Fixed Effect	Estimate	SE	95% CI	z value
Intercept	0.91	0.14	[0.64, 1.18]	6.65***
Linear	0.27	0.33	[−0.38, 0.91]	0.82
Quadratic	−1.28	0.10	[−1.47, −1.09]	−12.91***
Cubic	0.19	0.10	[0, 0.38]	1.96
Quartic	−0.04	0.05	[−0.14, 0.06]	−0.85
Intercept:Spatial	−0.13	0.01	[−0.14, −0.11]	−19.99***
Linear:Spatial	−0.79	0.04	[−0.86, −0.71]	−20.1***
Quadratic:Spatial	−0.03	0.04	[−0.11, 0.05]	−0.7
Cubic:Spatial	−0.27	0.04	[−0.34, −0.19]	−6.73***
Quartic:Spatial	0.02	0.04	[−0.06, 0.1]	0.47
Intercept:Linguistic	0.02	0.01	[0, 0.03]	2.31*
Linear:Linguistic	−0.71	0.04	[−0.79, −0.64]	−17.87***
Quadratic:Linguistic	−0.01	0.04	[−0.09, 0.07]	−0.26
Cubic:Linguistic	−0.12	0.04	[−0.2, −0.04]	−3.05**
Quartic:Linguistic	0.29	0.04	[0.21, 0.37]	7.25***

Note: Model estimates for peak shape are given on the logit scale. *** $p < .001$, ** $p < .01$, * $p < .05$.

which fixations began to decline was later for competitors than controls (resulting in later peak latencies).

The attenuated competition observed under linguistic and spatial load is consistent with Huettig et al.'s (2011) proposal that working memory supports the integration of characteristics associated with the visual objects, including linguistic labels, visual features, and their spatial location in the visual scene. In order to experience interference from seeing a *flag* when searching for a *flower,* one must encode the visual form of each object, retrieve each associated label, and register their phonological similarity, either through the integration of representations in working memory or implicitly through spreading activation. In either case, in order for this process to manifest as visual fixations to competitor objects, one must also encode the arbitrary spatial location of the competitor and bind it with the distracting characteristics. Taxing linguistic working memory may therefore reduce the salience of phonological features that could influence *what* draws a look (resulting in less sustained competitor fixations), while taxing spatial working memory may impede the recognition of *where* to look.

These proposed mechanisms are further supported by our finding that linguistic load had a relatively minor impact on target fixations, as visual features alone should motivate looks to its location. Spatial load, on the other hand, delayed both competitor and target fixations, as would be expected if spatial working memory supported the use and maintenance of information regarding the objects' locations in the search display.

As noted, however, an alternative (or likely additional) function of working memory in visual search is the maintenance of the target template between the time when the target is acquired (i.e., target preview onset) and when it must be identified (i.e., search display onset). In the current paradigm, the observed effects of spatial load are unlikely to be a result of impaired template maintenance as participants did not have access to useful spatial information until search display onset. Though it could be argued that the maintenance of the spatial grid pattern impeded the maintenance of the visual target template, there is evidence that visual search efficiency is unaffected by a concurrent visual working memory load in which color and shape information of particular objects must be held in memory (Woodman, Vogel, & Luck, 2001). The effects of linguistic load, on the other hand, could conceivably have resulted from reduced (but not entirely inhibited) subvocal rehearsal of the target label while waiting for the search display. Experiment 2 was therefore designed to investigate this possibility and further assess the robustness and automaticity of language activation during visual processing by manipulating the *need* to maintain a target template in memory.

5. Experiment 2: The influence of memory *demands* on language activation

There is substantial evidence that visual, linguistic, and semantic biases in visual search can vary as a function of task-demands (e.g., de Groot, Huettig, & Olivers, 2017; Downing, 2000; Soto & Humphreys, 2007; Zelinsky & Murphy, 2000). Some evidence suggests that activation of linguistic features may be contingent upon the explicit need to encode visual objects into memory. When subjects are instructed to study a visual display in preparation for a short-term memory test, they spend longer looking at objects with multi-syllable names than at objects with single-syllable names (Noizet & Pynte, 1976; Zelinsky & Murphy, 2000). However, Zelinsky and Murphy (2000) found that this looking pattern did not occur when subjects completed a simpler visual search task (Zelinsky & Murphy, 2000). Therefore, when memory for the linguistic label is not required and inner speech is not employed, the linguistic forms of visually-presented objects may not be activated. Though the results from Experiment 1 demonstrated that language-based competition can emerge even when verbal working memory is taxed, the possibility remains that participants had sufficient cognitive resources and motivation to subvocally rehearse the target label in order to maintain a linguistic template of the object during the delay preceding the onset of the search display.

To determine whether language influences visual scene processing even when it is not needed to manage temporal demands, Experiment 2 utilized the same non-linguistic search task from Experiment 1, but with an explicit manipulation of the duration of time that the target must be held in memory. In other words, rather than limiting the capacity to hold the target label in memory through the addition of a second task, the present study limits the need to hold the target in memory altogether by reducing or even eliminating the time between the onset of the target cue and the search display. In Long- and Short-Delay conditions, participants were presented with a target object which was then removed, requiring them to remember the target object for a long or short period before visual search. In a No-Delay condition, the target cue and search array appeared simultaneously, eliminating all demands on working memory for the maintenance of a target template. To the extent that language activation is an automatic and pervasive process that is not contingent upon the need to remember or rehearse an object label, we predicted that participants would experience competition (indexed by looks to objects with phonologically-similar names; e.g., *flower-flag*) across all three delay conditions.

6. Method

6.1. Participants

Twenty-four native English speakers (9 males) participated in Study 2; participants in Study 2 did not participate in Study 1. Participants ranged in age from 19 to 28 years (mean age = 22.46, *SD* = 2.57) and reported normal or corrected-normal vision and no history of hearing impairments. Native English status was confirmed by self-report measures on the *Language Experience and Proficiency Questionnaire* (Marian et al., 2007). See supplementary materials for participant demographics and cognitive information.

6.2. Design and materials

Study 2 was conceptualized as a 2 × 3 repeated-measures design, with item type (Competitor, Control) and target delay (Long, Short, None) as within-subject variables.

The same set of stimuli used in Study 1 were used in Study 2. Also as in Study 1, the experiment included 90 critical trials and 135 filler trials designed to mask the phonological manipulation. Unlike Study 1, trials were not blocked by condition; Long-, Short-, and No-Delay trials were interspersed and each of the 30 critical and 45 filler stimuli sets was repeated across conditions; participants were not informed about the duration of the delay prior to trial onset. Stimulus sets were not repeated on consecutive trials and the position of items was randomized across trial types. Trials were arranged in a pseudo-randomized order that was fixed between participants; half of the participants received the stimuli in the reverse order.

All trials began with a fixation cross, which was displayed for 500 ms. In the two *Delay conditions,* participants were then presented with a preview screen containing only the target picture for 1000 ms. To determine whether the effects of phonological competition observed in the No-Load condition in Experiment 1 are replicated when the time during which the target must be held in memory is reduced (from 1000 ms in Experiment 1), the target preview display in the two Delay conditions of Experiment 2 was followed by a fixation cross for either a long delay period of 750 ms (*Long-Delay condition*), or a short delay period of 250 ms (*Short-Delay condition*). The fixation cross was then replaced by the four-object search display. In the *No-Delay condition*, participants were not shown a preview of the target; instead, the target appeared in the center of the screen at the same time as the four-object search display (i.e., target, competitor, control, and filler) and remained on the screen throughout the duration of the trial. In all conditions, the search display remained on the screen until the participant provided a response. See Fig. 5 for a sample trial structure. In all conditions, participants were instructed to locate the target from the search array and to click on it as quickly as possible.

6.3. Apparatus and procedure

The apparatus and eye-tracking procedure were the same as in Experiment 1. Following the experiment, participants were asked to provide names for each of the target and competitor items seen throughout the study and images that were named incorrectly or were unnamed were discarded individually for each participant on a trial-by-trial basis before further analyses (6.94% of trials discarded).

6.4. Data analysis

Eye-movements were analyzed using complementary analyses of peak shape and latency using logistic and linear mixed effects regressions, respectively. As in Experiment 1, the number of fixations to critical objects (targets, competitors, controls) was first summed across trials at each time point (between 0 ms and 2000 ms following display onset), and then divided by the total number of trials (30 per condition). The calculated proportion of looks to each critical object was therefore independent of any additional fixations that were made to other areas of the display, including to the central target cue in the No-Delay condition. Trials that were responded to incorrectly (3.5% of trials; mean accuracy = 96.5%, *SD* = 4.15), or trials in which the response time was two standard deviations above or below the mean (3.9% of trials; mean RT = 1284.04 ms, *SD* = 120.62) were excluded from the analysis. There were no significant differences between Long- and Short-Delay conditions for either accuracy ($p = .211$) or response time ($p = .792$). The No-Delay condition had marginally higher accuracy (M = 97.4%, SD = 3.83) compared to the Long- (M = 95.2%, SD = 5.09; $t(46) = 2.26$, p = .073), but not Short-Delay condition (M = 96.9%, SD = 3.17, $t(46)$ = 0.54, p = .851), and significantly longer response times (M = 1339.33 ms, SD = 110.45) compared to both Long- (M = 1251.6 ms, SD = 112.18; $t(46) = 6.50$, $p < .0001$) and Short-Delay conditions (M = 1261.18 ms, SD = 123.69; $t(46) = 5.84$, $p < .0001$).

7. Results

7.1. Competitor and control fixations

7.1.1. Peak shape

The effects of phonological competition and delay condition were examined using growth curve analysis, with fixed effects of Competition (Control −0.5 vs. Competitor +0.5), Delay condition, and their interaction on all time terms. As we were interested in examining the effect of no delay relative to any delay, as well as the impact of different lengths of delays relative to each other, Delay condition was modeled with a Delay-Any contrast (No-Delay [−0.67] vs. the average of Short- [+0.33] and Long-Delay [+0.33]), and a Delay-Length contrast (Short- [−0.5] vs. Long-Delay [+0.5]). The model additionally included a random intercept for subject and by-subject random slopes for all time terms. As in Experiment 1, random effects of item were not included as the proportion of fixations at each time point were determined by aggregating across trials.

Model comparisons between full and depleted models revealed significant main effects of Competition (Likelihood Ratio Test, $\chi^2(1)$ = 405.39, $p < .001$) and Delay ($\chi^2(2) = 1149.49$, $p < .001$), indicating that the proportion of fixations differed between competitors and controls and across delay conditions. There were also significant interactions between Competition and Delay on the intercept ($\chi^2(2)$ = 430.59, $p <$ .001), and on the linear ($\chi^2(2)$ = 27.83, $p <$ .001), quadratic ($\chi^2(2)$ =

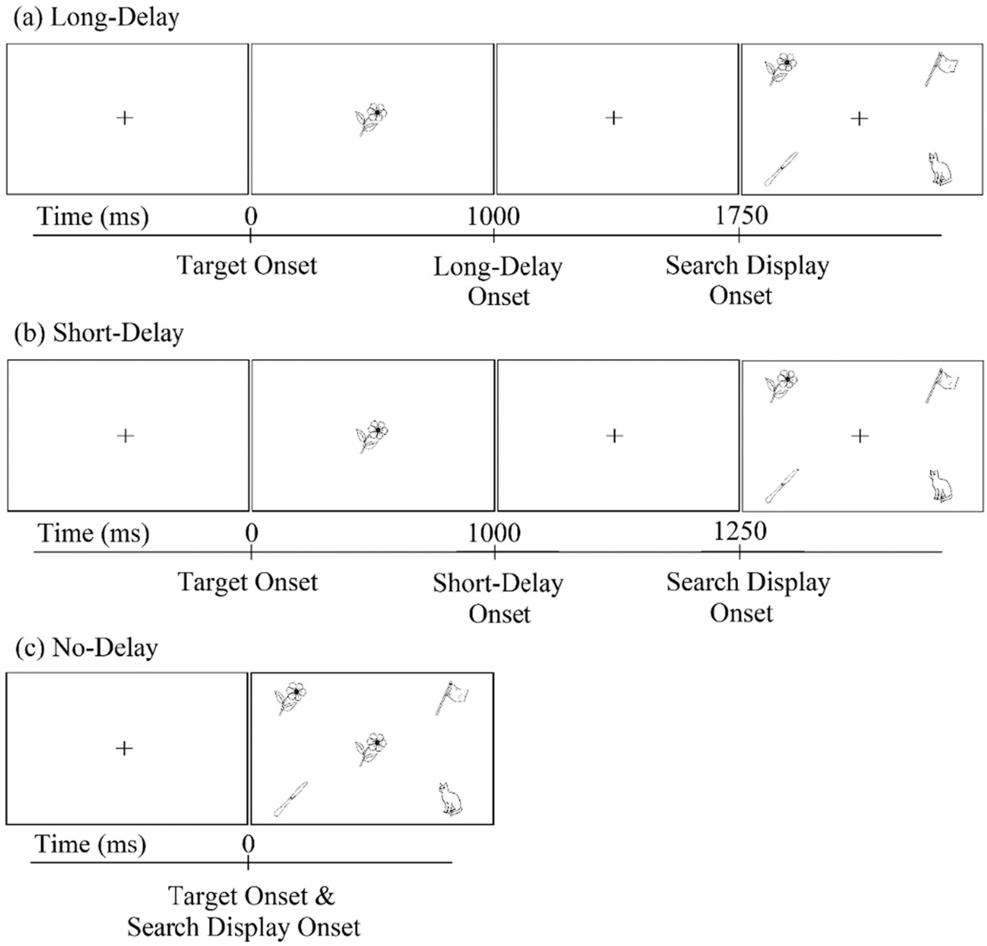

Fig. 5. Sample trial structure for the Long (a), Short (b), and No-Delay (c) conditions. The target (e.g., *flower*) was present in the search display along with a phonological competitor (e.g., *flag*) and control and filler items (e.g., *knife, cat*) which did not overlap phonologically. Participants were instructed to click on the target object as quickly as possible.

125.22, $p < .001$), and cubic ($\chi^2(2) = 10.03$, $p = .007$) time terms, indicating that the effects of competition differed across the delay conditions. Parameter estimates (Table 4) reveal that the effects of competition on the intercept and on the linear and quadratic time terms significantly differed between the No-Delay condition and the averaged Short- and Long-Delay conditions (i.e., the Delay-Any contrast). This suggests that previewing the target picture prior to the onset of the search display significantly affected the magnitude of competitor (vs. control) activation, as well as the shape of competitor (vs. control) activation over time, particularly in the center of the window. The length of time of the delay after previewing the target (i.e., Delay-Length contrast) had a large impact on viewing behavior during the visual search task. A significant effect of Delay-Length on the intercept captures the overall increase (averaged across competitors and controls) in fixations in the Long-Delay condition compared to the Short-Delay condition, while an effect of Delay-Length on the quadratic term reflects the steeper curve around the peak in the center of the window in the Long-Delay condition compared to the Short-Delay condition. Effects of competition on the quadratic and cubic time terms also significantly differed between the Long- and Short-Delay conditions, indicating that the rise and fall of competitor (vs. control) fixations varied as a function of Delay-Length, both in the center of the window (quadratic) and towards the tail ends (cubic).

Follow-up analyses examined the simple effects of phonological competition on the rate and magnitude of visual fixations for each of the three Delay conditions. As in Experiment 1, Tukey-adjusted pairwise comparisons of competitor and control fixations in each delay condition were made for each of the time terms. Model estimates for the No-Delay, Short-Delay, and Long-Delay conditions are presented in Table 5 and the proportions of competitor and control fixations in each condition are depicted in Fig. 6.

There was a significant effect of phonological competition on the intercept in the No-Delay condition, indicating that the proportion of competitor fixations exceeded that of controls. An effect of competition on the intercept was also found in the Short-Delay condition, though to a lesser extent, and no effect of competition on the intercept was found in the Long-Delay condition. There were additionally significant effects of competition on the quadratic term in all three conditions, reflecting differences in the rate at which competitor vs. control fixations increased and decreased around the peak (see Fig. 6). The previously observed Competition x Delay-Any effect on the quadratic term (see Table 4) reflects the relatively larger effect of competition on the quadratic term in the No-Delay condition, which manifested as more gradual and sustained fixations around the peak for competitors relative to controls. Visual inspection suggests that competitor fixations in the two delay conditions were more sharply peaked relative to No-Delay, and negatively skewed relative to controls.

Significant effects of competition were found on the cubic term in the Long-Delay condition and No-Delay conditions, reflecting differences in competitor and control fixations at the tail ends of the windows. The pattern of competitor and control fixations at the tail ends were relatively more similar in the Short-Delay condition, as reflected by the null effect of competition on the cubic term, as well as the previously observed Competition x Delay-Length effect on the cubic term (see

Table 4
Growth curve analysis of fixation peak shapes to competitor and control pictures in varying delay conditions.

Fixed Effect	Estimate	SE	95% CI	*z* value	*p*
Intercept	−5.24	0.48	[−6.18, −4.31]	−10.98	<0.001***
Linear	5.18	2.22	[0.82, 9.53]	2.33	0.020*
Quadratic	−8.08	1.70	[−11.41, −4.74]	−4.74	<0.001***
Cubic	1.61	0.72	[0.19, 3.02]	2.23	0.026*
Quartic	−0.77	0.45	[−1.65, 0.12]	−1.70	0.089
Intercept: Competition	0.29	0.01	[0.26, 0.32]	19.63	<0.001***
Linear:Competition	−0.02	0.12	[−0.24, 0.21]	−0.16	0.876
Quadratic: Competition	1.96	0.11	[1.74, 2.18]	17.27	<0.001***
Cubic:Competition	−0.82	0.10	[−1.01, −0.62]	−8.18	<0.001***
Quartic: Competition	−0.06	0.10	[−0.25, 0.13]	−0.59	0.556
Intercept:Delay-Any	0.00	0.02	[−0.04, 0.03]	−0.27	0.786
Linear:Delay-Any	−0.48	0.13	[−0.74, −0.21]	−3.52	<0.001***
Quadratic:Delay-Any	1.71	0.14	[1.44, 1.97]	12.55	<0.001***
Cubic:Delay-Any	−0.63	0.11	[−0.85, −0.41]	−5.59	<0.001***
Quartic:Delay-Any	1.13	0.11	[0.92, 1.35]	10.31	<0.001***
Intercept:Comp: Delay-Any	−0.65	0.03	[−0.71, −0.58]	−19.19	<0.001***
Linear:Comp:Delay-Any	1.40	0.27	[0.87, 1.93]	5.19	<0.001***
Quadratic:Comp: Delay-Any	−2.80	0.27	[−3.34, −2.27]	−10.32	<0.001***
Cubic:Comp:Delay-Any	−0.02	0.23	[−0.46, 0.42]	−0.08	0.933
Quartic:Comp: Delay-Any	0.27	0.22	[−0.16, 0.7]	1.23	0.219
Intercept:Delay-Length	0.58	0.02	[0.54, 0.61]	33.41	<0.001***
Linear:Delay-Length	−0.09	0.13	[−0.34, 0.15]	−0.75	0.455
Quadratic:Delay-Length	−0.57	0.12	[−0.8, −0.33]	−4.78	<0.001***
Cubic:Delay-Length	−0.20	0.11	[−0.43, 0.02]	−1.81	0.070
Quartic:Delay-Length	−0.11	0.11	[−0.33, 0.12]	−0.94	0.346
Intercept:Comp: Delay-Length	−0.06	0.03	[−0.13, 0.01]	−1.76	0.078
Linear:Comp:Delay-Length	−0.19	0.25	[−0.68, 0.31]	−0.74	0.459
Quadratic:Comp: Delay-Length	0.74	0.24	[0.27, 1.2]	3.11	0.002**
Cubic:Comp:Delay-Length	−0.71	0.23	[−1.15, −0.27]	−3.14	0.002**
Quartic:Comp: Delay-Length	0.42	0.23	[−0.02, 0.87]	1.87	0.061

Note: Model estimates are given on the logit scale. *** $p < .001$, ** $p < .01$.

Table 4). In sum, statistical analyses and visual inspection suggest that when participants were given an opportunity to preview the target (i.e., Short-Delay or Long-Delay conditions), competitor activation was more sharply peaked and negatively skewed, suggesting low initial levels of activation followed by a swift activation and resolution of competition. When there was no preview of the target, competitor activation was sustained at peak activation for a longer period of time.

7.1.2. Peak latency

To examine the effects of phonological competition and delay condition on the timing of peak fixations, peak latencies were analyzed using a generalized linear mixed effect regression with an inverse

Table 5
Growth curve analysis of fixation peak shapes to competitor and control pictures in varying delay conditions.

Fixed Effect	Estimate	SE	95% CI	*z* value
A) Long Delay				
Intercept:Competition	0.05	0.02	[−0.01, 0.11]	2.3
Linear:Competition	0.35	0.16	[−0.10, 0.81]	2.24
Quadratic:Competition	1.4	0.15	[0.97, 1.82]	9.29***
Cubic:Competition	−1.18	0.14	[−1.58, −0.78]	−8.36***
Quartic:Competition	0.25	0.14	[−0.16, 0.65]	1.74
B) Short Delay				
Intercept:Competition	0.11	0.03	[0.03, 0.19]	4.04**
Linear:Competition	0.54	0.2	[−0.02, 1.10]	2.78
Quadratic:Competition	0.66	0.18	[0.14, 1.18]	3.59**
Cubic:Competition	−0.47	0.18	[−0.97, 0.04]	−2.64
Quartic:Competition	−0.18	0.18	[−0.69, 0.33]	−1.01
C) No Delay				
Intercept:Competition	0.72	0.03	[0.64, 0.81]	25.11***
Linear:Competition	−0.95	0.24	[−1.63, −0.27]	−3.99**
Quadratic:Competition	3.83	0.25	[3.13, 4.53]	15.66***
Cubic:Competition	−0.8	0.2	[−1.36, −0.25]	−4.11***
Quartic:Competition	−0.24	0.19	[−0.78, 0.30]	−1.26

Note: Model estimates are given on the logit scale. P-values were corrected for multiple comparisons using the Tukey method. *** $p < .001$, ** $p < .01$, * $p < .05$.

gaussian and identity link. The model included fixed effects of Competition (Control −0.5 vs. Competitor 0.5), Delay (Delay-Any: No-Delay −0.67 vs. Short-Delay 0.33 and Long-Delay 0.33; Delay-Length: Short-Delay −0.5 vs. Long-Delay 0.5), and their interactions, as well as a random intercept for subject. Model comparisons revealed a significant main effect of Delay condition ($\chi^2(2) = 10.26$, $p = .011$), as well as a significant interaction between Competition and Delay ($\chi^2(2) = 21.77$, $p < .001$). Parameter estimates revealed a significant effect of Delay-Any (*Estimate* = −125.93, *SE* = 22.54, $z = -5.59$, $p < .001$, 95% CI [−170.11, −81.75]), with later peak latencies (averaged across competitors and controls) for the No-Delay condition compared to the average of the Long-Delay and the Short-Delay conditions. Pairwise comparisons on the marginal means showed that fixations peaked significantly later in the No-Delay condition (692.2 ms) than in both the Long-Delay (564.5 ms, $z = -5.15$, $p < .001$) and Short-Delay (567.9 ms, $z = 4.99$, $p < .001$) conditions (see Fig. 7).

The effect of Competition also differed between the No-Delay condition and the average of the Short- and Long-Delay conditions (*Estimate* = −138.05, *SE* = 45.08, $z = -3.06$, $p = .002$, 95% CI [−226.4, −49.7]). Pairwise comparisons revealed that peak latencies were earlier for competitor pictures compared to controls by 66.2 ms in the Long-Delay condition ($z = -2.25$, $p = .024$) and by 79.8 ms in the Short-Delay condition ($z = -2.68$, $p = .007$), whereas peak latencies did not differ between competitor and control fixations in the No-Delay condition ($z = 1.63$, $p = .103$). Long- and Short-Delay conditions did not differ from each other in either overall peak latency (averaged across competitors and controls) or the effect of Competition (*ps* > 0.05).

7.2. Target fixations

7.2.1. Peak shape

In order to examine the effect of Delay time on the shape of target fixations, a growth curve model was designed with a fixed effect of Delay plus interactions on all time terms, as well as random effects of subject on all time terms. Model comparisons revealed a significant effect of Delay on the intercept ($\chi^2(2) = 152.64$, $p < .001$), as well as on the linear ($\chi^2(2) = 174.41$, $p < .001$), quadratic ($\chi^2(2) = 149.95$, $p < .001$), cubic ($\chi^2(2) = 112.48$, $p < .001$), and quartic ($\chi^2(2) = 25.52$, $p < .001$) time terms. Parameter estimates (Table 6) showed a significant effect of Delay-Length on the intercept, indicating that target activation was greater under Long-Delay than Short-Delay (Fig. 8, *left*). We additionally

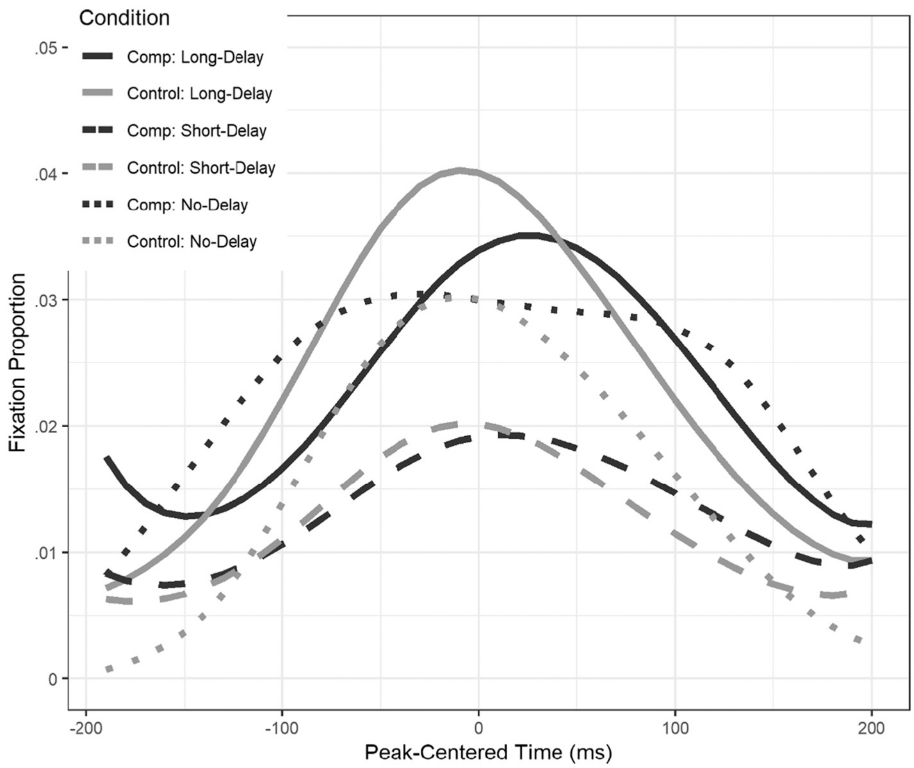

Fig. 6. Visual fixation peak shapes. Significant effects of competition were found for the intercept in the No-Delay (dotted lines) and Short-Delay conditions (dashed lines), indicating that competitors (black) were fixated more than controls (gray). When the search target was presented at the same time as the visual display (No-Delay, dotted lines), fixation peaks were sustained longer (effects of competition on the linear, quadratic, and cubic terms), whereas advance preview of the target resulted in sharper peaks (effects of competition on the quadratic term for both Short- and Long-Delay). Longer delays between presentation of the search target and the display (solid lines) resulted in greater fixations than a short delay (dashed lines; main effect of Delay-Length on the intercept). Lines represent best fit logistic regression model estimates. Timecourses were re-centered for each subject with 0 ms corresponding to peak activation for each condition.

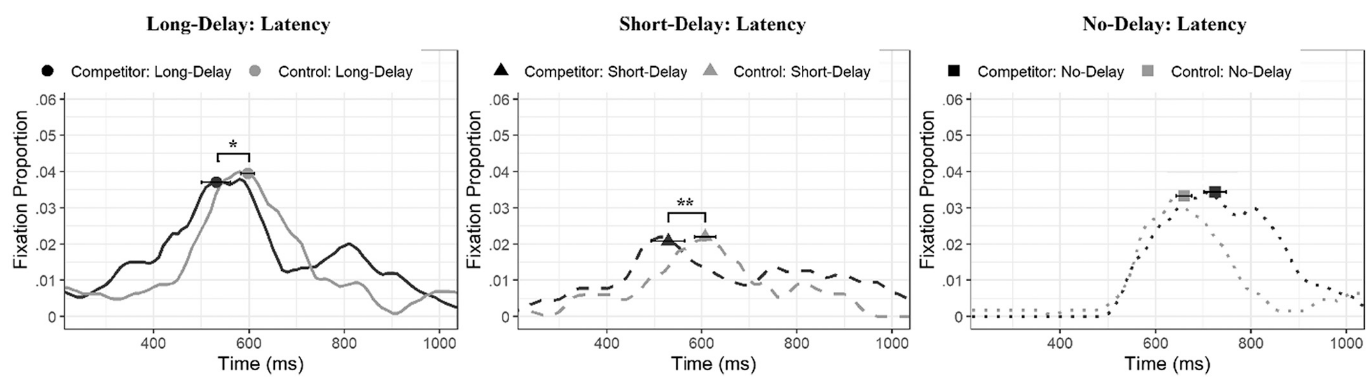

Fig. 7. Eye-tracking latency across delay times. Peak latencies revealed earlier activation of phonological competitors (black) than controls (gray) in Long-Delay (solid lines, circles) and Short-Delay (dashed lines, triangles) conditions. Competitor and control peak latencies did not differ in the No-Delay condition (dotted lines, squares). Curves represent observed data, dots represent peak latency, and horizontal lines represent standard error.

found significant effects of both the Delay-Any and Delay-Length contrasts on the linear, quadratic, cubic, and quartic time terms, the combination of which reflect differences in the steepness of the curves in the center of the window, as well as at the tail ends. Visual inspection indicates that preceding peak onset (0 ms), the rate of target activation was faster in the Long-Delay condition compared to the Short- and No-Delay conditions. Following peak onset, target deactivation was more gradual (i.e., activation was more sustained) in the No-Delay condition relative to the two delay conditions, and deactivation was more gradual in the Long-Delay condition relative to the Short-Delay condition.

7.2.2. Peak latency

In order to examine the effects of Delay condition on target activation, peak latencies for target fixations were analyzed using a generalized linear mixed-effects model with an inverse gaussian and identity link. Peak latencies identified using the half-area method revealed a significant main effect of Delay ($\chi^2(2) = 17.61$, $p < .001$). Parameter estimates showed a significant effect of the Delay-Any contrast (*Estimate* $= -80.5$, $SE = 10.95$, $z = -7.35$, $p < .001$, 95% CI [−101.97, −59.04]), indicating that the latency of peak activation differed between the No-Delay condition and the average of the Long- and Short-Delay conditions. Tukey-adjusted pairwise comparisons of model estimates revealed that fixations to the target in the No-Delay condition (1119 ms) peaked significantly later than in both the Long-Delay (1041 ms; $z = 6.29$, $p < .001$) and Short-Delay conditions (1036 ms; $z = 6.67$, $p < .001$). There was no effect of the Delay-Length contrast, indicating that peak latencies did not differ between the Long- and Short-Delay conditions (*Estimate* $= 4.49$, $SE = 11.70$, $z = 0.38$, $p = .701$, 95% CI [−18.45, 27.43]. Similar effects were found using the peak amplitude method reported for target fixations in Experiment 1. There was a significant main effect of Delay ($\chi^2(2) = 11.86$, $p = .005$), a significant simple effect of the Delay-Any contrast (*Estimate* $= -67.0$, $SE = 25.16$, $z = -2.66$, $p = .007$, 95% CI [−116.29, −17.69]), and no effect of Delay-Length (*Estimate* $= -8.16$, $SE = 27.25$, $z = -0.30$, $p = .764$, 95% CI [−61.57, 45.24]; see Fig. 8,

Table 6
Growth curve analyses for visual fixations to target pictures in varying delay conditions.

Fixed Effect	Estimate	SE	95% CI	z value
Intercept	1.17	0.16	[0.85, 1.49]	7.25***
Linear	−0.13	0.3	[−0.73, 0.46]	−0.44
Quadratic	−1.33	0.14	[−1.61, −1.06]	−9.4***
Cubic	−0.04	0.08	[−0.19, 0.12]	−0.5
Quartic	0.05	0.04	[−0.04, 0.13]	1.11
Intercept:Delay-Any	0.01	0.01	[0, 0.02]	1.43
Linear:Delay-Any	−0.48	0.04	[−0.55, −0.41]	−12.96***
Quadratic:Delay-Any	−0.45	0.04	[−0.52, −0.38]	−12.04***
Cubic:Delay-Any	−0.18	0.04	[−0.25, −0.1]	−4.72***
Quartic:Delay-Any	0.12	0.04	[0.04, 0.19]	3.11**
Intercept:Delay-Length	0.09	0.01	[0.07, 0.1]	12.28***
Linear:Delay- Length	0.1	0.04	[0.02, 0.19]	2.41*
Quadratic:Delay- Length	−0.1	0.04	[−0.19, −0.02]	−2.37*
Cubic:Delay- Length	0.41	0.04	[0.32, 0.49]	9.44***
Quartic:Delay- Length	0.17	0.04	[0.09, 0.26]	4.01***

Note: Model estimates for peak shape are given on the logit scale. *** $p < .001$, ** $p < .01$, * $p < .05$.

right). Pairwise comparisons revealed that peak latencies were later in the No-Delay condition (1113 ms) than in the Long-Delay (1042 ms, $z = 2.49$, $p = .012$) and Short-Delay conditions (1050 ms, $z = 2.19$, $p = .028$).

Taken together, the results from the peak shape and latency analyses suggest that the opportunity to preview the target prior to the onset of the search display resulted in earlier activation of the target, whereas simultaneous presentations of the target and search display resulted in more sustained target activation following peak activation.

8. Discussion

The results from Experiment 2 demonstrated that competition between visual objects occurred even when objects did not need to be remembered. Particularly, in the No-Delay condition, participants in the current study could successfully perform the task using a simple visual-feature match between the centrally-presented target and the four-object search display (e.g., searching for the curve-shaped object when looking for the *bell*). Nevertheless, participants fixated phonological competitors significantly more than non-overlapping controls, and over different timescales. This confirms that language is activated by visual object processing, even in non-linguistic tasks requiring no maintenance of a target in memory.

We saw evidence for language activation during visual processing under all three delay conditions, but there were also differences in how language was activated in each case. When participants were provided with the search cue prior to the search display (i.e., Long-Delay and Short-Delay conditions), phonological competitors were fixated 66–80 ms earlier than non-overlapping controls, but with a similar shape of the rise and fall of activation between competitors and controls. In contrast, when the search cue and the search display were presented simultaneously (i.e., the No-Delay condition), competitors and controls rose in activation at the same time, but competitors peaked later and remained active longer.

Our finding that phonological competitors in the No-Delay condition were initially fixated at comparable rates as control objects is consistent with prior Visual World studies indicating that a sufficient preview period is necessary to elicit preferential fixations to phonological competitors relative to controls. For instance, Huettig and McQueen (2007) observed that when participants passively listened to a phrase that included a target object (e.g., “Eventually she looked at the *beaker* that was in front of her”), fixations to phonological competitors (e.g., a picture of a *beaver*) preceded that of visual competitors (e.g., a *bobbin*), semantic competitors (e.g., a *fork*), and controls (e.g., an *umbrella*) when the visual display was presented at the onset of the sentence (Experiment 1), but not when the visual display was presented only 200 ms prior to the onset of the target word (Experiment 2). Based on the time it took participants to disambiguate the target and phonological competitor labels (~ 190 ms post target word onset), the authors reasoned that shortening the preview time reduced looks to phonological competitors because the acoustic signal would already be inconsistent with the phonological competitor by the time the names of visual objects could be

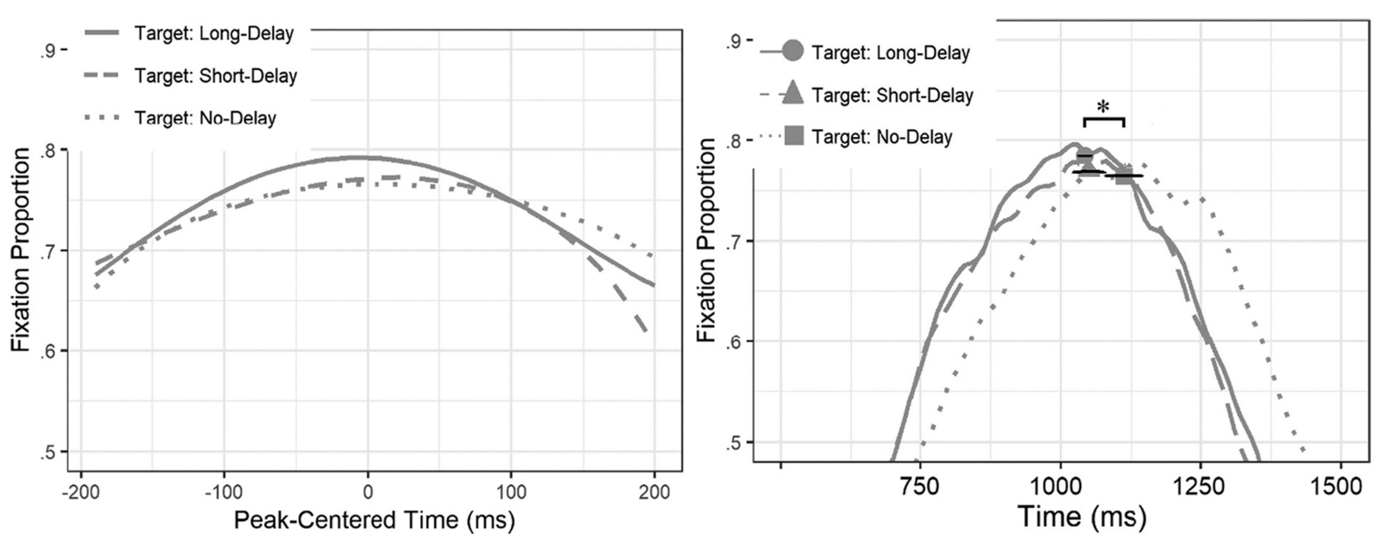

Fig. 8. *Left:* Peak shapes revealed reduced target activation in the Short-Delay condition (dashed line) relative to the Long-Delay condition (solid line; effect of Delay-Length on the intercept). Significant effects of Delay-Any and Delay-Length on the linear, quadratic, cubic, and quartic terms indicate that the steepness of the fixation curves differed between No-Delay and the two delay conditions, as well as between Short- and Long-Delay in the center and tail ends of the window. Visual inspection suggests that target activation was faster preceding peak onset (0 ms) in the Long-Delay condition (solid line) compared to the Short- (dashed line) and No-Delay (dotted line) conditions. Target activation following peak-onset was more sustained in the No-Delay condition relative to the two delay conditions, and activation was more sustained in the Long-Delay condition relative to the Short-Delay condition. Lines represent best fit logistic regression models. Timecourses were re-centered for each subject with 0 ms corresponding to peak activation for each condition. *Right*: Peak latencies revealed that target fixations peaked later with No-Delay (dotted lines, square) relative to Long- (solid lines, circle) and Short-Delays (dashed lines, triangle). The timing of Long- and Short-Delays did not significantly differ. Curves represent observed data, dots represent peak latency, and horizontal lines represent standard error.

retrieved. Ferreira, Foucart, and Engelhardt (2013) similarly observed that participants were able to use visual context and pragmatic knowledge to disambiguate syntactically ambiguous statements (e.g., "Place the book on the chair in the bucket") when a visual display was previewed for 1000 ms, but not 200 ms, prior to an ambiguous instruction. In the case of the present investigation, the timing of competitor fixations relative to controls varied as a function of whether or not the visual *target* was presented prior to or concurrently with the search display. The 1000 ms target preview periods in the Long- and Short-Delay conditions provided participants with ample time to retrieve the label of the target, which could have subsequently facilitated the activation of cohort labels and the salience of the phonological competitor once it was displayed. Presenting the target concurrently with the visual display, on the other hand, appears to have delayed the activation and matching of target and competitor labels, resulting in comparable rates at which competitors were fixated relative to controls.

Notably distinct from Huettig and McQueen (2007), however, reducing the time between display and target onsets did not fully eliminate preferential fixations to phonological competitors. Despite the comparable timing of initial fixations to competitors and controls in the No-Delay condition of the present study, evidence of phonological competition still emerged in the form of more sustained fixations to objects whose labels overlapped with that of the target. Indeed, given that in all conditions, the target could be identified based on visual features alone and was (with the exception of the No-Delay condition) known prior to the onset of the search display, the present findings suggest that ambiguity regarding the correct referent is not necessary for object labels to influence attention during non-linguistic visual search. In other words, advance notice of the target caused participants to fixate phonological competitors *earlier* than controls, whereas simultaneous presentation caused participants to fixate competitors for a *longer duration*.

In addition, the length of the delay between the target search cue and the search display had a noticeable effect on overall viewing behavior. When participants had to remember the search cue for a longer period of time, they fixated all objects, particularly competitors and controls, more than when there was a short delay. This change in overall viewing behavior may be due to degradation of the search cue as the delay increased, resulting in more diffuse attention over the entire search display in the Long-Delay condition.

This interpretation could also help explain the differences in the type of phonological competition observed in the No-Load condition of Experiment 1 (which had a delay of 1000 ms between target presentation and the search display) and the two delay conditions in Experiment 2 (delays of 750 ms and 250 ms). While phonological competition in the No-Load condition of Experiment 1 emerged primarily in terms of the *magnitude* and *duration* of competitor activation relative to controls, competition in the delay conditions of Experiment 2 primarily manifested as *earlier* fixations to competitors relative to controls. Though it should be noted that Experiment 1's No-Load condition is not directly comparable to Experiment 2's two Delay conditions (due to the fact that trials were blocked by condition in Experiment 1, but not in Experiment 2), when all three activation patterns are considered, we see that the effect of phonological competition on peak latency decreases with increasing delay times (see Fig. 9).

The largest latency effect was observed for the Short-Delay condition (Experiment 2; 250 ms delay between target preview and search), in which competitors preceded controls by 79.8 ms, followed by the Long-Delay condition (Experiment 2; 750 ms), with a competitor lead of 66.2 ms, and lastly the No-Load condition (Experiment 1; 1000 ms), with a competitor lead of 26.88 ms. To the extent that longer delays resulted in more diffuse attention over the display at the initial presentation of the search objects, it may have reduced the speed at which shared features of targets and competitors were matched, while increasing the overall activation of the competitor once the match was made. More concretely, it is likely the case that the total level of target label activation resulted from a combination of seeing the visual target during the preview period as well as in the search display itself. With a relatively short delay between preview and search (e.g., 250 ms), the linguistic representation activated during the preview period may remain salient and quickly bias attention towards the phonological competitor. With increasing delay time, the effects of phonological competition may be arising more from a combination of preview activation (which may have degraded over time) as well as search display activation, resulting in relatively delayed, but ultimately greater consideration of the phonological competitor. Though more work is needed to confirm these speculations, the present findings demonstrate that looking at visual objects activates linguistic representations even when there is no need or opportunity for subvocal rehearsal of the target label.

9. General discussion

Results from two non-linguistic visual search tasks demonstrate that the linguistic forms of visually-presented objects become activated, even when subvocal rehearsal of object labels is impeded by a concurrent linguistic memory task, and even when those objects do not need to be remembered. When conducting a visual search in which a competing distractor picture's name (e.g., *flag*) overlapped phonologically with the

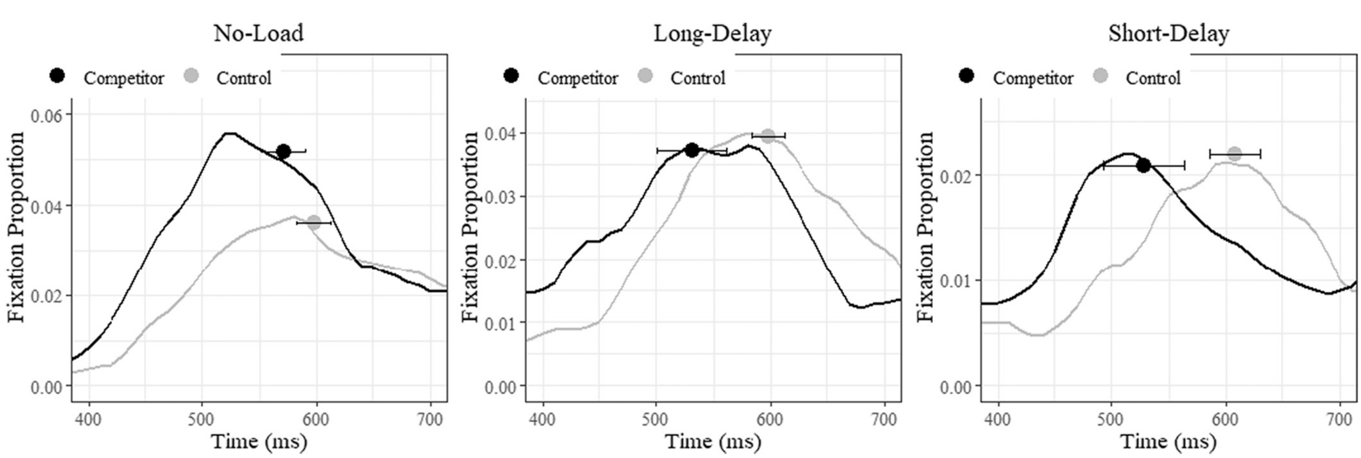

Fig. 9. Competitor (black) and control (gray) fixations in the No-Load (Experiment 1), Long-Delay (Experiment 2), and Short-Delay conditions (Experiment 2). Effects of competition on peak latency increased with shorter delays between the target preview and search display onset, with the largest competitor lead in the Short-Delay condition (delay: 250 ms), followed by the Long-Delay condition (delay: 750 ms), and then the No-Load condition (delay: 1000 ms). Curves represent observed data, dots represent peak latency, and horizontal lines represent standard error.

target's name (e.g., *flower*), participants looked at competitors more often and at different times than control items that did not overlap (e.g., *knife*), even though language input was never provided to participants. This phenomenon was observed regardless of the capacity (Experiment 1) or need (Experiment 2) to meaningfully encode the target label prior to the visual search. Because the target and competitor items employed in this experiment were only related in shared word-initial phonological overlap, and because this phonological overlap was never overtly presented to participants, the difference in looking patterns between competitors and controls can be attributed to activation of items' linguistic forms by visual input alone.

Though our findings revealed significant language activation regardless of memory capacity or demands, the degree and timing of activation did vary across conditions, suggesting that language-vision interactions are likely subject to a combination of top-down attentional guidance and automatic cascaded activation. A similar conclusion was drawn by de Groot, Huettig, and Olivers (2017), who manipulated task-demands by having participants memorize a word that was either relevant (the target itself) or irrelevant (to be recalled in a later memory test) for a subsequent visual search task. They observed that while visual and semantic interference was greater when the memorized word was directly relevant to the search task, participants were biased towards competitors in both conditions. The authors therefore concluded that top-down attentional mechanisms may modulate the weight that is placed on particular representations depending on task-specific goals, but that visual and semantic features associated with linguistic representations can become automatically activated as a result of learned associations. Here, we demonstrate that this activation is bidirectional, such that linguistic information is activated by visual representations, whether or not it serves a function for the task at hand.

Beyond the maintenance of targets and the modulation of activation, our findings from Experiment 1 are consistent with Huettig et al.'s (2011) proposal that working memory plays a key role in binding different features associated with objects during visual search. Though we found that imposing linguistic load had little effect on visual fixations to the target, spatial memory load delayed and reduced looks to both competitor and target objects. While linguistic, visual, and semantic features of an object can become closely associated (and automatically activated) through repeated co-activations, the spatial location of a given object is arbitrary and must be integrated with other relevant features of the target before fixations can be biased towards a specific place. As noted by Huettig et al. (2011), however, the mechanisms whereby spatial information is integrated with long-term memory representations are underspecified in investigations utilizing both visual search and visual world paradigms. The authors therefore outline a model with working memory as the "nexus" at which linguistic, semantic, and visual representations in long term memory can be bound to particular locations. Our finding that spatial load reduced the rate and magnitude of both target and competitor fixations provides support for the feature binding function of working memory.

This pattern contrasts with prior work demonstrating that taxing working memory can *increase* competition (e.g., Walenchok et al., 2016; Zelinsky & Murphy, 2000). Notably, Walenchok et al. (2016) and Zelinsky and Murphy (2000) found that greater memory load was associated with greater phonological competition during non-linguistic visual search tasks that were similar to those utilized in the present experiments. In both cases, however, the increased memory load resulted from the need to maintain multiple visual targets in memory (e.g., three or more different objects, one of which may appear in the search display) rather than the maintenance of unrelated linguistic representations (e.g., a string of digits). In the case of the former, the greater phonological competition observed under higher load may indeed be due to strategic encoding of object labels, something which cannot be entirely ruled out in the present experiment for trials under normal viewing conditions (i.e., No-Load in Experiment 1 and Long- and Short-Delay in Experiment 2). The key findings from the Linguistic-Load (Experiment 1) and No-Delay trials (Experiment 2), however, are that phonological competition is still observed even with limited capacity or need for subvocal rehearsal. In other words, our claim is not that participants do not engage in strategic maintenance of target labels when they have the capacity and motivation to do so, but rather that intentional rehearsal is not a necessary condition for linguistic activation.

One condition that *is* likely necessary, however, is sufficient experience associating the visual object with its linguistic label (see Lupyan et al., 2020). The basis for automatic activation in frameworks such as Huettig and McQueen's (2007) *Cascaded Activation Model* is the strength of associations between visual, linguistic, and semantic features of an object which are built over a lifetime of experience. Though seeing a picture of a *flower* is likely to automatically activate the word "flower" for a native English speaker, we would not necessarily expect to find such an effect with newly acquired or unfamiliar lexical information (Chabal & Marian, 2015). Indeed, while Zelinsky and Murphy (2000) found that participants activated the names associated with eight novel faces when they all needed to be memorized for a later task, there was no indication of linguistic activation when a single face needed to be identified from a subsequent search display. Though on the surface, the conditions of the single face trials resemble that of our No-Load and Long- and Short-Delay trials, a key difference is that Zelinsky and Murphy's (2000) face-name pairings were arbitrary and recently encoded, while our participants had a lifetime of experience associating visual objects with their corresponding word forms.

The idea that information stored in long-term memory can have immediate and automatic impacts on visual processing is not new. Shape (Yee, Huffstetler, & Thompson-Schill, 2011), color (Olivers, 2011), and semantic (Cooper, 1974; Huettig & Altmann, 2005; Yee & Sedivy, 2006) features of images all affect how visual scenes are processed (for a full discussion see Huettig, Mishra, & Olivers, 2012). Though activation of stored linguistic information had been posited (Huettig et al., 2011), it could not be ruled out that language-based effects were attributed to the inherently linguistic nature of most search tasks (Görges et al., 2013; Meyer et al., 2007) or to intentional task-based strategies adopted by participants (Chabal & Marian, 2015). Here, we provide evidence that linguistic form is indeed a feature that is activated by visual scenes, even when strategic language use is unnecessary.

In sum, we show that activation of language during visual processing is robust and persists under a variety of constraints and conditions (Marian, 2023). We suggest that language-based competition is not task-dependent or an artifact of intentional rehearsal strategies, but rather reflects automated activation of linguistic representations. We conclude that language alters fundamental aspects of the human experience and is ubiquitously activated during visual processing.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

Research reported in this publication was supported in part by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R01HD059858 to Viorica Marian. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank two anonymous Reviewers for their helpful comments and suggestions.

Appendix A. Supplementary data

Supplementary materials for this article can be found online at https://doi.org/10.1016/j.cognition.2021.104994.

References

Allopenna, P., 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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., Machler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Bates, E., Andonova, E., D'Amico, S., Jacobsen, T., Kohnert, K., Lu, C., & Pleh, C. (2000). Introducing the CRL international picture naming project (CRL-IPNP). *Center for Research in Language Newsletter, 12*(1).

Bles, M., & Jansma, B. M. (2008). Phonological processing of ignored distractor pictures, an fMRI investigation. *BMC Neuroscience, 9*, 1–14. https://doi.org/10.1186/1471-2202-9-20

Blumenfeld, H. K., & Marian, V. (2005). Covert bilingual language activation through cognate word processing: An eye-tracking study. *In Proceedings of the Annual Meeting of the Cognitive Science Society, 27*(27), 286–291.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html.

Chabal, S., & Marian, V. (2015). Speakers of different languages process the visual world differently. *Journal of Experimental Psychology: General, 144*(3), 539–550. https://doi.org/10.1037/xge0000075

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*(1), 75–84. https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology, 6*, 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

de Groot, F., Huettig, F., & Olivers, C. N. (2017). Language-induced visual and semantic biases in visual search are subject to task requirements. *Visual Cognition, 25*(1–3), 225–240. https://doi.org/10.1080/13506285.2017.1324934

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science, 11*(6), 467–473. https://doi.org/10.1111/1467-9280.00290

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

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

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

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96*(1), 23–32. https://doi.org/10.1016/j.cognition.2004.10.003

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

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica, 137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

Kahneman, D. (1973). *Attention and effort* (Vol. 88). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Kiesel, A., Miller, J., Jolicoeur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology, 45*(2), 250–274. https://doi.org/10.1111/j.1469-8986.2007.00618.x

Klauer, K. C., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology, 43*(4), 648–655. https://doi.org/10.1016/j.jesp.2006.06.003

Lenth, R. V. (2021). Emmeans: Estimated marginal means, aka least-squares means. In *R Package Version 1.7.0*. https://CRAN.R-project.org/package=emmeans.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Logie, R. H., Gilhooly, K. J., & Wynn, V. (2016). Counting on working memory in arithmetic problem solving. *Memory & Cognition, 22*(4), 395–410. https://doi.org/10.3758/BF03200866

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Preprint retrieved from*. https://doi.org/10.31234/osf.io/pztmc

Lupyan, G., & Swingley, D. (2012). Self-directed speech affects visual search performance. *Quarterly Journal of Experimental Psychology, 65*(6), 1068–1085. https://doi.org/10.1080/17470218.2011.647039

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science, 3*(2), 113–139.

Marian, V. (2023). *The Power of Language*. New York: Dutton, Penguin Random House. ISBN: 9780593187074.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marian, V., & Spivey, M. (2003a). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition, 6*(2), 97–115. https://doi.org/10.1017/S1366728903001068

Marian, V., & Spivey, M. (2003b). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173–193. https://doi.org/10.1017/S0142716403000092

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

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

Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*(1), 271–280. https://doi.org/10.1016/j.cognition.2008.02.003

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297. https://doi.org/10.1037/0033-2909.132.2.297

Noizet, G., & Pynte, J. (1976). Implicit labelling and readiness for pronunciation during the perceptual process. *Perception, 5*(2), 217–223. https://doi.org/10.1068/p050217

Olivers, C. N. L. (2011). Long-term visual associations affect attentional guidance. *Acta Psychologica, 137*(2), 243–247. https://doi.org/10.1016/j.actpsy.2010.07.001

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*(2), 150–164. https://doi.org/10.1016/S0022-5371(82)90521-7

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

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

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 369–398. https://doi.org/10.1111/1467-9280.00151

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

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology, 38*(5), 816–827. https://doi.org/10.1017/S0048577201000610

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision, 5*(1), 8. https://doi.org/10.1167/5.1.8

Walenchok, S. C., Hout, M. C., & Goldinger, S. D. (2016). Implicit object naming in visual search: Evidence from phonological competition. *Attention, Perception, and Psychophysics, 78*(8), 2633–2654. https://doi.org/10.3758/s13414-016-1184-6

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America, 104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44*(12), 1411–1426. https://doi.org/10.1016/j.visres.2003.11.024

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin & Review, 11*(2), 269–274. https://doi.org/10.3758/BF03196569

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science, 12*(3), 219–224. https://doi.org/10.1111/1467-9280.00339

Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2011). Function follows form: Activation of shape & function features during object identification. *Journal of Experimental Psychology: General, 140*(3), 348–363. https://doi.org/10.1037/a0022840

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 32*(1), 1–14. https://doi.org/10.1037/0278-7393.32.1.1

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science, 11*(2), 125–131. https://doi.org/10.1111/1467-9280.00227