AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001019

"Where Are the . . . Fixations?": Grammatical Number Cues Guide Anticipatory Fixations to Upcoming Referents and Reduce Lexical Competition

Violet A. Brown[1], Neal P. Fox[2], and Julia F. Strand[3]

[1] Department of Psychological & Brain Sciences, Washington University in St. Louis
[2] Department of Science, Crystal Springs Uplands School
[3] Department of Psychology, Carleton College

OPEN DATA OPEN MATERIALS PREREGISTERED+ TC

Listeners make use of contextual cues during continuous speech processing that help overcome the limitations of the acoustic input. These semantic, grammatical, and pragmatic cues facilitate prediction of upcoming words and/or reduce the lexical search space by inhibiting activation of contextually inappropriate words that share phonological information with the target. The current study used the visual world paradigm to assess whether and how listeners use contextual cues about grammatical number during sentence processing by presenting target words in carrier phrases that were grammatically unconstraining ("Click on the . . .") or grammatically constraining ("Where is/are the . . ."). Prior to the onset of the target word, listeners were already more likely to fixate on plural objects in the "Where are the . . ." context than the "Where is the . . ." context, indicating that they used the construction of the verb to anticipate the referent. Further, participants showed less interference from cohort competitors when the sentence frame made them contextually inappropriate, but still fixated on those words more than on phonologically unrelated distractor words. These results suggest that listeners rapidly and flexibly make use of contextual cues about grammatical number while maintaining sensitivity to the bottom-up input.

Keywords: contextual cues, sentence processing, spoken word recognition, visual world paradigm

Supplemental materials: https://doi.org/10.1037/xlm0001019.supp

Recognizing spoken language requires overcoming ambiguity at multiple levels. On a sensory level, the presence of background noise often obscures the speech signal such that the source of particular frequency bands becomes ambiguous. However, ambiguity in the speech signal is not limited to adverse listening situations—even in pristine listening conditions there are one-to-many mappings between acoustic features and phonemic categorization. Further, semantic (e.g., *to, two,* and *too*) and syntactic (e.g., "I saw the man with the telescope") ambiguities add even more complexity to the process of speech recognition. Though some of these challenges exist for recognizing isolated syllables and words, the speech that we encounter most frequently—continuous streams of spoken words in sentences—contains ambiguity at all of these levels.

Despite multiple sources of ambiguity, people are typically able to parse continuous speech rapidly and efficiently. What strategies do listeners use to resolve these ambiguities during online sentence processing? A broad consensus is that listeners integrate bottom-up and top-down cues present in the transient sensory input and interpret them online to efficiently guide their behavior (see, e.g., Magnuson, 2019). Listeners can make use of semantic, pragmatic, and grammatical information from early in an utterance to form predictions about upcoming referents (Kukona et al., 2014) or reduce activation of contextually inconsistent lexical candidates

Violet A. Brown https://orcid.org/0000-0001-5310-6499
Neal P. Fox https://orcid.org/0000-0003-0298-3664
Julia F. Strand https://orcid.org/0000-0001-5950-0139

This work was supported by Carleton College and the National Science Foundation through a graduate research fellowship awarded to Violet A. Brown (DGE-1745038) and by the National Institutes of Health (NIDCD) through a National Research Service Award awarded to Neal P. Fox (F32-DC015966). We are grateful to members of the Carleton Perception Lab for data collection and to Naseem Dillman-Hasso, Hunter Brown, and ZhaoBin Li for technical assistance and helpful conversations.

The data are available at https://osf.io/4ztd9/
The experiment materials are available at https://osf.io/4ztd9/
The preregistered design is accessible at https://osf.io/zgu9d

Correspondence concerning this article should be addressed to Violet A. Brown, Department of Psychological & Brain Sciences, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, United States. Email: violet.brown@wustl.edu

(Dahan & Tanenhaus, 2004). For instance, given the sentence "The boy ate the . . ." listeners might show preference for edible objects (Kukona et al., 2014; Weber & Crocker, 2012), or in the sentence "They began to . . ." they may show preference for verbs (Strand et al., 2017). Importantly, these contextual cues need not be fully restrictive; listeners remain sensitive to the bottom-up acoustic-phonetic input even when it is semantically (Kukona et al., 2014) or grammatically (Strand et al., 2017) inconsistent with the preceding context. Thus, bottom-up processing appears to proceed incrementally, even in the face of context that biases interpretation against a given word.

A common method for assessing the incremental nature of online speech processing is the visual world paradigm (VWP), in which participants' eye movements are monitored as they listen to sentences while viewing a display of images (often a 3×3 grid with images in each of the four corners), one of which is referenced in the sentence (see Magnuson, 2019 for a review). In this paradigm, eye fixations are an index of lexical activation; that is, where a participant is looking is assumed to indicate which words are active in their mental lexicon (Tanenhaus et al., 1995). The VWP can detect graded activation for multiple lexical items, so it is often used to assess the dynamics of lexical competition—the idea that multiple items tend to be activated simultaneously in the listener's mental lexicon and compete for recognition (Luce et al., 2000; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994). In a classic study, Allopenna et al. (1998) demonstrated that cohort competitors sharing an onset with the target word (e.g., *beetle* for the target word *beaker*) provided more competition than rhyme competitors (e.g., *speaker*), but rhyme competitors were still activated to a greater extent than distractor items (e.g., *carriage*), despite the fact that the onset phonemes of the rhyme competitor were acoustically inconsistent with the input. A particular strength of the VWP is that it provides a sensitive measure of the activation of multiple lexical candidates before, during, and after presentation of a target word, thereby providing information about the time course by which listeners incorporate contextual information.

Anticipating Upcoming Utterances

Given that spoken language unfolds over time, listeners may be able to make use of context to predict or anticipate upcoming words such that lexical candidates that are contextually appropriate demonstrate increased activation before presentation of the target word. In research using the VWP, this is assessed by presenting contextually constraining sentences and assessing whether participants tend to preferentially fixate on contextually congruent objects prior to the onset of the target word (see Kukona et al., 2011). For example, when presented with a display containing images of edible and nonedible objects, participants fixate on edible objects earlier when the sentence contains the word *eat* (e.g., "The boy will eat the cake") than when it contains the less predictive word *move* (Altmann & Kamide, 1999). Further, listeners are sensitive not just to *selectional restrictions* (i.e., the noun is semantically restricted to be an argument of the verb, as was the case in Altmann & Kamide, 1999), but to more general semantic properties of the agent of the sentence. For example, Kamide and colleagues (2003) showed more fixations to a picture of a motorbike when participants heard the phrase "the man will ride the . . ." than when they heard the phrase "the girl will ride the . . ." Although both motorbikes and carousels can be ridden, participants anticipated the referent based on more general properties of the agent of the sentence. Further, there is some evidence that strongly constraining semantic context can lead to anticipatory effects not just for upcoming words, but for cohort competitors of those words. For example, when presented with the phrase "In order to have a closer look, the dentist asked the man to open his . . ." participants fixated on a phonological competitor ("mouse") of the target word ("mouth") more than on unrelated distractors (Ito et al., 2018; Kukona, 2020).

In addition to semantic context, listeners may also use situational and pragmatic context to predict upcoming referents (Barr, 2008b; Chambers et al., 2004; Hanna & Tanenhaus, 2004; Magnuson et al., 2008). For instance, Hanna and Tanenhaus (2004) demonstrated that when interacting with a confederate whose hands were full, listeners expanded their lexical search space to include objects close to the confederate, even if an identical object was closer to the participant. Similarly, another study showed that participants were more likely to fixate on objects that they knew were visible to the speaker (in "common ground"), indicating that they anticipated that the speaker was more likely to refer to those objects (Barr, 2008b).

There is less VWP work on whether listeners use grammatical cues to predict upcoming words. The majority of research on grammatical context using the VWP has focused on the dynamics of lexical competition (which unfold after the onset of the target word), rather than on anticipation specifically. However, visual inspection of the data from those papers suggests that grammatical gender (Dahan et al., 2000) and information about part of speech (Strand et al., 2017) do not appear to lead to preferential fixations to items that are congruent with the preceding context (though see Lew-Williams & Fernald, 2007, 2010 for evidence that Spanish-speaking children and adults are faster to orient to words that are consistent with a gender-marked determiner). For example, prior to target word onset, a grammatically masculine determiner did not lead to preferential fixations to grammatically masculine nouns in French speaking participants (Dahan et al., 2000).

Some work, however, has suggested that grammatical cues about the number of objects that are likely to be referenced can guide listeners to anticipate upcoming words (Kouider et al., 2006). In English (like many languages), speakers are typically required to specify whether a referent is singular or plural, and the accompanying verb must be conjugated accordingly. For example, a speaker who is referencing plural objects would say "There *are* the . . ." rather than "There *is* the . . . ." Thus, the conjugated form of the verb is an informative cue about an upcoming referent that listeners might use to predict words before they begin. Some previous research has indicated that both adults and children use information about grammatical number from subject-verb agreement to anticipate upcoming words. For example, verb morphology in combination with lexical quantifiers (e.g., "Look, there are some . . ." vs. "Look, there is a . . .") can guide anticipatory fixations to numerically congruent objects in 24-month-old children (Kouider et al., 2006), and verb morphology alone (e.g., "Where are the . . ." vs. "Where is the . . .") is sufficient to guide anticipatory fixations to numerically consistent objects in 2.5-year-old children and adults (Lukyanenko & Fisher, 2016). However, using similar paradigms, other studies have not found evidence for anticipatory eye movements driven by grammatical number. For example, when French toddlers heard "look, the$_{plural}$ [object]" they did not fixate more on plural objects than singular ones prior to the onset of the target word (Robertson et al., 2012). Similarly, adults hearing "Where *are* the [objects]" did not preferentially look toward plural objects more than singular objects prior to target word onset (Riordan et al., 2015).

It is not entirely clear why some research has shown anticipatory effects for grammatical context and some has not. One possible explanation for the discrepant findings is that for most of these studies (Dahan et al., 2000; Robertson et al., 2012; Strand et al., 2017), the grammatical cues occurred immediately before noun onset, so there may not have been enough time for listeners to use the grammatical cues predictively. In the case of the study conducted by Riordan and colleagues (2015), design and analytical choices may have precluded the effect from emerging. For example, that study used a small number of stimuli that appeared multiple times across the conditions, and the preview window (in which participants viewed the VWP images prior to hearing the instructions) was quite long, which could have encouraged participants to approach the task differently than they might in everyday listening conditions. Thus, this experiment will address these methodological issues to assess whether information about grammatical context—specifically grammatical number—can guide anticipatory fixations to upcoming utterances.

Integrating Context With Bottom-Up Input

In addition to using context to anticipate upcoming referents, listeners may use contextual cues to facilitate integration of the word into the sentence context after bottom-up input from the target word has begun to unfold (Forster, 1989; Marslen-Wilson, 1989). That is, contextual information may constrain how the incoming speech input is interpreted (see Federmeier, 2007; Lukyanenko & Fisher, 2016 for more on the distinction between using context to predict vs. integrating it with the incoming input to constrain lexical activation). A substantial body of work utilizing a wide range of methodologies—including event-related potential research (Friederici et al., 1996), phoneme classification tasks (Fox & Blumstein, 2016), and lexical decision tasks (Colé & Segui, 1994)—has demonstrated that listeners rapidly integrate some types of grammatical cues with the bottom-up input.

The distinction between anticipation (also referred to as prediction) and integration is subtle but important for understanding the mechanism by which listeners make use of context. Whereas the effects of prediction are detectable prior to word onset, contextual constraints on activation of lexical candidates are only apparent after listeners begin to process the acoustic-phonetic input from the target word, and should therefore only be discernible after the onset of the target word. Note, however, that activation at different time points need not imply that different underlying processes are responsible for anticipatory and integration effects—it may well be that the anticipation and integration describe the same phenomenon at different time points. Indeed, prior work has demonstrated that lexical activation fluctuates as the sentence unfolds such that contextual effects may or may not be observed in both time windows; listeners appear to make use of some forms of context for prediction but not integration (Barr, 2008b), whereas others are used for integration but not prediction (Dahan et al., 2000). In other words, different types of context (e.g., semantic, grammatical) may exert influence on lexical activation over different time courses. Regardless of whether the same process underlies anticipation and integration, in order to understand how listeners are making use of contextual cues, it is necessary to examine both time windows.

In VWP research, contextual constraints on integration with the bottom-up input manifest as a reduction in fixations to contextually inappropriate cohort competitors after the target word has begun to unfold (Altmann & Kamide, 1999; Dahan & Tanenhaus, 2004; Kukona et al., 2014; Weber & Crocker, 2012). For instance, when presented after an unconstraining context like "Sam chose the . . . *button*," participants fixated on a cohort competitor like *butter* more than a phonologically unrelated distractor, whereas when the target word was presented in a semantically constraining sentence like "Sam fastened the . . . *button*," fixations to the contextually inappropriate cohort competitor *butter* were reduced (Brock & Nation, 2014). These effects have also been demonstrated with some forms of grammatical context—listeners are less likely to fixate on cohort competitors of the target word when they are inconsistent with the grammatical context (Dahan et al., 2000; Strand et al., 2017). However, no research to date has assessed whether cues about grammatical number modulate lexical competition. Therefore, a second aim of the current research is to assess whether and how information about grammatical number is integrated with the bottom-up input.

Finally, assuming we find evidence that grammatical number affects processing of the target word as the speech unfolds, a subaim is to assess whether the context was sufficiently strong to reduce activation of inconsistent cohort competitors to the level of distractors. Previous research has shown that although contextual cues can reduce activation of contextually inappropriate cohort competitors, in some situations listeners remain sensitive to the bottom-up input (e.g., when there is sufficient acoustic-phonetic support for the competitor or the competitor is a high frequency word; Dahan & Tanenhaus, 2004; Strand et al., 2017; Weber & Crocker, 2012). These findings are consistent with the account that bottom-up and top-down inputs are continuously integrated during speech processing (see, e.g., Dahan & Tanenhaus, 2004; Magnuson et al., 2008). That is, even in contexts such as "They began to . . ." that strongly bias the final word to be a verb (e.g., *run*), listeners do not fully disregard the possibility that the final word is grammatically inappropriate (e.g., *rug*). Thus, this experiment will assess the strength of the grammatical number cue by testing whether residual activation for cohort competitors persists, even when the preceding grammatical number context renders them unlikely.

Current Study

The goal of the current study was to evaluate when and how listeners make use of contextual cues about grammatical number. We assessed whether grammatical number cues can (a) guide anticipatory fixations prior to target onset and (b) reduce or even completely extinguish competitor preference after the onset of the target word when the numerically constraining sentence context renders the competitor unlikely. Consistent with previous research (Dahan et al., 2000; Dahan & Tanenhaus, 2004; Strand et al., 2017; Weber & Crocker, 2012), participants viewed grids with four images as they listened to sentences, and we monitored their eye movements as the sentence unfolded. Each grid contained a target (e.g., *beagles*), a cohort competitor with a different number (e.g., *beaker*), a numerical lure that was phonologically unrelated to the target (e.g., *archers*), and a distractor that was both phonologically and numerically distinct from the target (e.g., *archway*). Plurality (singular vs. plural) rather than object type (e.g., target vs. competitor) was of interest in the anticipatory analysis because the anticipatory analysis window ends before any phonological information about the target word can be discerned. In contrast, the cohort analysis focused on activation of the competitor (e.g., *beaker*) relative to the distractor (e.g., *archway*).

For the anticipatory analysis, we hypothesized that participants would show more fixations to plural (e.g., *beagles* and *archers*) relative to singular objects (e.g., *beaker* and *archway*) prior to target onset when the preceding context biased the final word to be a plural rather than singular noun. For the cohort analysis, we hypothesized that there would be fewer fixations to the numerically inappropriate cohort competitor relative to the distractor in the constraining than the unconstraining context. However, we expected that in constraining contexts, there would still be greater activation of the cohort competitor (e.g., *beaker*) than the distractor (e.g., *archway*), despite poor contextual fit. These results would suggest that grammatical number guides anticipatory fixations and that it reduces but does not eliminate activation of the numerically inappropriate cohort competitor.

Method

Stimuli, data, and code for all analyses are provided at https://osf.io/4ztd9/, and preregistration documentation is available at https://osf.io/zgu9d. In line with Simmons and colleagues' (2012) 21-word solution—which is intended to increase transparency about research practices—we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Participants

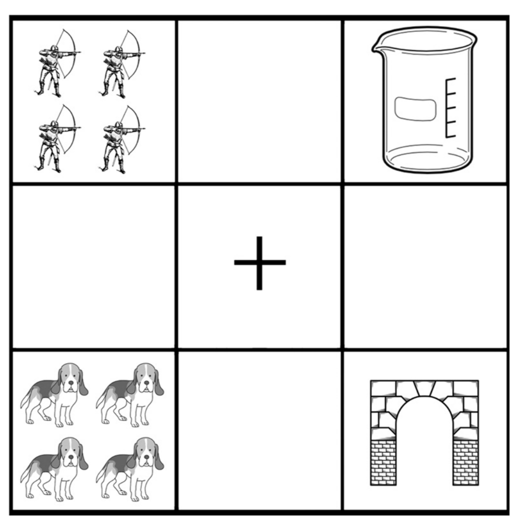
Data from 135 native English speakers with normal hearing and normal or corrected-to-normal vision were included in analyses for this experiment. In addition to these included participants, and in accordance with preregistered exclusion criteria, four participants were excluded from analyses because the proportion of frames with usable data across the combined anticipatory and cohort windows was more than 3 standard deviations below the mean proportion of frames with usable data (i.e., these participants had too much missing data). We preregistered that we would exclude participants whose response (click) accuracy was lower than 85%, but no participants met this exclusion criterion. We had preregistered a final sample size of 136 to be consistent with our prior work (Strand et al., 2017), but excluding four participants resulted in a final sample size of 135, and campus closures precluded us from collecting data from an additional participant. Carleton College's Institutional Review Board approved the procedures. The study took approximately 60 minutes to complete, and participants were compensated $11 for their time.

Materials

To select the 288 words we used in this study, we consulted a phonologically transcribed dictionary (Balota et al., 2007) and identified pairs of two-syllable nouns that overlapped phonologically at onset (e.g., *beagle* and *beaker*). At least one member of each pair had regular English plural morphology (e.g., plural of *beagle* is *beagles*). One word from each pair was randomly assigned to be plural—except in cases where one of the words in the pair did not have regular English plural morphology, in which case the word with regular English plural morphology was assigned to be plural—yielding 144 word pairs that were phonological onset competitors (i.e., cohort competitors) but differed in their plurality (e.g., *beagles* and *beaker*). The sets of words that made up the singular and plural members of cohort competitor pairs were matched on average phoneme length, phonological neighborhood density (taken from Balota et al., 2007), and lexical frequency (log frequencies taken from Brysbaert & New, 2009), with all $ps > .64$ using two-tailed paired *t*-tests. These comparisons were made on the singular forms of the words (e.g., singular *beagle* and *beaker*), as plural forms typically have lower frequency, longer length, and fewer neighbors than singular forms. Note that another design choice would have been to include multiple forms of each target in each grid (once in a plural form and once in a singular form) and assess activation of both. A benefit of the target and competitor pairing we used here is that it enables us to compare our pattern results with prior work on cohort activation.

Each cohort competitor pair was randomly matched to another cohort competitor pair, yielding 72 sets of items consisting of two cohort pairs (e.g., *beagles*, *beaker*, *archers*, *archway*; see Appendix for the full stimulus list) that were used to create the VWP displays (grids) that participants saw (see Figure 1 for an example grid). In this way, when one image was designated as the target word (e.g., *beagles*), the other three images in the grid could be classified as the cohort competitor (*beaker*; matching in onset but not number), the numerical lure (*archers*; matching in number but not onset), and the

Figure 1
Example VWP Grid



Note. In the example described in the text, beagles is the target, beaker is the cohort competitor, archers is the numerical lure, and archway is the distractor. VWP = visual world paradigm.

distractor (*archway*; matching in neither onset nor number). Manual review ensured that the randomly matched cohort competitor pairs were not phonologically related and that the sets of four did not contain semantically related or visually similar objects. Because the item sets included two pairs of cohort competitors (e.g., *beagles/beaker* and *archers/archway*), the numerical lure and distractor were always cohort competitors of one another. This property of the experimental design prevents participants from assuming (prior to target onset) that two objects that share an onset are more likely to be targets than the other objects in the VWP grid, a strategic issue that prior work has avoided by instead including additional filler trials (Dahan et al., 2000; Dahan & Tanenhaus, 2004).

Speech Stimuli

Speech stimuli were recorded using a Blue Yeti microphone and were matched on RMS amplitude using Adobe Audition (Version 10.1.1.11). Stimuli were produced by a female native speaker of American English without a discernible regional accent. All 288 critical words and eight additional practice stimuli (corresponding to two practice grids) were produced in isolation from the three carrier phrases, which included two numerically constraining carrier phrases ("Where is the . . . [*singular target*]?" and "Where are the . . . [*plural target*]?") and one unconstraining carrier phrase ("Click on the . . . [*singular/plural target*]"). Unconstraining trials contained no information about target identity or number until the onset of the target word, whereas numerically constraining trials contained a copula (*is* or *are*) that indicated whether the target would be singular or plural prior to target onset. Importantly, on numerically constraining trials, the target number was never incongruent with the information indicated by the carrier phrase, so the copula provided listeners with a reliable grammatical number cue.

Carrier phrases were recorded separately from the target words to ensure that no phonological information about the target was available before target word onset (Salverda et al., 2014). The three carrier phrases were digitally spliced such that both the constraining and unconstraining carrier phrases used the same acoustic token of *where*, all three frames ended in the same acoustic token of *the*, and for all three frames the acoustic onsets of *is*, *are*, and *on*, as well as the onset of *the*, were time-aligned (see Figure 2). Sentences were constructed by selecting the carrier phrase (1,148 ms), inserting 50 ms of silence after it, then adding each target word. Thus, the same token was used as the target word in the constrained and unconstrained trials. The average length of the target words was 788 ms, and the average divergence point—the point at which a given word (e.g., *beagles*) became disambiguated from its competitor (e.g., *beaker*)—was 315 ms. To assess divergence points, two coders listened to each audio file in Audacity (Version 2.3.2) and assessed the point at which the phonologically related pairs begin to differ from one another. These values (for all words and both coders) were averaged. In cases of large differences in estimated divergence points between the coders, a different set of coders independently reassessed the divergence point and those values replaced the original values. These divergence points were used to determine the end of the time window for the cohort analysis (see Results).

Images

Visual stimuli consisted of 288 freely available black-and-white drawings that were used to create the 72 grids. Drawings were selected from existing databases (Cycowicz et al., 1997; Duñabeitia et al., 2017) and supplemented from online sources. VWP grids were created with Adobe Photoshop 2017 (see Figure 1 for an example grid). Each grid contained two singular objects (e.g., a drawing of a *beaker* and a drawing of an *archway*) and two sets of plural objects. Each set of plural objects was represented by a square array of four smaller copies of the drawing of the singular target word. When presented to participants on a screen, each square of the grids measured approximately 90 mm on each side, with singular and plural images filling the same total visual area. Each of the 72 grids was seen exactly once by every participant, so a given image always appeared in the same location on the grid, but the object type (target, cohort competitor, numerical lure, or distractor) of each of the four images varied across participants. For a given participant, each object type appeared equally often in each position in the grid (e.g., the target appeared in each of the four positions exactly 18 times for all participants) and the arrangement (relative positioning) of all four item types was balanced across grids.

Design

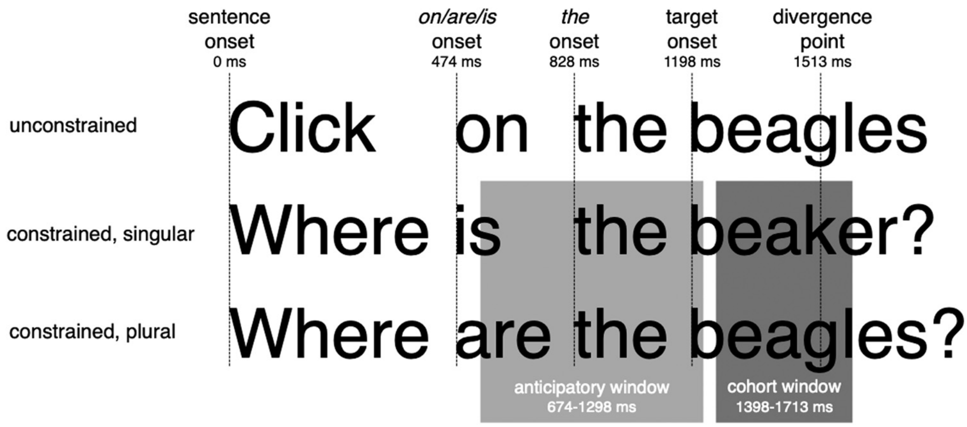
Participants completed 72 trials in which they viewed a VWP grid while they listened to a sentence that ended in a singular or plural target word (36 trials of each). Plural and singular trials were equally divided across numerically unconstraining ("Click on the . . . [*singular/plural target*]") and numerically constraining ("Where is the . . . [*singular target*]?" or "Where are the . . . [*plural target*]?") carrier phrases. Trials across all conditions (singular/plural targets following constraining/unconstraining contexts) were presented in a randomized, intermixed order for each participant.

Each participant saw every grid exactly once, but with one of eight different aurally-presented sentences (see Table 1). Each of the four images in the grid served as the target for different participant groups, and served as the target in both a numerically unconstraining (e.g., "Click on the . . .") and a numerically constraining ("Where is/are the . . .") sentence context. Each grid appeared in all eight groups, corresponding to different verbal instructions, and each participant was assigned to one of the eight groups. This design ensured that participants could not employ strategic processing based on the number and composition of the stimuli. Although any given participant saw each grid only once (with only one of the images in the grid serving as the target), across participants, every object in every grid served as a target in both a constraining and an unconstraining carrier phrase.

Procedure

Prior to beginning the main experiment, participants were trained on the words assigned to each image to ensure that, for example, everyone knew the word was *beagle* and not *dog* (see Strand et al., 2017 for a similar procedure). Participants were first sequentially presented with the pictures that would be used in the main experiment along with a printed label of what the image referred to in a pseudorandomized order. After

Figure 2
Anticipatory and Cohort Windows For Each Trial Type



Note. Schematic of the timing of the three carrier phrases. The two shaded regions represent the windows of time used in the analysis of the constrained trials (see Results section for more information).

this initial self-paced familiarization phase, participants were sequentially presented with each picture in a different pseudorandomized order and asked to type its label into a text box. If participants responded correctly, that image was removed from the training set so they did not see it again. If they responded incorrectly, the appropriate name appeared for two seconds on the screen, and the image was presented again at the end of training. This phase continued until the participant had successfully learned all the picture/label pairings. Throughout this training paradigm, participants only saw a single drawing for each image paired with the singular form of the noun, regardless of whether the object would appear in its plural or singular form during the subsequent eye-tracking experiment. The likelihood that an image was repeated during training varied across images and across participants, with the average image being repeated .4 times (see accompanying R script for details).

For the main experimental task, participants sat a comfortable distance from a 24-in. PC monitor connected to a Dell Precision T5600 computer. Fixation information was collected at 60 Hz with a Tobii X2-60 eye-tracker. Stimulus presentation, data collection, and calibration were controlled via SuperLab (version X6). Speech stimuli were presented via Sennheiser HD 280 Pro headphones at a comfortable listening level. Participants were told "You will see objects on the screen and hear instructions about them. Sometimes each image will appear on its own and sometimes there will be multiple copies of it. Listen closely and click on the object or objects that the speaker refers to." Fixations anywhere within the borders of each image's square in the grid were counted as fixations to that image. Every time a participant clicked on an image, a custom script returned the cursor to the center of the screen to encourage fixations to the central cross at the beginning of every trial. Each trial began with a blank VWP grid presented for 1,000 ms, at which point the images were added to the grid for 500 ms prior to the onset of the audio file. The next trial was initiated when the participant clicked on an image. Participants completed two practice trials (one with a plural target and one with a singular target, both presented in a constraining context) prior to the start of the experiment.

Results

According to preregistered criteria, individual trials were only excluded when the participant clicked on the incorrect image (fewer than 1% of trials), which may have resulted from inattention to the task. However, 94.20% of incorrect clicks were to competitors—compared to just 4.27% to numerical lures and 1.33% to

Table 1
Eight Sets of Instructions for the Example Grid Shown

Group	Numerical constraint	Carrier phrase	Target number	Target	Cohort competitor	Numerical lure	Distractor
1	Unconstrained	Click on the ____.	plural	beagles	beaker	archers	archway
2	Constrained	Where are the ____?					
3	Unconstrained	Click on the ____.	singular	beaker	beagles	archway	archers
4	Constrained	Where is the ____?					
5	Unconstrained	Click on the ____.	plural	archers	archway	beagles	beaker
6	Constrained	Where are the ____?					
7	Unconstrained	Click on the ____.	singular	archway	archers	beaker	beagles
8	Constrained	Where is the ____?					

Note. All participants saw every grid with one of eight different aurally-presented sentences (unconstraining versus constraining sentence contexts with each of the four images as the target).

distractors—suggesting that errors did not result from random clicking. Note that there were no incorrect clicks that fell outside the grid because the trial only ended when participants clicked on one of the four objects.

In the unconstrained trials—in which participants had to rely solely on bottom-up information about the speech to identify the target word—the data showed patterns typical of VWP research on lexical competition (see Figure 3). Fixation proportions to each of the four objects were evenly distributed at approximately 25% each prior to the onset of the target word. As the target word unfolded, fixations to target words steadily rose while fixations to objects that did not share phonological overlap (i.e., the numerical lure and distractor) fell. Fixations to the cohort competitor rose briefly, then fell following disambiguation from the target word.

Two main analyses were conducted in which we tested for an effect of grammatical number information in the sentence context on (a) anticipatory fixations and (b) cohort competition. Both the anticipatory and lexical competition analyses were conducted using linear mixed effects models via the *lme4* package (Version 1.1-23) in R (Version 4.0.2; Bates et al., 2014). Statistical significance was assessed via likelihood ratio tests comparing nested models differing only in the effect of interest. Where appropriate, *p*-values were obtained from model summary outputs via the *lmerTest* package (Kuznetsova et al., 2017).

Analyses were conducted on data from two critical windows: an anticipatory window and a cohort window (see Figure 2). The anticipatory window was intended to represent the time that listeners have grammatical cues about plurality but no phonological information about the target. It is generally assumed that it takes listeners 200 ms to program and launch an eye movement (Fischer, 1992; Rayner et al., 1983); thus, the start of the analysis window was 200 ms after the onset of *is* or *are* and ended 100 ms after the target word onset. We opted to offset the end of the anticipatory window by 100 ms rather than 200 ms to ensure that no phonological information from the target was included. The cohort window was intended to approximate the time that includes phonological overlap between targets and competitors and therefore began 200 ms after the onset of the target word and ended 200 ms after the average divergence point of the target and competitor.

For each trial, we computed the number of frames with fixations on each image in the analysis window (classified by item type: target, cohort competitor, numerical lure, or distractor). Logit-transformed fixation rates with an offset of .5 were computed after collapsing across frames within the analysis window. Fixations outside the grid and frames with missing data were included when determining the total number of frames in a given trial for use in the empirical logit calculation (see Barr, 2008a). We had preregistered that we would collapse across either participants or items (following the recommendations of Barr, 2008a) to account for the fact that frames within a particular trial for a particular participant are correlated (i.e., a participant's fixation location in one frame is highly predictive of their fixation location in the next frame). However, given that we did not plan to analyze the time course of fixations within each analysis window, we realized it would be more appropriate to aggregate the data across frames in each window, obviating the need for collapsing across either participants or items to reduce dependencies in the data. This decision did not change the conclusions drawn from the analyses (see online supplemental materials). For each analysis (anticipatory and cohort), the dependent variable was a difference in logits, representing a fixation preference score (see each analysis below for more details). The independent variables were (a) the plurality of the sentence (singular vs. plural) in the anticipatory analysis and (b) constraint (unconstrained vs. constrained) in the cohort analysis. Deviation coding was used for plurality (singular = $-.5$; plural = .5), and treatment coding was used for constraint (unconstrained = 0; constrained = 1).

For all analyses, participants and items (words) were included as random effects.[1] All analyses included random intercepts for both grouping factors, and we attempted to model random slopes when they were justified by the design (Barr et al., 2013). Unless otherwise noted, results from the maximal model are reported. In cases of nonconvergence or singularity, we altered control parameters (e.g., changed the optimizer) and set correlations among random effects to zero, and when that failed, we selectively removed random slopes (see accompanying R script). Before removing any random slopes, we conducted likelihood ratio tests to ensure that removal of the random slope did not significantly reduce model fit.

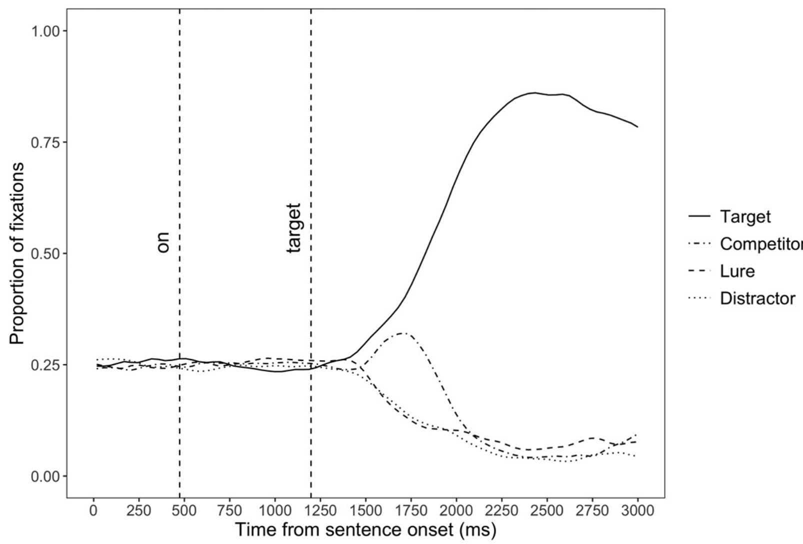
Given the large number of ways in which eye-tracking data can be analyzed, both the anticipatory and cohort analyses included several exploratory sensitivity analyses in which different analytical techniques were adopted to assess the robustness of the effects we observed to a variety of analytical techniques. In general, sensitivity analyses involve changing the analytical technique, exclusion criteria, model specification, or other inputs to the analysis, and assessing the impact of these decisions on the outcome of the analysis (see, e.g., Thabane et al., 2013). These analyses are especially useful when there are multiple ways of testing a hypothesis in a given data set. Therefore, these analyses help to ensure that we did not simply "get lucky" with the particular analysis we chose. These additional analyses rendered patterns of results that are consistent with those reported here and are discussed in more detail in the online supplemental materials and in the additional R script at https://osf.io/4ztd9/.

Anticipatory Analysis

The anticipatory analysis was conducted exclusively on constrained trials because the purpose was to assess whether participants make use of the copula (*is* or *are*) in the grammatically constraining context to guide anticipatory eye movements prior to target onset. The dependent variable in this analysis was the difference in the logit-transformed fixations to plural objects minus those to singular objects during the anticipatory window ("plurality preference"), and the independent variable was the plurality of the sentence context (singular or plural). Object types that were matched in number but not onset (e.g., *beagles* and *archers*; *beaker* and *archway*) were not differentiated in this analysis because listeners could have no knowledge during the anticipatory window about the phonological properties of the target's onset. Larger values of the dependent variable indicate that prior to hearing any phonological information about the target, participants were more likely to fixate on plural objects than singular objects. Note that although there are many other reasonable outcome

[1] We had preregistered that we would include words within grids as nested random effects but during analysis realized that that was not appropriate given the design of the experiment (see accompanying code for more details).

Figure 3
Proportion of Fixations in Unconstrained Trials



Note. Fixation proportions to each of the four objects in the VWP for the unconstrained trials (i.e., "Click on the . . ."). The dashed vertical lines represent the onset of the word *on* in the carrier phrase (left) and the onset of the target word (right).

measures that could be used to assess whether listeners anticipated the upcoming referent based on the preceding context (e.g., congruity preference, calculated as the difference between logit-transformed fixations to objects that matched the number indicated by the sentence context and objects that mismatched the number), the sensitivity analyses in the accompanying R script indicate that none of these outcomes produced a different pattern of results.

We hypothesized a significant effect of plurality such that there would be greater plurality preference in the anticipatory window when the numerical context was consistent with the target being plural ("Where are the . . .") relative to singular ("Where is the . . ."). Consistent with our hypothesis, a likelihood ratio test of the model including the plurality of the sentence as a fixed effect provided a better fit for the data than a model without it, indicating that cues to grammatical number affected plurality preference ($\chi^2_1 = 34.44$, $p < .001$). Examination of the summary output for the full model indicated that participants tended to fixate on plural relative to singular objects to a greater extent when the context indicated that the target word would be plural ($B = .75$, $SE = .12$; $t = 6.21$, $p < .001$), see Figure 4 and Table 2. The mean plurality preference in the plural context (difference in logits) was .88 ($SD = 3.65$) whereas the mean plurality preference in the singular context (difference in logits) was .13 ($SD = 3.63$). Though not central to our question of interest, the intercept term was significant ($B = .50$, $SE = .07$, $t = 7.41$, $p < .001$), indicating that participants tended to fixate more on plural relative to singular objects overall, perhaps because the plural objects were more visually complex.

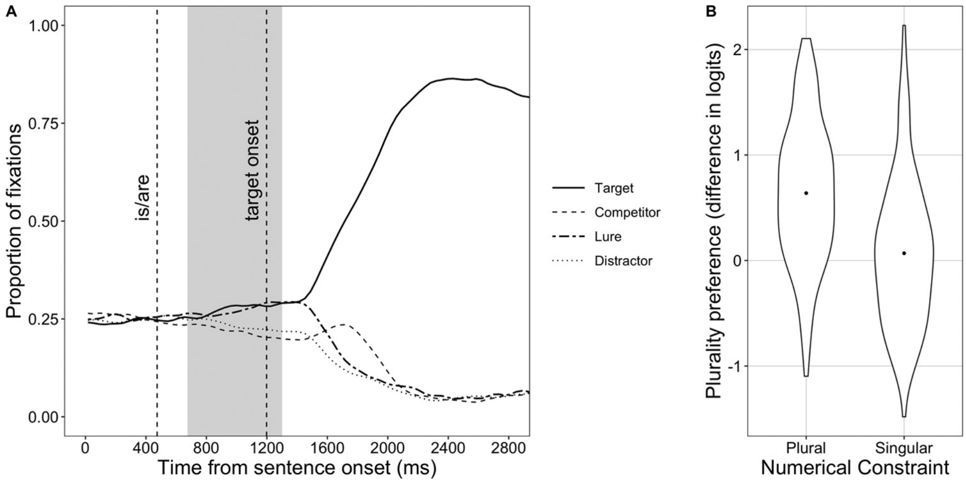
Cohort Analysis

In the next set of analyses, we assessed whether the presence of numerically constraining information in the carrier phrase could influence lexical competition during processing of the target word. The dependent variable for this analysis was the difference between logit-transformed fixations to the cohort competitor (e.g., *beaker* if the target word is *beagles*) and the distractor (e.g., *archway*), both of which were numerically incongruent with the target (Strand et al., 2017; Thothathiri & Snedeker, 2008), within the cohort window. Importantly, we did not include the numerical lure as a distractor in the cohort analysis because there could be residual fixations to the numerical lure following anticipatory fixations to this object in the numerically constraining contexts. If the numerical lure was treated as a distractor object, then early fixations to the numerical lure would tend to inflate the average rate of fixations to distractors, which would artificially deflate true competitor preference effects during the cohort window. The decision not to combine fixations to numerical lures and distractors was made prior to beginning the study (see the preregistration form for details) and was therefore not informed by our findings in the anticipatory analyses. Larger values for the difference between empirical logit-transformed fixation rates for cohort competitors and distractors indicate greater "competitor preference," suggesting that the cohort competitor's lexical representation has been activated as a result of phonological overlap with the onset of the target word. The final model included random intercepts for participants and items as well as by-participant random slopes for constraint, but did not include by-item random slopes because the maximal model encountered estimation issues.

Given the robust emergence of competitor preference effects throughout the literature on spoken word processing, we predicted that there would be a significant competitor preference in the unconstrained condition, as indicated by a significant and positive intercept (recall that the unconstrained condition is the reference level according to our dummy coding scheme). Consistent with previous research, results confirmed the presence of classic cohort

Figure 4

Proportion of Fixations and Plurality Preference for Anticipatory Analysis



Note. (A) Proportion of fixations to each of the four objects in the numerically constrained trials. The dashed vertical lines represent the onset of *is* or *are* (left) and the target word (right). The shaded region represents the anticipatory analysis window. (B) Violin plot showing by-participant plurality preference (difference in logits) for singular and plural targets. The dot represents the mean plurality preference and the shape represents the distribution of responses. Although plurality preference may be expected to be negative in the singular condition, the general trend to fixate on plural objects more than singular ones (perhaps due to image complexity) may be obscuring this effect. Note that although the data in the analysis were collapsed across frames within the anticipatory window but not across participants or items, the data in panel B were grouped by participants for visualization purposes.

competition effects—in the unconstrained condition, there were more looks to the cohort competitor than the distractor (B = .36, SE = .04, t = 8.41, $p < .001$).

If listeners use syntactic information from preceding context to guide lexical processing, then the competitor preference that is present in the unconstrained condition may be weakened in the constrained condition because the cohort competitor (*beaker*) does not represent a grammatically congruent continuation of the sentence ("Where are the *bea* . . ."). Thus, we also predicted that there would be a significant negative effect of numerical constraint on competitor preference, consistent with weaker competitor preference during the cohort window after a numerically constraining sentence context. Consistent with this prediction, the effect of constraint was significant (χ^2_1 = 8.25, p = .004) and the slope was negative (B = −.17, SE = .06, t = −2.91, p = .004), indicating that participants fixated less on the numerically inappropriate competitor in the constrained context than the unconstrained context. The mean competitor preference in the unconstrained condition (difference in logits) was .36 (SD = 2.71) whereas the mean competitor preference in the constrained condition (difference in logits) was .19 (SD = 2.56).[2]

Table 2

Proportion of Fixations to Each of the Object Types in the Anticipatory and Cohort Windows

Window	Condition	Object type	*M* (*SD*)
Anticipatory	Constrained	Target	0.27 (0.01)
		Lure	0.27 (0.01)
		Competitor	0.22 (0.01)
		Distractor	0.23 (0.01)
Cohort	Constrained	Target	0.38 (0.07)
		Competitor	0.22 (0.01)
		Distractor	0.17 (0.04)
	Unconstrained	Target	0.33 (0.04)
		Competitor	0.28 (0.03)
		Distractor	0.19 (0.03)

In addition to this main cohort analysis, we also conducted a preregistered analysis to test whether the numerical context was sufficiently constraining to reduce competitor fixations to the level of the distractor fixations. That is, the previous set of analyses indicated that numerical context reduced activation of numerically inconsistent words, but it did not assess whether this reduction was so extreme as to actually eliminate activation of the competitor (see Strand et al., 2017 for an analogous analysis). For this analysis, we recalculated the empirical logit fixation rates for the competitor and distractor using only constrained trials, and again defined "competitor preference" as the difference in these empirical logits. We then built a model with competitor preference as the outcome and a fixed effect for the intercept. If the competitor still shows some residual activation in the constraining context despite being incongruent with the preceding numerical context, this competitor preference would be indicated by a significant positive intercept. If, however, the numerical context was sufficiently constraining that fixations to the competitor were effectively

[2] See online supplemental materials for an exploratory analysis that did not find evidence for a relationship between competitor preference and the magnitude of anticipation.

eliminated, the intercept should not differ from 0 (i.e., there should be no preference to fixate on the competitor relative to the distractor).

Results of this analysis revealed that logit-transformed fixation rates to competitors exceeded those to distractors, despite both objects being numerically inconsistent with the preceding context, indicated by a significant positive intercept ($B = .19$, $SE = .04$, $t = 4.48$, $p < .001$). Taken together, the results of the cohort analysis indicate that numerical context reduced but did not eliminate activation of the numerically incongruent competitor; that is, even in the face of grammatical context that renders the competitor unlikely, listeners remain sensitive to the bottom-up acoustic and consider numerically incongruent but phonologically supported lexical items (see Figure 5 and Table 2).

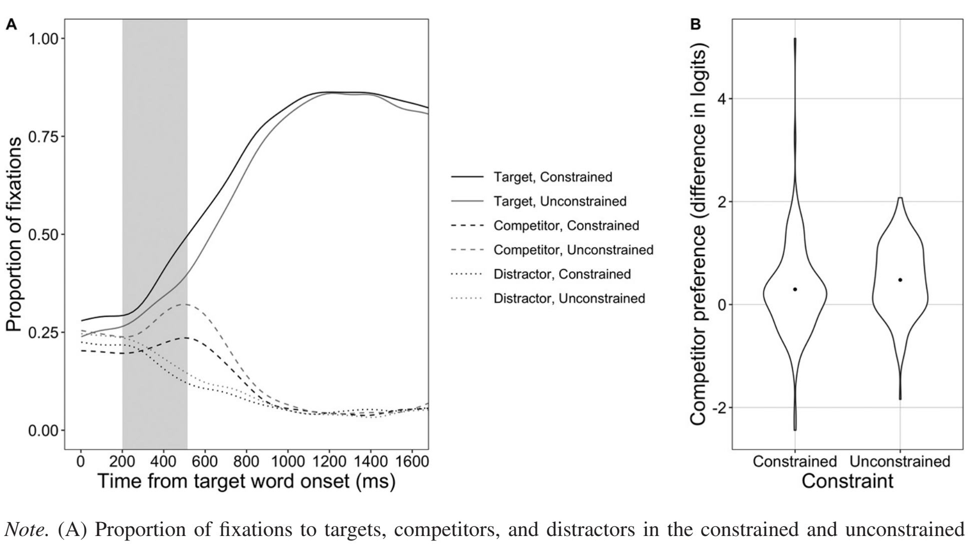
Discussion

In the face of ambiguous and rapidly unfolding acoustic input, listeners make efficient use of contextual cues. These cues may facilitate speech processing by allowing listeners to anticipate or predict words that are likely to occur later in the utterance and/or by inhibiting activation of contextually inappropriate lexical candidates that receive bottom-up acoustic-phonetic support by virtue of their phonological overlap with the target word. The current eye-tracking study assessed whether fixations to singular or plural nouns were affected by the conjugation of the preceding verb. Results showed that listeners made use of grammatical number cues for both anticipating upcoming words and integrating them with the preceding sentence context, but even in numerically constraining contexts, activation of contextually inappropriate cohort competitors persisted above the level of distractor objects. Thus, cues to grammatical number are immediately integrated with the bottom-up input but not so restrictively that listeners disregard lexical candidates that are incongruent with the context.

The first key finding of this experiment was that listeners tended to look at contextually congruent referents even before the onset of the target word. These anticipation effects have been demonstrated elsewhere for both semantic (Kamide et al., 2003) and grammatical (e.g., "Where are the good cookies?" Lukyanenko & Fisher, 2016) context. However, other studies have indicated that listeners appear not to use grammatical contextual cues predictively (Dahan et al., 2000). There are at least two explanations for these discrepancies. The first is that listeners use different forms of context differently. For instance, when hearing "The boy will eat . . ." the possible referents are limited to a relatively small set of edible objects, whereas for phrases such as "They began to . . ." the possible referents include all verbs (a much larger set of objects). It may be that anticipatory effects are stronger for forms of context that are more strongly constraining (i.e., those that limit the lexical search space to a relatively small set of candidates). The second possible explanation for discrepancies in anticipatory effects across various forms of context is that anticipation effects may only emerge in situations in which there is sufficient time between the contextual cues and the target word. For instance, in research on grammatical gender (Dahan et al., 2000) and grammatical number (Robertson et al., 2012) in French, the gender- or number-marked determiner often comes immediately before the target word, leaving little time for participants to initiate anticipatory eye movements to the target word. Thus, it may be that anticipatory effects can only be detected over a longer time window.

Figure 5

Proportion of Fixations and Competitor Preference for Cohort Analysis



Note. (A) Proportion of fixations to targets, competitors, and distractors in the constrained and unconstrained conditions. Fixations to numerical lures are omitted from this figure for simplicity. Though fixations to targets, like those to numerical lures, were not relevant for this analysis, they are included in the figure for reference. (B) Violin plot showing by-participant competitor preference (difference in logits) for constrained and unconstrained targets. The dot represents the mean competitor preference and the shape represents the distribution of responses. Note that although the data in the analysis were collapsed across frames within the cohort window but not across participants or items, the data in panel B were grouped by participants for visualization purposes.

Future research should attempt to distinguish between these possibilities by testing whether the contexts that have *not* shown patterns of anticipatory looks may indeed show those patterns when the speech is slowed or additional words are added between the constraining verb and the target word (see Lukyanenko & Fisher, 2016 for an example of this technique).

The second major finding of this study was that cohort competitors were fixated on less in contexts that made them unlikely. This suggests that listeners downgraded activation of words that were phonologically but not contextually appropriate. These results are consistent with the continuous integration view of speech processing (Dahan & Tanenhaus, 2004; Magnuson et al., 2008). According to this account, listeners utilize all available bottom-up and top-down cues when processing continuous speech, and immediately and continuously integrate information from these sources to guide perception. Support for this view has been found in the semantic realm for studies conducted in German (Weber & Crocker, 2012) and French (Dahan & Tanenhaus, 2004), in the grammatical realm to distinguish between nouns and verbs in English (Strand et al., 2017) and grammatical gender in French (Dahan et al., 2000), and in the pragmatic realm using an artificial lexicon (Magnuson et al., 2008), among others. In each of these studies, contextual cues were immediately integrated with the incoming acoustic-phonetic input to constrain activation of contextually inappropriate lexical candidates that share phonological information with the target.

The third key finding was that although grammatically constraining context reduced activation of contextually inappropriate cohort competitors, listeners remained sensitive to the bottom-up input and preferentially fixated on the competitors more than the phonologically unrelated distractors. Why might listeners consider lexical candidates that are clearly inconsistent with the preceding context? Although predictions derived from context are often correct (we would not be expected to use contextual cues otherwise), predictions can be defied, so it may be beneficial to incorporate contextual cues flexibly rather than rigidly. This is not the first study to show residual activation for contextually inappropriate cohort competitors (e.g., see Dahan & Tanenhaus, 2004; Strand et al., 2017; Weber & Crocker, 2012). However, in previous work, activation for the competitors was increased via an experimental manipulation—either by cross-splicing acoustic-phonetic information from the competitor into the target to increase bottom-up support for the competitor (Dahan & Tanenhaus, 2004; Strand et al., 2017) or by specifically selecting low frequency targets with high frequency competitors to bias listeners to fixate on competitors (Weber & Crocker, 2012). Thus, the current experiment demonstrates a form of context that can reduce but not eliminate activation of contextually inappropriate competitors without researchers deliberately increasing activation of the competitor relative to the target. Though methodological differences preclude making direct comparisons across studies, these results suggest that grammatical number cues may be less restrictive than some forms of context (e.g., semantic context), though they may still be more restrictive than others (e.g., pragmatic context; Barr, 2008b).

One explanation for the finding that grammatical number cues are used less restrictively than some other forms of context is that grammatical number cues may have relatively weak cue validity. In English, although "is" is typically associated with singular nouns and "are" is typically associated with plural nouns, this rule is violated frequently; for example, English has many instances of mass nouns (e.g., "Where is the luggage?"), pluralia tanta (e.g., "Where are the tongs?"), and noun phrases that may modify the interpretation of plural nouns (e.g., "Where is the herd of sheep?"; Riordan et al., 2015), so "is" and "are" are often not associated with singular and plural nouns, respectively. Given this weak cue validity, listeners may place less weight on grammatical number cues relative to other types of contextual cues. Future research should attempt to assess whether forms of context that are violated more frequently lead to less restrictive processing.

The results we report here are consistent with the claim that listeners make use of grammatical number cues to anticipate upcoming words and integrate those cues with the unfolding acoustic input. However, it is worth noting that the way speech was presented in this study (and indeed in all VWP studies) is quite far removed from natural listening situations (see Huettig et al., 2011 for a review). One criticism of the VWP is that viewing a display with a limited number of objects may lead to strategic processing, particularly if the sentence is spoken slowly. In this experiment—in which the visual displays preceded the speech—listeners may have deliberately generated predictions to a greater extent than is possible in naturalistic settings. Further, cues to grammatical number were never violated in this study, which may have increased their cue validity in this context. This is not to imply that anticipatory effects simply reflect artifacts of the VWP—evidence from ERP research (e.g., Friederici et al., 1996) and lexical decision tasks (e.g., Colé & Segui, 1994; Jakubowicz & Faussart, 1998) also suggest that listeners anticipate grammatically appropriate lexical candidates—but rather that effects observed in VWP studies may be stronger than what might be expected in the real world, in which response options are not limited and violations are common.

In addition to concerns about how the VWP may exacerbate effects of anticipation, it has been argued that viewing a display with a limited number of objects may lead to preactivation of the lexical candidates, increasing the magnitude of subtle phonological effects relative to how they operate in everyday conversations (Huettig et al., 2011; though see Tanenhaus et al., 2000). Familiarizing participants with the words in the training phase may further inflate bottom-up influences because those lexical items have been preactivated (Ito et al., 2018). The results reported here suggest that listeners make use of grammatical number cues in idealized situations, but future work should assess the generalizability of these findings to more naturalistic conditions and assess the boundaries of these effects. For example, manipulating the number of objects in the visual display, whether participants were familiarized with the images beforehand, and the speaking rate could provide insight into the extent to which our findings are reliant upon the particular design choices we made and our implementation of the VWP. Further, future research could manipulate cue validity by including pluralia tanta and mass nouns in addition to nouns with regular plural forms to assess the extent to which these findings are dependent upon the consistency of the grammatical cue in the experimental context.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements:

Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. 10.1006/jmla.1997.2558

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. 10.1016/s0010-0277(99)00059-1

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. 10.3758/bf03193014

Barr, D. J. (2008a). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. 10.1016/j.jml.2007.09.002

Barr, D. J. (2008b). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40. 10.1016/j.cognition.2008.07.005

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. 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., & Green, P. (2014). Package "lme4" (Version 1.1-15) [Computer software]. R foundation for statistical computing. https://github.com/lme4/lme4/

Brock, J., & Nation, K. (2014). The hardest butter to button: Immediate context effects in spoken word identification. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *67*(1), 114–123. 10.1080/17470218.2013.791331

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. 10.3758/BRM.41.4.977

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(3), 687–696.

Colé, P., & Segui, J. (1994). Grammatical incongruency and vocabulary types. *Memory & Cognition*, *22*(4), 387–394. 10.3758/bf03200865

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*(2), 171–237. 10.1006/jecp.1996.2356

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480. 10.1006/jmla.1999.2688

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498–513. 10.1037/0278-7393.30.2.498

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *71*(4), 808–816. 10.1080/17470218.2017.1310261

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. 10.1111/j.1469-8986.2007.00531.x

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 31–45). Springer New York.

Forster, K. I. (1989). Basic issues in lexical processing. In W. D. Marsen-Wilson (Ed.), *Lexical representation and process* (pp. 75–107). MIT Press.

Fox, N. P., & Blumstein, S. E. (2016). Top-down effects of syntactic sentential context on phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(5), 730–741. 10.1037/a0039965

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1219–1248. 10.1037//0278-7393.22.5.1219

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115.

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

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. 10.1016/j.jml.2017.09.002

Jakubowicz, C., & Faussart, C. (1998). Gender agreement in the processing of spoken French. *Journal of Psycholinguistic Research*, *27*(6), 597–617. 10.1023/A:1023297620824

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. 10.1016/S0749-596X(03)00023-8

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular–plural distinction. *Language Learning and Development*, *2*(1), 1–25. 10.1207/s15473 341lld0201_1

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2153–2162. 10.1037/xlm0000935

Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 326–347. 10.1037/a0034903

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42. 10.1016/j.cognition.2010.12.002

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software. Articles*, *82*(13), 1–26.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198. 10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. 10.1016/j.jml.2010.07.003

Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*(3), 615–625. 10.3758/bf03212113

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. 10.1097/00003446-199802000-00001

Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition*, *146*, 349–370. 10.1016/j.cognition.2015.10.012

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

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition*, *108*(3), 866–873. 10.1016/j.cognition.2008.06.005

W. D., Marslen-Wilson (Ed.). (1989). Access and integration: Projecting sounds onto meaning. *Lexical representation and process* (pp. 3–24). MIT Press.

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

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

Rayner, K., Slowiaczek, M. L., Clifton, C., Jr., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(6), 912–922. 10.1037//0096-1523.9.6.912

Riordan, B., Dye, M., & Jones, M. N. (2015). Grammatical number processing and anticipatory eye movements are not tightly coordinated in English spoken language comprehension. *Frontiers in Psychology*, *6*, 590. 10.3389/fpsyg.2015.00590

Robertson, E. K., Shi, R., & Melançon, A. (2012). Toddlers use the number feature in determiners during online noun comprehension. *Child Development*, *83*(6), 2007–2018. 10.1111/j.1467-8624.2012.01828.x

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. 10.1016/j.jml.2013.11.002

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *SSRN*. 10.2139/ssrn.2160588

Strand, J. F., Brown, V. A., Brown, H. E., & Berg, J. J. (2017). Keep listening: Grammatical context reduces but does not eliminate activation of unexpected words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 962–973. 10.1037/xlm0000488

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

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. 10.1126/science.7777863

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., Thabane, M., Giangregorio, L., Dennis, B., Kosa, D., Borg Debono, V., Dillenburg, R., Fruci, V., Bawor, M., Lee, J., Wells, G., & Goldsmith, C. H. (2013). A tutorial on sensitivity analyses in clinical trials: The what, why, when and how. *BMC Medical Research Methodology*, *13*, 92. 10.1186/1471-2288-13-92

Thothathiri, M., & Snedeker, J. (2008). Give and take: Syntactic priming during spoken language comprehension. *Cognition*, *108*(1), 51–68. 10.1016/j.cognition.2007.12.012

Weber, A., & Crocker, M. W. (2012). On the nature of semantic constraints on lexical access. *Journal of Psycholinguistic Research*, *41*(3), 195–214. 10.1007/s10936-011-9184-0

(*Appendix follows*)

Appendix

All Stimulus Words, by Grid

Grid number	Pair 1, singular	Pair 1, plural	Pair 2, singular	Pair 2, plural
1	cribbage	crickets	saddle	satchels
2	termite	turtles	buckle	buckets
3	penny	pencils	spider	spires
4	peeler	pizzas	rivet	rivers
5	streetcar	streamers	toaster	togas
6	picnic	pickles	forehead	forests
7	maple	maypoles	salute	saloons
8	teapot	teachers	walrus	walnuts
9	outfield	hourglasses	circle	surgeons
10	shaker	shavers	comet	commas
11	angler	anklets	whisker	whistles
12	beetle	beaters	lettuce	letters
13	corner	corals	pillbox	pillars
14	raisin	ravens	pennant	pendants
15	lyre	lions	candy	candles
16	chalice	chapels	rhombus	rompers
17	tulip	tombstones	hamster	hampers
18	barber	barbells	cotton	cottages
19	antler	anvils	locket	lockers
20	carrot	carriages	bowtie	bolos
21	bullet	bulldogs	monarch	monsters
22	harness	harbors	bacon	bakers
23	texan	textbooks	rowboat	robots
24	dollar	dollies	rattle	radishes
25	jersey	juries	weaver	weasels
26	jacket	jackals	palace	pallets
27	knapsack	napkins	cobweb	cobblers
28	money	monkeys	sandwich	sandals
29	mustard	muzzles	level	levers
30	catfish	caverns	wafer	waiters
31	circus	circuits	orchid	organs
32	citrus	cities	gallon	galleys
33	archway	archers	beaker	beagles
34	fungus	funnels	anchor	ankles
35	scuba	scooters	carpet	cartwheels
36	album	alcoves	convent	concerts
37	parsley	parcels	meeting	meters
38	combine	compacts	hammock	hammers
39	boyscout	boilers	mustache	mustangs
40	beaver	beehives	snorkel	snowflakes
41	crater	cradles	necklace	networks
42	champagne	shampoos	attic	atoms
43	collar	collies	server	surfers
44	turban	turkeys	scalpel	scallops
45	reindeer	rainbows	cheater	cheetahs
46	torso	tortoises	leopard	lemons
47	lotion	locusts	ratchet	rackets
48	planter	planets	robber	robins
49	cabbage	cabins	insect	infants
50	boulder	bowlers	cherry	cherubs
51	gerbil	journals	parka	parsnips
52	rocket	rockers	arcade	armoires
53	llama	lobsters	taser	tables
54	banner	banjos	pedal	peppers
55	paddle	padlocks	nozzle	nostrils
56	checklist	checkbooks	dollhouse	dolphins

(*Appendix continues*)

Appendix (*Continued*)

Grid number	Pair 1, singular	Pair 1, plural	Pair 2, singular	Pair 2, plural
57	eagle	easels	panda	pantries
58	paper	paintings	salmon	salads
59	ferret	ferries	princess	printers
60	medal	meadows	panther	pansies
61	halo	haystacks	armor	armpits
62	zippo	zippers	chili	chickens
63	ladle	lasers	winner	windows
64	buzzer	buttons	sofa	sodas
65	puppy	puppets	army	armchairs
66	magnet	maggots	seagull	ceilings
67	lantern	lanyards	aloe	alleys
68	lattice	ladders	charcoal	chargers
69	corkscrew	corsets	musket	muscles
70	camel	cameras	outhouse	outlets
71	cedar	seahorses	coffee	coffins
72	marble	markers	peanut	peacocks

Received September 9, 2020
Revision received January 19, 2021
Accepted January 20, 2021 ■