Visuospatial Working Memory Load Reduces Semantic Prediction in the Visual World

Christopher Allison[a], Falk Huettig[a,b, c], Leigh Fernandez[a], Thomas Lachmann[a]

[a]University of Kaiserslautern-Landau

[b]Max Planck Institute for Psycholinguistics, Nijmegen

[c]Faculty of Psychology, University of Lisbon

--- *Language, Cognition, and Neuroscience* (in press) ---

Corresponding author:

Christopher Allison

allison@rptu.de

Abstract

Prediction in language is often about objects in the language users' visual surroundings. Previous research suggests that linguistic working memory limitations in such task environments constrain language-mediated anticipatory eye movements. In this study, we investigated the effects of visuospatial cognitive load on language-mediated predictive eye gaze behaviour in a diverse group of L2 English speakers using the visual-world paradigm. Participants completed three levels of an increasingly difficult visuospatial working memory task before hearing either semantically constraining or unconstraining sentences, choosing an object best fitting the sentence, and completing the working memory task. Evidence of L2 anticipatory eye gaze was observed in all conditions. Importantly, a significant effect of difficulty, especially in the higher-load condition, suggests that increasing visuospatial working memory reduces anticipatory eye gaze. We close by discussing the importance of (visual) working memory in visual world studies and highlight the inherently integrative nature of predictive processing during language-vision interactions.

Key words: anticipatory eye gaze, language, prediction, working memory, visual world paradigm

Introduction

Prediction in language is a major influence on language processing (e.g., Altmann & Mirković, 2009; Dell & Chang, 2014; Federmeier, 2007; Ferreira & Chantavarin, 2018; Hale, 2001; Hickok, 2012; Huettig, 2015; Huettig et al., 2022; Kuperberg & Jaeger, 2016; Levy, 2008; Norris et al., 2016; Pickering & Garrod, 2013; Van Petten & Luka, 2012). However, the majority of the previous research in predictive language processing has been with manipulations (e.g., unnaturally slow speech rates), settings (e.g., carefully controlled labs), and populations (e.g., monolingual native speakers) that maximize the ability for predictive processing. In recent years, the context-dependence of predictive processing has become a crucial question and the focus of much psycholinguistic research (Pickering & Gambi, 2018, for review).

One unresolved issue, however, is the extent to which prediction might be facilitated or impeded by contextual factors in which language processing occurs. This is an important question with considerable real-world relevance because every-day situations are typically far from ideal for language processing: noisy environments, unexpected input, and unfamiliar contexts are all common. For example, interpreting speech signals in noisy and distorted conditions is cognitively taxing (e.g., Stenfelt & Rönnberg, 2009). Similarly, Wagner et al. (2016) observed that processing degraded speech (manipulated to be similar to the speech one with a cochlear implant would hear) delayed the integration of semantic information. This delay, the authors proposed, may be a consequence of more effortful mapping between the auditory signal and relevant mental representation as a function of a higher degree of mismatch between them.

The ability to rapidly and flexibly link incoming auditory signals with stored mental representations is especially important considering one of the defining features of language: displacement (Hockett & Altmann, 1968). Language does not need to refer to objects that are physically co-present in the environment. Although such a characterization of language is undoubtedly correct, it is noteworthy that language often *does* refer to objects in the language users' surroundings. Thus, we as language users need to be able to rapidly and appropriately be able to apply such linkings. We may ask a dinner guest to "pass the salt", tell a visitor to "mind the step" or ask a child to look at "the cat with milk on its face". There is much evidence that individuals respond to such referential information by orienting their overt visual attention (their eye gaze) towards the mentioned object. In doing this, the linguistically activated mental representation (a *type* representation) is linked to a specific perceptual instance in the real world (a *token* representation; see Mishra et al., 2013 for discussion).

An experimental method that has proven to be particular useful for examining these kinds of language-visual interactions is the visual world paradigm (Tanenhaus et al., 1995). Look and listen tasks (Altmann & Kamide, 1999) in which participants listen to utterances in the context of a visual display without an explicit task, have revealed much evidence for language-mediated anticipatory eye movements to co-present objects that the language may soon refer to. Many of these studies have focused on predictions based on semantic information. In such studies, participants hear a sentence like "*The tailor trims the suit"*, where the target word *suit* is predictable based on semantic information from *tailor* and *trims*. There is ample evidence for such semantic predictions across many different speaker groups: children as young as 2 (Mani & Huettig, 2012), L2 speakers (Dijkgraaf et al., 2019), and people with dyslexia (e.g., Huettig & Brouwer, 2015) or autism (e.g., Huettig et al., 2023; Zhou et al., 2019) have shown the capacity for them. It is, however, important to note that certain aspects of the visual world paradigm interact with or directly encourage predictive processing. Such factors include the mere existence of a visual array and the timing and duration thereof, all of which may discourage more elaborative processing or aid the speed of recognition of spoken words (for review see Huettig et al., 2011).

While the capacity for semantic prediction in language comprehension is clear, there is still active debate regarding its nature. Naturally, contextual factors are particularly relevant when real-world language refers to objects or events in the immediate surroundings of the listener. Combining unfolding linguistic input with the processing of co-occurring objects in the visual environment requires the mapping between linguistic representations and visual object representations. It is also an inherently integrative process that likely requires cognitive resources on 'both sides' of the mapping process.

Limited cognitive resources are likely to be especially challenging for L2 speakers, given the fact that L2 language processing is generally more cognitively demanding (Hopp, 2022) than L1 language processing. L2 speakers often show delayed or weakened predictive gaze behaviour compared to L1 speakers (e.g., Karaca et al., 2021; Schlenter, 2023), are generally slower to predict (e.g., Ito, Pickering, & Corley, 2018), and even quite proficient L2 speakers remain unable to use certain cues for prediction (e.g., Mitsugi & MacWhinney, 2016). That being said, there are multiple studies showing that L2 semantic prediction can be comparable to that of L1 speakers (e.g., Abashidze, 2023; Fernandez et al., 2024; Hopp, 2015). Such findings support the prevailing theory that these differences in L2 predictive capabilities are quantitative and mediated by individual differences (Kaan, 2014; Schlenter, 2023) as opposed to being qualitatively different from L1 speakers. Given all this, resource limitations particularly in working memory (cf. Huettig & Janse, 2016) may hamper predictive gaze behaviour, especially for L2 speakers. It is important to note, however, that the exact contributions of working memory in predictive gaze behaviour remains unclear. While some studies (e.g., Kukona et al., 2016; Otten & Van Berkum, 2009) have found a tenuous relationship between working memory capacity and predictive behaviour, much research seems to suggest that working memory plays a role in predictive processes.

Some valuable insight into the cognitive costs of predictive processing can be seen in how the specific task can modulate predictions. Across two studies, Brothers et al. (2017) showed that (Experiment 1) semantic prediction can be strategically facilitated by asking participants to predict and showed that (Experiment 2) semantic prediction can be modulated by the reliability of predictive cues. Specifically in Experiment 2, they found evidence for prediction in a context where most of the other stimuli were predictable, and no evidence for prediction in a context where most of the other stimuli were unpredictable. While they did not directly test cognitive demand, the experiments suggest that even predictions based on shared semantic characteristics are rapidly subject to top-down, strategic influences. This further suggests a likelihood that such predictions could be influenced by changes in cognitive demand. Flexible, strategically influenced anticipatory processing implies at least some cost in generating or especially in maintaining semantic predictions, as one would expect an automatic, resource-free style of processing to remain consistent regardless of the task.

In an effort to more directly explore the role of cognitive demand on L2 prediction, Chun and Kaan (2019) and Chun et al. (2021) investigated L2 semantic predictive processing of syntactically complex sentences. Specifically, they increased the cognitive demand of a visual world eye-tracking task by increasing both the syntactic complexity of the auditory stimuli (i.e., using sentences with relative clauses complaining complex noun phrases, e.g., *I know the friend of the dancer that will open/get the present*) and by increasing the complexity of the visual display (i.e., using semi-realistic visual arrays containing two agents and three objects). In both studies, L2 listeners successfully used semantic information to predict an upcoming target word, even given the increased cognitive demand of the task. However, both studies found delays in L2 predictive processing. Chun and Kaan (2019) found that the L2 predictions occurred approximately 180 ms later than for L1 listeners, while Chun et al. (2021) found that L2 predictions occurred later for the syntactically complex sentences than they did for semantically equivalent, simple sentences (e.g., *The dancer will open/get the present*). The authors of both studies suggest that language processing for both L1 and L2 processing is thus constrained by the availability of cognitive resources.   Ito, Corley, and Pickering (2018) directly tested working memory resource limitations in L1 and L2 speakers by increasing the working memory demand during a visual world task. Specifically, they had participants remember a five-word list, perform a visual world trial, and then recall the list. Both L1 and L2 participants exhibited significantly reduced predictive gaze behaviour during the visual world task during the additional memory load condition, with L2 speakers only showing significantly increased looks to target 100 ms after target onset. This result is consistent with the notion that working memory capacity limitations influenced participant performance; those participants who performed visual world trials with more concurrent cognitive demands showed less predictive gaze behaviour than those who performed the same trials without extra cognitive demands.

It is important to note that the word-list manipulation used by Ito, Corley, and Pickering (2018) was linguistic in nature. When language refers to objects in the surrounding visual environment, linguistic working memory may not be the only type of working memory involved. For example, the classic working memory model of Baddeley (1992) includes a phonological loop (assumed to deal with linguistic input) and a visuospatial sketch pad (assumed to deal with visuospatial input). In line with such a view, the word-list working memory manipulation may have specifically loaded phonological working memory, thus leaving the possibility that capacity limitations in the visuospatial sketch pad could also constrain language-mediated anticipatory eye movements.

Current study

Here we tested whether increased visuospatial cognitive load can reduce language-mediated predictive gaze behaviour by using a visuospatial, within-participants, cognitive load manipulation during visual world eye-tracking trials. There is much evidence, for example from blank screen studies (Altmann, 2004; Ferreira et al., 2008; Spivey & Geng, 2001) that the visual arrays used in in visual world studies are spatially encoded: there is a tendency for participants to look at locations in the array that were previously occupied by relevant objects even when these objects are no longer visible. Combined with the observations that spatial information about an object is stored when objects are stored in visual working memory (e.g., Jiang et al., 2000), we hypothesized that a visuospatial cognitive load manipulation would interfere with predictive gaze behaviour.

We used a modified Corsi block tapping task (Corsi, 1972) to create no-load, lower-load, and higher-load conditions. The Corsi task is widely used to measure visuospatial working memory. It involves the spatial encoding of multiple objects and, importantly, shows no evidence of verbal reencoding (Vandierendonck et al., 2004). Thus, the Corsi task is a relatively “pure” visuospatial working memory task and any disruptions in predictive gaze behaviour due to this task are not likely to be due to direct linguistic interference.

Following a procedure similar to that of Ito, Corley, and Pickering (2018), participants first performed the visuospatial working memory manipulation, then performed a visual world trial, and ended by recalling the working memory information. Participants were presented with an array of 9 white squares and were required to encode the visuospatial location of 0 (no-load), 2 (lower-load) or 4 (higher-load) of these squares before performing the visual world trial and recall this sequence after the visual world trial.

Continuing the design of Ito, Corley, and Pickering (2018), we used a diverse group of L2 English participants. This study was not concerned with comparing L1 and L2 gaze behaviour. Since previous research suggests that L2 language processing is generally more demanding (Hopp, 2022), we simply reasoned it to be more likely for L2 speakers to encounter resource limitations and thus to observe an impact of cognitive load on predictive processing.

Methods

Participants

Forty-five L2 English speakers from the University of Kaiserslautern-Landau between ages 20 and 34 participated in the experiment and received either 10€ or participation credit. English proficiency was assessed via a subset of the Oxford Placement Test and only data from participants that scored above 50% on this assessment were included. One person scored below the inclusion threshold for the proficiency test and was excluded. Thus, data from 44 participants (mean age = 25.75, SD = 2.9, range = 20-34; mean proficiency = 74.09, SD = 10.65, range = 52-94) were analysed. These remaining participants had the following native languages: Turkish (9), Hindi (6), Malayalam (5), Marathi (4), Persian (4), Arabic (3), Telugu (3), Kannada (2), Tamil (2), Chinese (1), German (1), Greek (1), Gujarati (1), Indonesian (1), and Urdu (1). These participants started learning English at an average age of 6.5 (SD =3.48) and had the following self-rated English scores, with 10 representing a native-like level: speaking – 8.14 (SD = 1.42), understanding – 8.78 (SD = 1.10), reading – 8.97 (SD = 1.05) and writing – 8.09 (SD = 1.43). All participants reported normal or corrected to normal hearing and vision and none reported any neurological impairments. All participants provided written consent. The study was approved by the University of Kaiserslautern-Landau Ethics Committee of the Faculty of Social Sciences.

Materials

We used 48 auditory sentence pairs and visual arrays from Fernandez et al. (2024) for a look and listen visual world study. These sentences were recorded by an early 30’s male who was a native Scottish-English speaker. In each sentence pair, the predictability of the critical object was manipulated by changing the agent and the verb of the sentence (e.g., predictable: *The waiter brings the plate* or unpredictable: *The runner remembers the plate*). Each sentence pair had a corresponding visual array consisting of four objects (e.g., pictures of a plate, a scarf, a window, and a parking garage) with one object in each of the four corners (see Figure 1 for an example array). For the predictable sentence in the pair, one object was predictable (plate), one object was plausible but not predictable (scarf), and two objects were neither plausible nor predictable (garage, window). For the unpredictable sentences, all objects were plausible but unpredictable. All objects were 300x300 pixel greyscale drawings taken from the MultiPic database (Duñabeitia et al., 2018).

[figure 1 goes here]

Each spoken sentence consisted of five words (i.e., The Agent Verb The Object) and was 1903.07 ms long (see Table 1). The fixed length of each sentence was accomplished by manually expanding or compressing the recording of each word to the global mean of each sentence position (e.g., the mean utterance length of every word in the "Agent" position was calculated and all "Agent" words were normalized to this length). The resulting mean speech rate of the sentences was 3.47 (SD = 0.77) syllables per second (range 2.56 – 5.65 syllables per second).

[table 1 goes here]

Cognitive load was manipulated using a modified version of the visuospatial Corsi block tapping task (Corsi, 1972). Three cognitive load conditions were used: no-load, low-load, and high-load and the task was divided into an encoding and a recall phase with a Visual World trial in between. The encoding phase in each condition began with the presentation of 9 randomly located blank white squares. In the low- and high-load conditions, either 2 or 4 of the squares (respectively) were indicated by a 500ms colour change from white to dark grey and participants were instructed to remember the order and location of any indicated squares. In the no-load condition, participants saw the grid for 1500 ms (with no indications). In the low- and high-load conditions, participants saw the grid for 500 ms before a square was indicated for 500 ms. There was an interval of 500 ms between squares being indicated. In the recall phase, participants were again presented with the grid of 9 white squares and were required to click the squares in the order and location that was previously indicated (see Figure 2). The Corsi task was chosen specifically to minimize any linguistic interference during the Visual World trials.

We used a blocked, within-subjects design with increasing difficulty per block. Specifically, participants completed a block of no-load trials, followed by a block of low-load trials, and then a block of high-load trials. For each participant, the 48 sentence pairs were randomly assigned to one of the three cognitive load conditions, resulting in 16 trials per condition. From these trials, 8 predictable and 8 unpredictable trials were randomly chosen. This resulted in each participant being presented with a randomized, unique set of items in each of the conditions. Participants completed the Language and Social Background Questionnaire (LSBQ; Anderson et al., 2018). This questionnaire provided information about neurological and developmental disorders as well as information on when, where, and how participants learned and used English. Participants also completed a subsection of the Oxford Placement Test (OPT) as a measure of English proficiency.

Procedure

Participants were individually tested in a dedicated room with a 50 cm viewing distance to a 1024 x 768 pixel resolution CRT monitor and their eye movements were recorded using a head-mounted SR Research Eyelink 1000 sampling at 1000 Hz recording the right eye. The participants were instructed to remember the order and location of the squares, listen to the spoken sentence (presented through Philips Bass+ on-ear headphones), click the picture best represented by the sentence, and then choose the squares that were indicated at the start of the trial. The eye-tracker was then calibrated with a nine-point grid and participants completed two practice trials before the 48 experimental trials.

Each trial began with a drift correction in the centre of the screen followed by the block-dependent cognitive load manipulation. Participants had a 2000 ms preview of the visual array before listening to a predictable or unpredictable sentence. Participants then chose the most fitting picture after hearing the entire utterance. In the no-load condition, trials ended upon picking the picture. In the low- and high-load conditions, participants were then presented with the grid from the start of the trial and had to click the squares in the correct serial order. After finishing the 48 experiment trials, participants completed the LSBQ and OPT.

Results

Behavioural tasks

Accuracy in the comprehension task (i.e., selecting the correct object from the visual array) was 98.2% in the no-load condition, 98.7% in the low-load condition, and 99.1% in the high-load condition. Incorrect trials were excluded from further analysis. Accuracy in the cognitive load manipulation task (i.e., successfully recalling the order and location of the indicated squares) was 89.3% in the low-load condition and 75% in the high-load condition. Participants were significantly more accurate when completing the low-load condition than the high-load condition ($t(43) = 4.93$, $p < .001$), indicating that the high-load condition was indeed more difficult.

Eye-tracking analyses

Figure 3 shows the time course of the target fixation proportion in the predictable condition for each of the three cognitive load conditions. Timing was consistent between all sentences and thus the x axis (Time) represents the actual time in the sentence (each sentence was 1903 ms long). To account for saccade timing, 200 ms were added to both the verb onset (i.e., agent offset; 705 ms) and the target onset (1437 ms) and these times are marked by dotted lines on the graphs. These values defined the predictive timeframe that was analysed. We chose to start analysis at the agent offset as only then can we be sure that predictive gaze behaviour is based on the agent information and not a word with a phonologically similar onset (e.g., *whale* instead of *waiter*).Visual inspection shows clear evidence for prediction in all three conditions and reduced predictive gaze behaviour in the load conditions in the predictive window.

R (R Core Team, 2022), the VWPre package (Porretta et al., 2016), and the lme4 package (Bates et al., 2015) were used to process and analyse the eye-tracking data. Blinks and looks outside of the 300 x 300 pixel pictures of the visual array were recorded and included in the data. Proportion data for looks to each area of interest were calculated in 50 ms bins and transformed and the log-ratio for looks to target to looks to nontarget were calculated using the following formula: log((proportion of looks to target + 0.5) / (mean proportion of looks to nontarget + 0.5)). We analysed the log-ratio data for looks to target vs nontargets using a linear mixed effect model testing the effect of *difficulty* (no-load, low-load, high-load) and *predictability* (predictable, unpredictable). The data were aggregated across the pre-defined time window from the verb onset + 200 ms until target onset + 200 ms and grouped by subject, item, difficulty, and predictability.

The variables *difficulty* and *predictability* were dummy-coded with the reference levels of *no-load* and *predictable*, respectively. We fit a maximal model which resulted in the following: log-ratioTarget ~ difficulty*predictability + (1+difficulty*predictability|Subject) + (1+difficulty*predictability|Trial). We ran this model in the pre-defined time window and used |t|>2 as the threshold for a significant effect.

A summary of the results can be seen in Table 2. The analysis confirmed a significant effect of predictability (β = -0.34, SE = 0.05, t = -7.5; see Figure 3 and Figure 4 for visualisation). Model comparison confirms a significant negative effect of difficulty X^2 = 6.9, p = .03, indicating a significant reduction in predictive gaze behaviour with added difficulty. Particularly, predictive gaze behaviour was significantly reduced in the higher-load condition (β = -0.12, SE = 0.05, t = -2.6). We also found significant interactions between predictability and difficulty in both the low-load (β = -0.14, SE = 0.06, t = -2.4) and the high-load (β = -0.14, SE = 0.06, t = -2.5) conditions. A follow-up analysis of the predictable and unpredictable conditions separately reveals a significant effect of the high-load condition in the predictable condition (β = -0.12, SE = 0.04, t = -2.7) but no effect in the unpredictable condition (β = 0.02, SE = 0.04, t = 0.6). The follow-up analysis also reveals no significant effect of the low-load condition in the predictable condition (β = -0.05, SE = 0.04, t = -1.2). In the unpredictable condition, we see a likely spurious effect of the low-load condition (β = 0.1, SE = 0.04, t = 2.4). We say likely spurious for two reasons: (1) the condition is unpredictable and thus we cannot influence predictive gaze behaviour in this timeframe, and (2), visualization (see Figure 4) shows a short increase in “predictive” looks in the unpredictable, low-load condition that quickly returns to baseline, whereas the effects in the predictable condition (Figure 3) are consistent across the entire predictive timeframe.

We also ran the same model in the time window from target onset + 200 ms until the end of the sentence. This analysis showed a significant effect of predictability (β = -0.43, SE = 0.07, t = -6.3) in this post-target time window. However, we see no significant effect of either of the load conditions and no significant interaction between predictability and load in this time window.

[table 2 goes here]

[figure 3 goes here]

[figure 4 goes here]

Discussion

This study investigated the effects of increasing visuospatial cognitive load on anticipatory eye gaze behaviour in skilled L2 English speakers. To do this, we conducted a visual-world eye tracking experiment in which participants received a visuospatial cognitive load manipulation before listening to predictable or unpredictable sentences with a visual array of four pictures. Participants first completed a no-load block of trials in which they were presented with a random array of 9 blank squares before completing the visual world task but were not tasked with remembering the location of any of the squares. Then, participants completed a block of lower-load trials in which they were presented with a random array of 9 squares and were tasked to remember and recall two squares after the visual world trial. Finally, they completed a block of higher-load trials in which they had to remember and recall four of the squares.

First, the results of this study highlight the robustness of semantic prediction in L2 speakers. In all three conditions, participants were able to predict upcoming visual referents, i.e., they looked to the target before it was explicitly mentioned when listening to predictable sentences. This is also one of few studies to show L2 predictive gaze behaviour using a speech rate typical of real speech. Speech rate is widely underreported in visual world research and faster speech rates have been found to reduce predictive gaze behaviour in both L1 and L2 speakers (e.g., Huettig & Guerra, 2019; Fernandez et al., 2020). The standard spoken stimuli used in the visual world paradigm are often presented at speech rates that would be uncommon in real world environments in which spoken language is used in order to allow more time for predictions to occur. Our study suggests that, at least for semantic prediction based on semantically constraining information, slower speech rates are not necessary for prediction to occur and L2 speakers can predict upcoming information at real-life speech rates even with increased cognitive load.

Secondly, we found a robust reduction in predictive gaze behaviour in more cognitively demanding conditions, especially in the higher visual-load condition. Thus, increasing visuospatial cognitive load interferes with predictive gaze behaviour. This finding combines well with those of Ito, Corley, and Pickering (2018) to show the importance of working memory when language is used to refer to co-present objects. The two studies are methodologically similar, differing primarily in the type of load task used: either visuospatial in our case, or linguistic/phonological in their case. Taken together, these two findings suggest two main possibilities: either (1), that any type of additional task demand may lead to a delay of semantic predictions, or (2), that these two types of increased cognitive demand specifically interfere with the two main aspects of the visual world paradigm, namely the visuospatial encoding of the visual array and the phonological processing of the spoken stimuli. Furthermore, these two types of load task seem to lead to considerably different outcomes on predictive gaze behaviour, with a linguistic task almost completely eliminating gaze behaviour in L2 speakers (though further work is necessary to confirm this conclusion as it relies on a comparison across different experiments, labs, participants, and materials). If reliable, however, this suggests that the predictive disruptions are more likely to be some form of specific interference in language processing than a disruption due to a more domain general cognitive demand. This highlights the importance of the specifics of the cognitively demanding task when examining the effect of "cognitive load" on predictive gaze behaviour.

Further research is necessary to elucidate the mechanisms of the *visual* working memory influences on anticipatory eye movements. There are at least two accounts that are compatible with these results. For one, the findings fit with the aforementioned Baddeley model of working memory (Baddeley, 1992) in that predictive gaze behaviour is reduced both when "loading" the visuospatial sketchpad and when "loading" the phonological loop. Predictive eye gaze behaviour when language refers to visually co-present objects (as evidenced by participant performance in the visual world paradigm) is directly affected both by the visuospatial representations of the visual array and the phonological representations in the phonological loop. These processes may involve mappings between language-derived representations (from the spoken language input) and visually-derived representations (from the visual input) at several levels of representations (as proposed by Huettig & McQueen, 2007) or at the visual level only (as proposed by Dahan & Tanenhaus, 2004). Further experimentation is required to distinguish the specific mapping-levels involved.

A second type of account compatible with the present results has it that, instead of two dissociable processes being separately impacted, linguistic and non-linguistic representations activated by spoken and visual input respectively share a *common representational substrate* (Altmann & Mirković, 2009). The key idea underpinning this proposal is that unfolding language activates not only upcoming linguistic possibilities, but also upcoming conceptualizations of the event itself. Anticipatory eye movements in this view are a consequence of a *common code* reflecting a mental world comprised of joint (linguistic, visual, and conceptual) event representations. Consider, for example, hearing the classic visual world example of “the boy will eat …”. Are we predicting “the cake”, or are we predicting a likely event given the context? The common coding account suggests that the common event representations activated by seeing the cake and hearing “the boy will eat” are what directs predictive gaze behaviour towards the cake. Both phonological load and visuospatial cognitive load would hence interfere with anticipatory eye gaze behaviour according to this theoretical approach. Further research could usefully be conducted to distinguish the representational mappings and common event codes accounts.

To conclude, this study provides the first direct experimental evidence that increases in visuospatial cognitive demand interfere with predictive gaze behaviour. Language-mediated anticipatory eye movements as evidenced by the visual world paradigm are thus likely to require cognitive resources that are involved in visuospatial processing. These findings highlight the fact that language in the context of co-present visual objects requires the integration of both visual and linguistic representations, either through mapping across levels of representations or common event coding.

Disclosure of interest

The authors report no conflict of interest.

Data availability

The data that support the findings of this study are openly available on the Open Science Framework at https://osf.io/8vwz4/?view_only=bff9b52597354cc9829e14196b7d0932

References

Abashidze, D., Schmidt, A., Trofimovich, P., & Mercier, J. (2023). Integration of visual context in early and late bilingual language processing: evidence from eye-tracking. *Frontiers in Psychology*, 14(April). https://doi.org/10.3389/fpsyg.2023.1113688

Altmann, G. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*(2), B79-87. https://doi.org/10.1016/j.cognition.2004.02.005

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G., & Mirković, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science*, *33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, 93, 203-216. https://doi.org/10.1016/j.jml.2016.10.002

Chun E, Chen S, Liu S and Chan A (2021) Influence of syntactic complexity on second language prediction. In E. Kaan and T. Grüter (Eds.), *Prediction in Second Language Processing and Learning* (pp. 69–89). Amsterdam: John Benjamins Publishing Company.Chun, E., & Kaan, E. (2019). L2 Prediction during complex sentence processing. *Journal of Cultural Cognitive Science*, 3(2), 203-216. https://doi.org/10.1007/s41809-019-00038-0

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain.*

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. https://doi.org/10.1037/0278-7393.30.2.498

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience*, *34*(7), 881–900. https://doi.org/10.1080/23273798.2019.1591469

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 (2006)*, *71*(4), 808–816. https://doi.org/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. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G., & Allen, S. E. (2020). Effects of speech rate on anticipatory eye movements in the visual world paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology*, *73*(12), 2348–2361. https://doi.org/10.1177/1747021820948019

Fernandez, L. B., Hadley, L. V., Koç, A., Gamboa, J. C., & Allen, S. E. (2024). Is there a cost when predictions are not met? A VWP study investigating L1 and L2 speakers. *Quarterly Journal of Experimental Psychology*, *0*(0). https://doi.org/10.1177/17470218241270200

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007

Ferreira, F., & Chantavarin, S. (2018). Integration and Prediction in Language Processing: A Synthesis of Old and New. *Current Directions in Psychological Science*, *27*(6), 443–448. https://doi.org/10.1177/0963721418794491

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In NAACL '01: *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies.* https://doi.org/10.3115/1073336.1073357

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145. https://doi.org/10.1038/nrn3158

Hockett, C., & Altmann, S. (1968). A note on design features. In T. Sebeok (Ed.). *Animal Communication: Techniques of Study and Results of Research*, 61–72.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *IRAL - International Review of Applied Linguistics in Language Teaching,* 53(3), 277–306. https://doi.org/10.1515/iral-2015-0014

Hopp, H. (2022). Second Language Sentence Processing. *Annual Review of Linguistics*, *8*(1), 235–256. https://doi.org/10.1146/annurev-linguistics-030821-054113

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*, 105050. https://doi.org/10.1016/j.cognition.2022.105050

Huettig, F., & Brouwer, S. (2015). Delayed Anticipatory Spoken Language Processing in Adults with Dyslexia—Evidence from Eye-tracking. *Dyslexia*, *21*(2), 97–122. https://doi.org/10.1002/dys.1497

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

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

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

Huettig, F., Voeten, C. C., Pascual, E., Liang, J., & Hintz, F. (2023). Do autistic children differ in language-mediated prediction? *Cognition*, *239*, 105571. https://doi.org/10.1016/j.cognition.2023.105571

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264. https://doi.org/10.1017/s1366728917000050

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. https://doi.org/10.1016/j.jml.2017.09.002

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 683–702. https://doi.org/10.1037//0278-7393.26.3.683

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Karaca, F., Brouwer, S., Unsworth, S., & Huettig, F. (2021). Prediction in bilingual children: The missing piece of the puzzle. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 116–137). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.06kar

Kukona, A., Braze, D., Johns, C.L., Mencl, W. E., Van Dyke, J. A., Magnuson, J. S., Pugh, K. R., Shankweiler, D. P. & Tabor, W (2016). The real-time prediction and inhibition of linguistic outcomes: Effects of language and literacy skill. *Acta Psychologica*, 171, 72-84. https://doi.org/10.1016/j.actpsy.2016.09.009

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

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

Mitsugi, S., & MacWhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35. https://doi.org/10.1017/s1366728914000881

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18. https://doi.org/10.1080/23273798.2015.1081703

Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, 1291(0), 92-101. https://doi.org/http://dx.doi.org/10.1016/j.brainres.2009.07.042

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/s0140525x12001495

Porretta, V., Kyröläinen, A.-J., van Rij, J., & Järvikivi, J. (2016). VWPre: Tools for Preprocessing Visual World Data.

R Core Team (2022). R: A Language and Environment for Statistical Computing. https://www.R-project.org/

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*(2), 253–267. https://doi.org/10.1017/s1366728922000736

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*(4), 235–241. https://doi.org/10.1007/s004260100059

Stenfelt, S., & Rönnberg, J. (2009). The signal-cognition interface: Interactions between degraded auditory signals and cognitive processes. *Scandinavian Journal of Psychology*, *50*(5), 385–393. https://doi.org/10.1111/j.1467-9450.2009.00748.x

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

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, *95*(1), 57–79. https://doi.org/10.1348/000712604322779460

Wagner, A., Pals, C., Blecourt, C. M. de, Sarampalis, A., & Başkent, D. (2016). Does Signal Degradation Affect Top–Down Processing of Speech? In *Physiology, psychoacoustics and cognition in normal and impaired hearing,* 894(1), 297–306. https://doi.org/10.1007/978-3-319-25474-6_31

Zhou, P., Zhan, L., & Ma, H. (2019). Predictive Language Processing in Preschool Children with Autism Spectrum Disorder: An Eye-Tracking Study. *Journal of Psycholinguistic Research*, *48*(2), 431–452. https://doi.org/10.1007/s10936-018-9612-5

Tables

Table 1. Standardized format of the sentences, the normalized length of each word in the sentences (ms), and example sentence pair.

	THE	AGENT	VERB	THE	OBJECT	Total (ms)
Predictable	The	waiter	brings	the	plate	1903.07
Unpredictable	The	runner	remembers	the	plate	1903.07
Length (ms)	93.58	612.72	602.05	130.27	464.45	1903.07

Table 2. Results of the mixed effects model.

Effect	Estimate (SE)	t
Intercept	0.26 (0.04)	6.1
Predictability	-0.34 (0.05	-7.5
Lower-load	-0.5 (0.04)	-1.4
Higher-load	-0.12 (0.06)	-2.5
Lower x Pred	-0.14 (0.06)	-2.4
Higher x Pred	-0.14 (0.06	-2.5

Figures

Figure 1. Example of a visual array that accompanied the sentence pair The waiter/runner brings/remembers the plate.

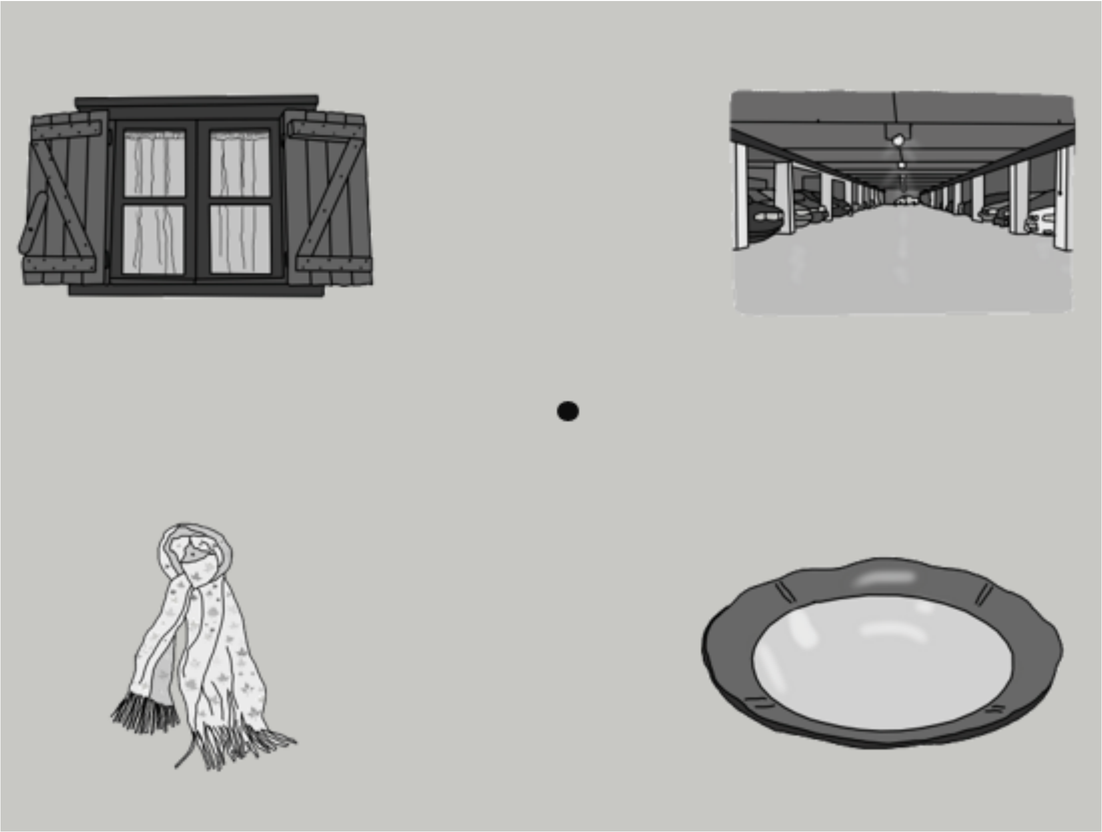


Figure 2. Example of the modified Corsi block task. Either two (low-load condition) or four (high-load condition) of the squares would be indicated by changing to a dark gray color for 500 ms. There was a 500 ms interval between the indication of the squares. Participants were instructed to remember the order and location of any indicated squares.

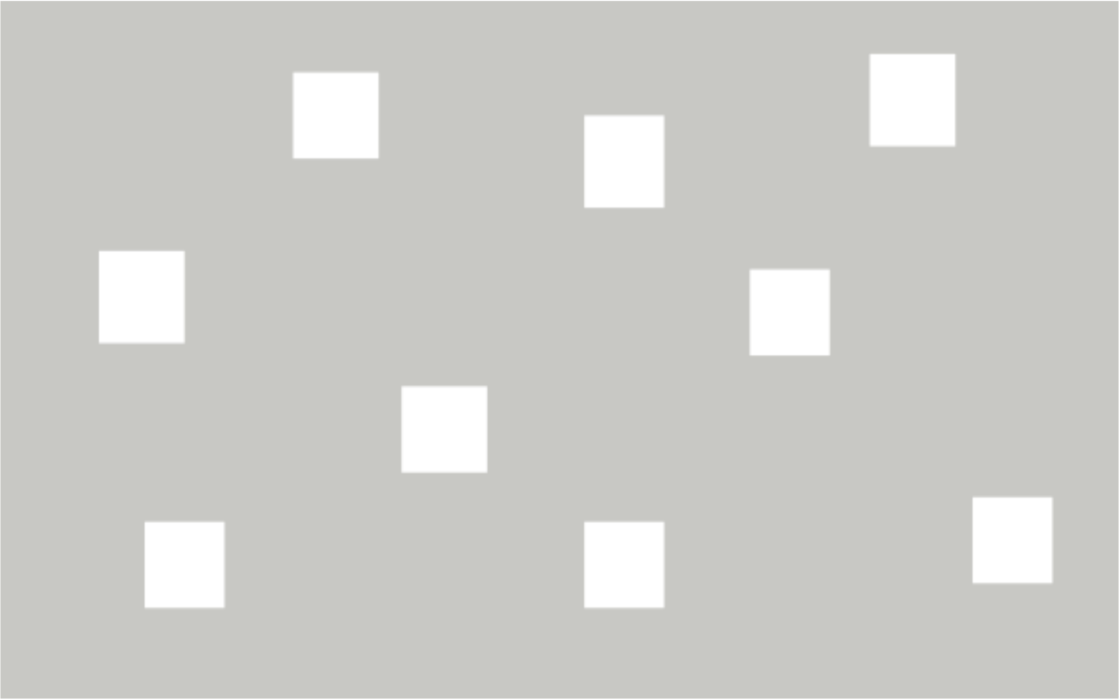


Figure 3. Fixation proportion data to the predictable target in the three cognitive load conditions. Looks in the no-load, lower-load, and higher-load condition are represented by green, yellow, and red lines, respectively. The dotted lines represent an averaged look to the distractors. The bands surrounding the lines represent ± 1 SE. Dotted grey lines at 293, 905, and 1637 ms represent agent onset + 200 ms, verb onset + 200 ms and target onset + 200 ms, respectively.

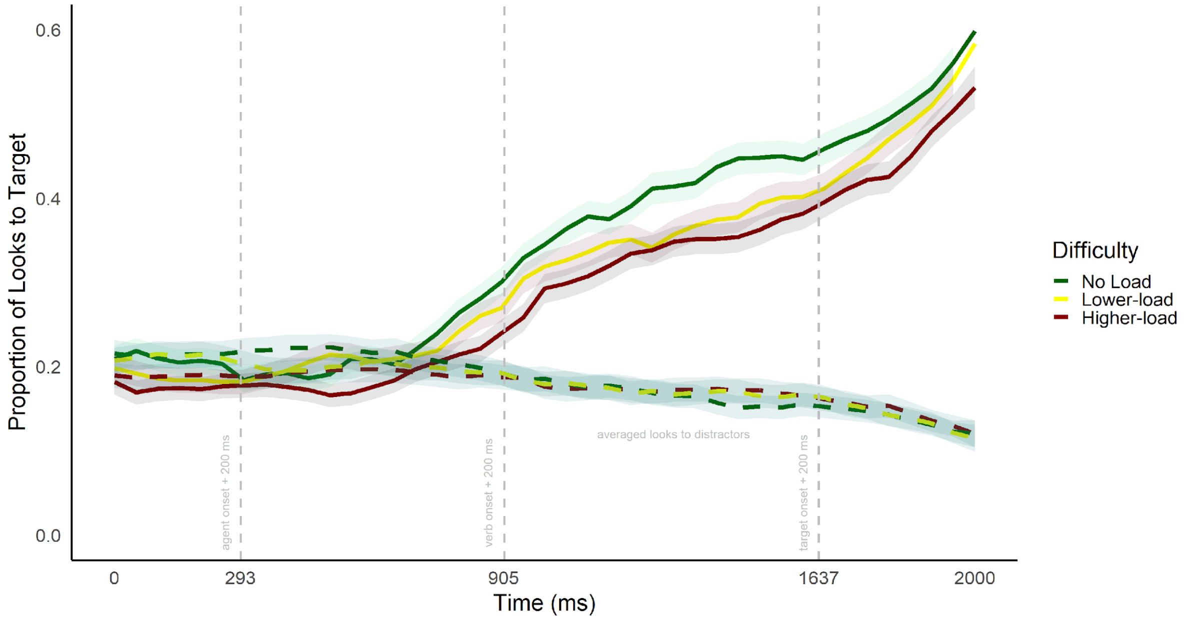
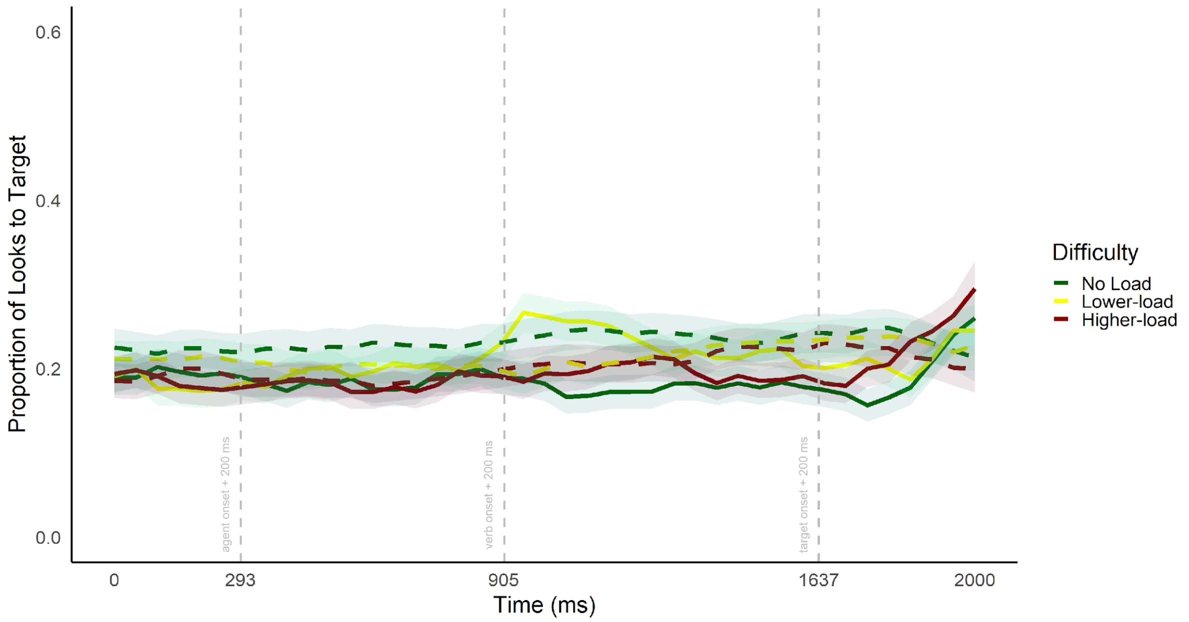


Figure 4. Fixation proportion data to the unpredictable target in the three cognitive load conditions. Looks in the no-load, lower-load, and higher-load condition are represented by green, yellow, and red lines, respectively. The dotted lines represent an averaged look to the distractors. The bands surrounding the lines represent ± 1 SE. Dotted grey lines at 293, 905, and 1637 ms represent agent onset + 200 ms, verb onset + 200 ms and target onset + 200 ms, respectively.


Proportion of Looks to Target
agent onset + 200 ms
verb onset + 200 ms
target onset + 200 ms
Time (ms)
Difficulty
No Load
Lower-load
Higher-load