# COGNITIVE DEMANDS OF PREDICTIVE LANGUAGE PROCESSING: THE ROLES OF SPEECH RATE AND VISUOSPATIAL WORKING MEMORY

Vom Fachbereich Sozialwissenschaften
der Rheinland-Pfälzischen Technischen Universität, Campus Kaiserslautern
zur Verleihung des akademischen Grades
Doktor rerum naturalium (Dr. rer.nat.)
genehmigte

D i s s e r t a t i o n

vorgelegt von
Christopher Todd Allison

| | |
|---|---|
| Tag der Disputation: | Kaiserslautern, 25.02.2026 |
| Dekanin: | Prof. Dr. Shanley Allen |
| Vorsitzende/r: | Prof. Dr. Johannes Lenhard |
| Gutachter/in: | 1. Prof. Dr. Thomas Lachmann |
| | 2. Prof. Dr. Falk Huettig |

DE 386

März 2026

## Statement of Authorship

Hiermit versichere ich,

- dass ich die vorgelegte Arbeit selbst angefertigt und alle benutzten Hilfsmittel in der Arbeit angegeben habe,
- dass ich diese Dissertation nicht schon als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht, und
- dass weder die gleiche noch eine andere Abhandlung der Dissertation bei einer anderen Universität oder einem anderen Fachbereich der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau veröffentlicht wurde.

4. März 2026, Kaiserslautern

Christopher Todd Allison

# TABLE OF CONTENTS

# LIST OF FIGURES AND TABLES

## Frame Text



## Study 1

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.

## Study 2



## Study 3

Table 3. Mean normalized constituent durations (ms)

Figure 1. Mean fixation proportion to all objects across language and sentence type (correctly answered items only).

Table 4. First-stage prediction GAMM output for the empirical logit during the prediction time window. Part A reports the parametric coefficients and Part B reports the smooth terms.

Figure 2. Contour plots for the empirical logit of the three-way interaction between time, speech rate and language group. Top left panel. Contour plot for L2 speakers. Top right panel. Contour plot for L1 speakers. In both of the top panels, negative values (i.e., green) indicate lower empirical logits (looks to the target) and positive values (i.e., pink) indicate greater empirical logits. Bottom left panel. The difference between the L2 and L1 speakers. In the bottom panel, the negative values indicate greater empirical logits (looks to the target) by L1 speakers relative to L2, and positive values indicate greater empirical logits by L2.

Figure 3. Divergence point and 95% confidence intervals superimposed on the fixation proportion of looks to the target and verb-related object.

Table 5. Second-stage prediction GAMM output for the empirical logit during the prediction time window. Part A reports the parametric coefficients and Part B reports the smooth terms.

Figure 4. Contour plots for the empirical logit of the three-way interaction between time, speech rate and language group. Top left panel. Contour plot for L2 speakers. Top right panel. Contour plot for L1 speakers. In both of the top panels, negative values (i.e., green) indicate lower empirical logits (looks to the target) and positive values (i.e., pink) indicate greater empirical logits. Bottom left panel. The difference between the L2 and L1 speakers. In the bottom panel, the negative values

indicate greater empirical logits (looks to the target) by L1 speakers relative to L2, and positive values indicate greater empirical logits by L2.

Figure 5. Divergence point and 95% confidence intervals superimposed on the fixation proportion of looks to the target and agent-related object.

## ABSTRACT

While there is now ample evidence of prediction in language processing, the overwhelming majority of this evidence comes from “prediction encouraging set-ups”. Thus, while many theories of language processing posit an inherent role of prediction in language processing, the reliance on prediction encouraging set-ups and the findings that some manipulations drastically reduce prediction (e.g. concurrent phonological demand) or that some groups (e.g., non-native speakers, illiterate adults, children, older adults) show reduced prediction have led some to question the ubiquity of predictive language processing. Two standard parts of the “prediction encouraging set-up” are slow speech rates and a minimally demanding processing context/environment. The three studies included in this dissertation manipulate both of those, specifically by using faster speech rates and a concurrent nonlinguistic working memory demand, during the also more demanding task of non-native predictive language processing. I report evidence that simple semantic predictions in non-native English speakers are generally robust to both increases in speech rate and concurrent working memory demands. However, either gaze behavior or predictions that require language/vision interactions as well as predictions that require combining multiple representations may be reduced as demand increases.

## SUMMARIES OF INCLUDED ARTICLES

### Study 1.

Allison, C., Huettig, F., Fernandez, L., & Lachmann, T. (2025). Visuospatial working memory load reduces semantic prediction in the visual world. *Language, Cognition, and Neuroscience*, 1-10. DOI: 10.1080/23273798.2025.2522272



**Author Contribution**

**Allison, C**: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Project administration. **Huettig, F**: Conceptualization, Formal analysis, Writing – Original Draft, Writing – Review & Editing, Supervision. **Fernandez, L**: Conceptualization, Validation, Formal analysis, Resources, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision. **Lachmann, T**: Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Funding acquisition.

Study 1 establishes the effect of a concurrent visuospatial working memory demand on L2 predictive processing. In this study, a diverse group of proficient L2 English speakers completed visual world eye-tracking trials. Before some trials, participants needed to encode the order and location of a number of squares (2 in some trials, 4 in others) that were indicated before the visual world trial and to recall these squares after the visual world trial. We found that L2 English speakers showed predictive looks to targets in all three conditions despite L2 language processing at normal speech rates with a concurrent visuospatial working memory task. Furthermore, we found a significant effect of the concurrent visuospatial working memory tasks: participants were less likely to predictively look towards a target word when there was a concurrent visuospatial demand. This effect was stronger (i.e., the working memory task was more disruptive) when more information had to be remembered (i.e., 4 squares instead of 2). However, due to the use of the visual world paradigm and the corresponding visual array, it is unclear if this reduction is happening at the level of word/feature pre-activation or if this effect is instead contingent on co-occuring visuospatially represented objects.

## Study 2.

Allison, C., Huettig, F., Lachmann, T., Allen, SEM, & Beck, AK. (in review). Are Visuospatial Working Memory Influences on Prediction in Speech Contingent on the Co-presence of a Relevant Visual World? An ERP study. *Language, Cognition, and Neuroscience.*



### Author Contribution

**Allison, C**: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Project administration. **Huettig, F**: Conceptualization, Formal analysis, Writing – Original Draft, Writing – Review & Editing, Supervision. **Lachmann, T**: Conceptualization, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision, Funding acquisition. **Allen, SEM**: Conceptualization, Methodology, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision, Funding acquisition. **Beck, AK**: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project administration

Study 2 investigates the effects of the same visuospatial working memory task as study 1. However, this time we used EEG and looked at the N400 ERP component so that we can see the effects without the visual array inherent to the visual world paradigm. This study also used a similar group of L2 English participants as study 1 to maximized comparability. Participants heard predictable or unpredictable sentences either with a concurrent visuospatial working memory demand (remember 4 squares) or without a concurrent demand. We found substantial evidence supporting an effect of predictability on average N400 amplitude and substantial evidence against an effect of a concurrent visuospatial working memory demand or an interaction between predictability and working memory. These findings suggest that the degree of pre-activation of a target word or its features is not reduced by a concurrent visuospatial working memory task and that, instead, the effects of visuospatial working memory that we see in eye-tracking studies are likely contingent on co-present visual objects.

## Study 3.

Fernandez L.B., Hadley L.V., Gamboa J.C.B., Allison C., & Allen S.E.M. (2025). The impact of speech rate on first- and second-stage prediction in L1 and L2 speakers. *Bilingualism: Language and Cognition*, 1-13. DOI: 10.1017/S1366728925100515

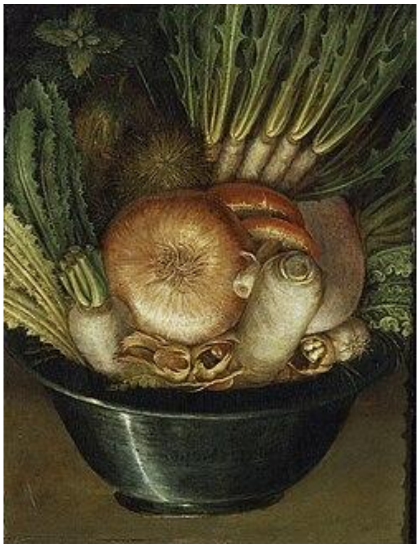


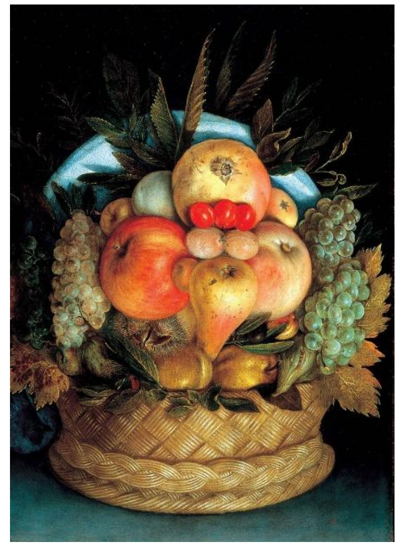
**Author Contribution**

**Fernandez, LB**: Conceptualization, Methodology, Investigation, Project Administration, Software, Validation, Formal Analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Hadley, LV**: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Funding Acquisition. **Gamboa, JCB**: Methodology, Investigation, Project Administration, Writing - Original Draft, Writing - Review & Editing. **Allison, C**: Conceptualization, Methodology, Formal Analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Allen, SEM**: Conceptualization, Methodology, Writing – Original Draft, Writing - Review & Editing, Supervision, Funding Acquisition.

Study 3 investigates the effects of speech rate on first and second stage predictions in L1 and L2 English speakers. The first stage predictions were based solely on semantic associations, while the second stage predictions required a combination of words to derive the correct target. We found that L1 and L2 speakers made both first- and second stage predictions similarly at slow (2.6 syllables per second) and middle (3.6 syllables per second) speech rates. However, faster speech rates (4.6 syllables per second) specifically affected L2 predictions in both stages of prediction. L2 speakers made slightly later first-stage predictions and made fewer and later second-stage predictions. Thus, while L2 speakers could predict at all speech rates across both stages, these predictions were delayed, reduced, and more variable at higher speech rates.

# CHAPTER 1. INTRODUCTION





Above are two oil paintings by Italian artist Giuseppe Arcimboldo, painted between the years 1570-1590. On the surface, these are simply paintings of a bowl of vegetables or a bowl of fruit. Turn the page upside down and have a second look. Nearly instantly, you will see something completely different. What would otherwise be an upside-down version of the same painting became something more. Here, we see an example of the constructive nature of the brain. That is, the brain constructs what we actually perceive by combining the incoming sensory input with our previous, experience-based world knowledge. What we actually see then is our brains' interpretation concerning the most likely state of the world. We know from our previous experience with the world that faces are quite variable but comprise a specific set of features. Thus, a certain arrangement of features triggers our strong perceptual bias for face perception (e.g., Tsao & Livingstone, 2008). The *context* of our reality has shaped the incoming sensory input. This effect of seeing meaningful stimuli in an otherwise meaningless medium is known as pareidolia and is often explained under the frameworks of predictive coding and the Bayesian brain (for comprehensive discussions, see e.g. Friston, 2010; Lupyan & Clark, 2015). Core to these frameworks is the idea that the brain generates probabilistic models of how the world works. These models are based on our experiences with the world, and by interacting with the world we refine or update these models. When interactions go as planned, our internal model is reinforced. When things do not fit, our internal model can be updated. Thus, under the predictive mind framework, processing has an inherent aspect of prediction.

Under the umbrella of processing is, of course, language processing. Predictive language processing has played a major role in psycholinguistic theory in the past few decades (e.g., Altmann & Mirković, 2009; Dell & Chang, 2014; Huettig, 2025; Pickering & Garrod, 2013), with countless papers showing that prediction is pervasive in language processing. For an example of how language must, minimally, involve some predictive mechanism sensitive to context, we can look at the classic garden-path sentence: "*The horse raced past the barn fell*" (Bever, 1970). The normal context of English sentences (i.e., how sentences we encounter are most commonly structured) influences the initial reading of this sentence and leads readers to difficulties when encountering the word "fell." The typical, predictable structure of the sentence is incorrect, and the unlikely continuation of the actual sentence structure leads to increased difficulty in processing the sentence. In other words, the context (how sentences are normally structured) influenced our processing (the difficulty of reading the sentence) *before* encountering some information (all of the sentence). The brain *has* to be acting predictively for this difficulty to occur. If we simply waited until the sentence was complete to commit to an interpretation, there would be no difficulty. Furthermore, a psycholinguist is much less likely to encounter this processing difficulty upon reading "*The horse raced past the barn fell*." This is the prototypical garden-path sentence that any psycholinguist would have encountered and thus, their previous experience with this sentence has updated their probabilistic model of how this sentence proceeds.

Garden-path sentences are but one of many types of evidence for predictive language processing. For example, reading based eye-tracking studies have shown that predictable words are fixated for less time and are more likely to be skipped (e.g., Rayner et al., 2011). A classic study by Altmann & Kamide (1999) showed that, when hearing sentences like "*The boy will eat the cake*" and looking at a variety of objects, participants will look towards a cake (an edible object) as soon as they hear "*eat.*" Kutas & Hillyard (1984) used electroencephalography (EEG) and showed that semantically predictable words are processed differently to unpredictable words. Similar findings on the effects of predictable input have been replicated countless times by now and are incredibly consistent at finding that language processing is sensitive to the predictability of the incoming stimuli.

The power of predictive language processing can be seen through ChatGPT and other transformer-based large language models (LLMs) which have, for better or worse, changed modern society. LLMs are, quite literally, prediction machines that have been trained on an immense amount of text from sources such as manuals, webpages, articles, books, etc. (analogous to how, in a predictive coding model, our internal representation of the world has been trained by all of the input we have received). This data helps the LLM learn the statistical regularities of language and forms the baseline context for the language model. The goal of the LLM is then to predict the most likely continuation given the context, essentially using its preexisting world knowledge (the dataset on which it is trained) to predict what the most statistically likely series of words would be in response to new input. The LLM approach of using a rich context to predict the most likely series of words given a new input has been sufficient to produce what many consider to be very humanlike output. Even experimentally, ChatGPT has been shown to replicate human behavior in 10 out of the 12 psycholinguistic experiments presented to it by Cai et al. (2025). Interestingly, ChatGPT was susceptible to effects that are thought to require grammatical representations even though it is not trained on syntax, suggesting that mechanisms for predicting the next word can even lead to an emergent syntactic representation. In other words, although it is only designed to predict the next most likely word, these models can form new internal representations (learn!) that they are never explicitly trained on. As another example, Li et al. (2023) trained a GPT model (again, a model that predicts the most likely continuation given the context) on legal move sequences in a board game (Othello). They were then able to show that this model formed internal representations of the board state and derived the rules of the game. While obviously not fully representative of human processing, ChatGPT (and other similar language models) have shown that a purely statistically driven, predictive processing approach can quite well simulate human learning and language processing in some cases. Clark (2013) describes the brain as "a prediction machine" and the humanlike performance of these literal prediction machines offers some interesting support for such an idea.

Given the current evidence, it is at least evident that prediction can play a role in language processing. However, there are still questions regarding the ubiquity of prediction (is everyone always predicting?; e.g., Huettig & Mani, 2016), the cognitive requirements of prediction (does working memory play a role in prediction?; e.g., Ryskin et al., 2020), or the mechanisms through which predictions actually occur (one stage? two stages? production system?  e.g., Huettig, 2015). The findings presented in this thesis offer meaningful insights into these questions.

## 1.1 Prediction in Spoken Language Comprehension

### *1.1.1 Seeing the future*

What is prediction? This question has been surprisingly difficult to answer. As noted by Kuperberg & Jaeger (2016), the meaning of prediction has taken on a variety of meanings between subfields and between researchers. For the following discussion, I will define linguistic prediction similarly to Huettig (2015) and Kuperberg & Jaeger (2016): linguistic prediction is the pre-activation of at least some linguistic feature (e.g., a word, features associated with a word, the role of a word in a sentence, etc.) that is shaped by context (e.g., pre-existing knowledge, current processing environment) before this linguistic feature is encountered. Put somewhat more simply, linguistic prediction is the activation of *some* part of a *thing* before the *thing* is encountered.

As discussed by Kuperberg & Jaeger (2016), one of the major limitations for examining linguistic predictions, and indeed a major historical reason why its role in language processing was previously largely disregarded, is the limitation of *how* to look at predictions. Predictions are, necessarily, about the *future*. For a long time, there was little "direct" evidence of prediction. While there was plenty of evidence that more predictable words were processed faster or differently from less predictable words (in reaction times, e.g., Stanovich & West, 1983; in eye-tracking measures, e.g., Ehrlich & Rayner, 1981; in neural responses, e.g., Kutas & Hillyard, 1980), these methodologies all suffered from the same critical problem. In each of these methodologies, the effect of predictability could, at best, be measured at the *onset* of the predictable item. Thus, it was historically easy to argue that these effects were not due to pre-activation of a word, but instead due to a facilitated integration of a word into its current context.

This changed in 1995 with the seminal study by Tanenhaus et al. This study revitalized a previously used paradigm by Cooper (1974) in which participants have their eyes tracked while listening to auditory input and looking at a display containing multiple objects. A similar method was then used by Altmann & Kamide (1999) to become what is today one of the most influential paradigms in psycholinguistics, the Visual World Paradigm (VWP; for a further comprehensive review of this task, see Huettig, Rommers, et al., 2011). The VWP is an eye-tracking paradigm that consists of two major parts: (1) an auditory stimulus and (2) some images that are visible while listening to the auditory stimulus. Participants hear a sentence that is typically either a task ("*Click on the banana"*) or a general sentence ("*The boy will eat the cake"*). The most common version of the display is a visual array consisting of 4 items, typically situated near the corners of a screen. These objects may either be a target (what the sentence refers to) a competitor (something meant to attract some kind of attention, for example a semantically related object) or a distractor (a neutral, unrelated object). The key feature of this paradigm is that, by tracking where a person looks as a sentence unfolds, we can see evidence for participants looking towards predictable objects *before* encountering the word that refers to that object. For example, when Altmann & Kamide (1999) presented participants with the sentence "*The boy will eat the cake"* and a visual display consisting of a boy, a cake, and several non-edible objects (a ball, a train, and a car), it was clear that participants were looking at the image of the cake as soon as they heard the word "*eat*" (i.e., *before* hearing the word "*cake*"). This paradigm enabled the first, direct evidence of prediction of a target word that was undeniably not due to integratory mechanisms.

Alongside eye-tracking and the VWP, the other most common methodology for looking at spoken language predictive processing is by using EEG and measuring the N400 ERP. The N400 is a neural response that occurs around 400 milliseconds after the onset of a word and the amplitude of the N400 is reduced when encountering a more predictable word. Thus, if participants listen to sentences while being recorded with EEG and the N400 response is measured at the target word onset, predictable target words will elicit a smaller N400 response than unpredictable target words. While evidence for the N400 as a neural response sensitive to predictability dates back to a study by Kutas & Hillyard in 1980, this was still often interpreted as evidence of integration due to the measurement occurring at target onset. However, an influential 2005 study by DeLong et al. showed evidence of a modulated N400 amplitude *before* target onset by measuring the N400 at the onset of an English article (a/an) preceding a predictable target noun. In this study, participants were presented sentences such as "*The day was breezy so the boy went outside to fly a/an kite/airplane.*" They found the expected reduction of the N400 for the predictable target word ("*kite*") compared to the unpredictable ("*airplane*"), but, crucially, also found a reduction of the N400 for the article corresponding to the predictable target ("*a*") compared to the unpredictable target ("*an*"). This study provided the first direct evidence for the N400 being at least partially a measure of predictability, as participants must have predicted a word that started with a consonant for an effect to occur at "*a*". Despite the large influence of this study, there is considerable controversy around this finding. For example, Ito et al. (2017a) failed to replicate this finding, DeLong et al. (2017) responded to this failed replication with critiques on the methodology used, Ito et al. (2017b) responded to these critiques, Yan et al. (2017) reanalyzed the data from both studies and found support for the effect, and Nieuwland et al. (2018) failed to replicate the effect in a 9 laboratory, 334 participant replication.

Luckily, other research has continued to provide evidence that at least one major contributor of the N400 response is an index of pre-activating a word or semantic features associated with a word. For example, careful and intentional manipulation of the predictability and plausibility of a word (e.g., Mantegna et al., 2019; Nieuwland et al., 2019) have been able to show dissociable effects of prediction and integration and neuroimaging data has localized N400 effects to the posterior middle temporal cortex, a region associated with storing and activating mental representations (Lau et al., 2008). Thus, purely integratory accounts of the N400 are not sufficient. Today, one of the most common understandings of the N400 is that it represents the change in the current activation state of semantic memory and that each incoming word alters this state (Kutas & Federmeier, 2011). In this framework, a word that is already predictable given the context has a more active state in semantic memory when it is encountered and thus elicits relatively less change (i.e., a smaller N400) than an unpredictable word. It is noteworthy that some new computational models formalize the N400 as the prediction error produced as the brain derives meaning from a newly encountered word. In such a framework, a predictable word produces a smaller prediction error which manifests as a smaller N400 (e.g., Eddine et al., 2024). However, for the scope of the studies covered here, either interpretation is sufficient to emphasize that modulations in the N400 are indicative of changes in the predictability of an encountered target word.

To summarize, eye-tracking and EEG are the two most common ways to examine predictive processing. While eye-tracking in the visual world paradigm offers a quite "true" measure of prediction, insofar as it allows us to see evidence of looks to a target before it is mentioned, this methodology inherently requires concurrent visual objects which may further constrain or shape the cognitive mechanisms involved in predictive processing. On the other hand, EEG allows us to examine language processing without this extra visual information, but with a measurement that only occurs at target onset. However, this measurement (the N400) is consistently associated with the degree of pre-activation of a word. In the studies included in this dissertation, both of these methodologies were used to offer unique insights into predictive language processing.

*1.1.2 The many flavors of prediction*

Language is full of statistical regularities that our brain can use to aid in predictions (e.g., Saffran & Kirkham, 2018). For example, in any given language certain sound combinations are possible while others are not. The structure of a language (grammar) only allows for certain categories of word in certain relationships to other words. Some words have meanings that are more closely related to other words. Evidence suggests that, not only can each of these cues (phonology, syntax, and semantics) be used for prediction, predictions at each of these levels of representation may be happening simultaneously and interacting (e.g., Heilbron et al., 2022).

The most common ways to manipulate the predictability of a sentence are through either syntactic manipulations or semantic manipulations. In languages with grammatical genders (e.g., Spanish, German, Dutch), these gendered articles can be used to predict a likely upcoming target. For example, in the Dutch sentence "*Kijk naar de afgebeelde fiets"* ("look at the displayed bicycle")*, "de"* is an article that is compatible with "*fiets"*. If participants hear this sentence in the VWP, they will begin to preferentially look at objects compatible with the gender before hearing the actual object (e.g., Huettig & Guerra, 2019).

Due to the lack of grammatical gender in English, by far the most common way prediction is observed in English is through semantic manipulations. For example, the semantic constraints from the aforementioned sentence "*The boy will eat the…"* leads participants to look towards any edible items that are in a concurrent display (e.g., Altmann & Kamide, 1999).

Furthermore, these semantic predictions can proceed through multiple different pathways. For example, a study by Kukona et al. (2011) showed that participants may first make incorrect or incompatible predictions based on semantic constraints. When participants in their study heard the sentence "*Toby arrests the crook,"* they found that participants looked towards an image of a police officer upon hearing "*arrest.*" Thus, the semantic relationship between *arrest* and *police officer* drew gaze behavior even though "*Toby arrests the…*" is unlikely to be followed by *police officer*. Multiple-system models (models that assume that predictions can occur through multiple pathways and not just one singular mechanism) of prediction typically consider this style of prediction to be fast, automatic, uncontrolled, and guided by spreading activation ("the dumb route to prediction" - Huettig, 2015; prediction-by-association - Pickering & Gambi, 2018).

Most models of predictive language processing also consider a second, slower, more intentional pathway to prediction. Studies by Corps et al. have managed to show evidence for two distinct stages of prediction in both L1 (native; 2021) and L2 (non-native; 2022) English speakers. They presented participants with sentences like "*I would like to wear the nice dress/tie*" that were spoken by either a male or a female voice. This sentence was accompanied by a visual array containing images of a dress (stereotypically feminine + wearable), a tie (stereotypically masculine + wearable), a hairdryer (stereotypically feminine) and a drill (stereotypically masculine). Participants first showed semantically constrained looks equally towards wearable objects (dress and tie; stage 1) and then towards objects that were gender compatible (i.e., dress if spoken by a woman and tie if spoken by a man; stage 2). The usage of perspective consistent predictions requires previous knowledge of stereotypically male and female associations to be combined with the semantic constraints of a verb. These second stage predictions are thus often seen to be due to a more deliberate, integratory or combinatory mechanism than the first stage predictions. While these more demanding predictions have been shown to exist, the exact pathway through which they proceed is unclear. For example, Pickering & Gambi (2018) suggest that such predictions are driven by the same mechanisms that are used in language production. Meanwhile, Huettig (2015) suggests that production-, association-, combinatorial-, and simulation-based mechanisms are all required to account for the current findings of predictive language processing and that these mechanisms may all be interacting together simultaneously depending on the task context.

To summarize, predictions are able to formed at multiple levels of representation (e.g., phonologic, syntactic, semantic) and proceed through multiple pathways (e.g., more automatic or more effortful paths). The studies included in this thesis are primarily concerned with semantic predictions and assume that there are at least one primarily associative mechanism and one primarily combinatory mechanism through which predictions may be formed.

### *1.1.3 Are you thinking what I'm thinking?*

Everyone may predict, but not everyone predicts equally. The strength and reliability of predictive language processing can vary based on many factors. Age, for example, seems to play an important role in prediction. While children as young as two have been shown to be able to predict (e.g., Mani & Huettig, 2012), the extent of these predictions is limited. Interestingly, Gambi et al. (2018) found that it is the high-level predictions (e.g., based on semantics) that children are more capable of than the low-level predictions (e.g., based on more perceptual input like phonology), suggesting that children learn to predict (they learn words and concepts, then they start predicting based on conceptual information) instead of children predicting to learn (they predict based on incoming perceptual input, prediction error drives learning). Multiple frameworks of predictive processing suggest that prediction is a driving force for learning (e.g., Dell & Chang, 2014; Friston, 2010). Thus, if children are capable of higher-level predictions (e.g., based on semantics) as early as age 2, but still do not reliably use more perceptually based predictions (e.g., based on phonology) even at age 5, then children seemingly learn without a necessary role for prediction.

On the other end of the age spectrum, results on predictions in older adults are quite varied. From a theoretical point of view, older adults have a longer time to accrue language experience and thus could be expected to show more prediction due to this experience. On the other hand, increasing age comes with various cognitive declines, including specifically declines in working memory and processing speed which have, at least in some cases, been previously linked to predictive processing (e.g., Huettig et al., 2016). Indeed, when controlling for the effects of working memory and processing speed, Huettig et al. (2016) found no effect or even a slightly positive effect of age on prediction. However, a variety of studies have found reduced (e.g., DeLong et al., 2012; Federmeier & Kutas, 2019; Fernandez et al., 2025) or qualitatively different (e.g., Rayner et al., 2006) prediction in older adults. Thus, there seems to be a "sweet spot" for predictive capabilities in early adulthood, where both the predictive mechanisms and the cognitive functioning may be at their peaks.

Similarly, L1 (native) and L2 (non-native) speakers of a language differ in their predictive capabilities. For example, grammatical gender cues are readily usable by L1 speakers for predictive processing but are often not usable for L2 speakers. Even when L2 speakers perform at ceiling level for offline comprehension tasks, they may still be less able to use grammatical gender as a predictive cue than L1 speakers (e.g., Hopp, 2013; Grüter et al., 2012). However, Hopp (2016) found that sufficiently advanced L2 speakers can predict using grammatical gender similarly to L1 speakers. Semantic cues, on the other hand, are readily usable by L2 speakers (e.g., Chambers & Cooke, 2009; Ito et al., 2018). For example, Dijkgraaf et al. (2016) found that Dutch-English bilinguals could use semantic information to predict similarly in both Dutch and English and that their L2 English predictions were no different from monolingual English speakers.

Despite these processing differences, Kaan (2014) suggests that L1 and L2 predictive processing are fundamentally the same. Instead, she suggests that both L1 and L2 predictive processing are moderated by factors such as language exposure (given enough exposure, L2 and L1 should predict similarly; Hopp, 2016), quality of lexical representation (L1 speakers with weaker lexical representations predict less, e.g., Mishra et al., 2012), or cognitive resources (L2 processing is more cognitively demanding than L1 processing, Hopp, 2010). Indeed, neural evidence from Newman et al. (2012) suggests that language exposure (more vs. less) is more relevant than language group (L1 vs. L2) and eye-tracking evidence from Ito et al. (2018) suggests that limiting cognitive resources similarly reduces L1 and L2 predictive behavior. Despite multiple models of predictive processing suggesting an important role of cognitive resource capacities (e.g., Hopp, 2010; Huettig, 2016; Kaan, 2014; McDonald, 2006) only one study (Ito et al., 2018) had previously shown evidence that directly manipulating working memory resources can reduce predictive processing.

Thus, children, older adults, and L2 speakers of a language all show reduced prediction compared to younger L1 adults. As discussed by Ryskin et al. (2020), it is not necessarily clear if these differences arise due to differences in working memory capacity or due to differences in language exposure. Importantly, the way that L2 speakers predict (especially in regards to semantic predictions) should be fundamentally the same as L1 speakers. Thus, L2 language processing may be seen as an extra cognitive demand that we take advantage of throughout these studies.

## 1.2 Motivations and Research Questions

The main source of motivation for the following series of experiments can be summed up by a quote from Huettig & Mani (2016): "… it may be problematic that most experimental evidence for predictive language processing comes from 'prediction-encouraging' experimental set-ups." Two standard parts of the "prediction-encouraging experimental set-up" are slow speech rates and a minimally demanding processing context/environment (i.e., a quiet laboratory setting in which participants hear a simple sentence and select from a small subset of items). Thus, an overarching goal of this series of research was to examine prediction in a less “prediction-encouraging” (i.e., more demanding) context and to see if such manipulations reduced the likelihood or magnitude of predictive processing. By understanding what types of demand influence predictive language processing, we can gain more understanding of the mechanisms that enable predictive processing. Understanding the limits of prediction brings us closer to understanding the cognitive architecture that enables prediction. Indeed, each of the studies in this thesis induce multiple cognitive demands which would not be present in a standard prediction experiment. Participants were faced with concurrent working memory demands, non-native language processing, and speech rates above what are often used in spoken language prediction research.

All three of the studies in this dissertation feature spoken sentences with normal speech rates. While many studies have shown strong evidence that people *can* predict upcoming sentences, these results are almost exclusively in the context of “prediction-encouraging” set-ups. A part of the typical “prediction-encouraging" set-up is the slow speech rate of the spoken sentences (this is discussed further in section 2.2.2). There is an ongoing debate on the role prediction plays in language processing which focuses on whether prediction is necessary and obligatory for language processing or is, instead, an optional process that is sometimes beneficial (e.g., Huettig & Mani, 2016; Ryskin & Nieuwland, 2023). If slow speech rate is required for consistent prediction, then it seems quite unlikely that prediction could be a mandatory facet of language processing. Experiment 3 of this collection takes this one step further and focuses specifically on the effects of speech rate on the stages of prediction to see if different predictions are differentially affected by speech rate.

Two of the three studies presented in this dissertation use a concurrent visuospatial working memory task, specifically a digitized version of the Corsi block tapping task (Corsi, 1970), to increase the demands of prediction. In this task, participants are presented with an array of 9 squares and are indicated a certain number of these squares to remember and later recall in the same order as previously indicated. This task was chosen with two main goals in mind. The first goal was to minimize the recruitment of language in the cognitively demanding task. Thus, the Corsi block tapping task represents a cognitively demanding task that does not require explicit naming or rehearsal of linguistic or phonological information. Secondly, one of the motivations of this research was to test the automaticity of predictive processing. Bargh (1994) described what he called the "four horsemen of automaticity," i.e., four defining characteristics of an automatic process in the brain. These four characteristics are *intention* (are you aware of when this process starts?), *control* (are you able to stop or alter this process once it is initiated?), *awareness* (do you realize this process is happening?), and *efficiency* (does this process require mental effort?). Of particular interest here is the aspect of *efficiency*, or whether predictions are resource demanding. If generating predictions is resource demanding, a reduction in the availability of cognitive resources would be expected to reduce prediction. Thus, reductions in prediction due to a concurrent visuospatial working memory task could potentially be evidence towards prediction being a resource demanding process.

With these motivations in mind, I ask the following questions:

*RQ1. Do L2 speakers predict despite a variety of challenging situations?*

Evidence from Huettig & Guerra (2019) and Ito et al. (2018) suggest that normal speech rates and concurrent working memory demands (respectively) largely predictive gaze behavior. However, real-life language processing is at normal speech rates and often situated in more demanding situations (i.e., not in a quiet controlled laboratory setting). Thus, if prediction is an inherent or ubiquitous aspect of language processing, one should expect to find evidence for prediction at normal speech rates and in more demanding processing contexts. To this end, I ask two more specific research questions.

- RQ1.1 Do L2 speakers need a "slow" speech rate to predict?
- RQ1.2 Do L2 speakers show consistent evidence of prediction with increased demands across stages of prediction and different measurements of prediction?

*RQ2. Does increasing the cognitive demand of a prediction task reduce prediction?*

Whereas RQ1 is more broadly concerned with the ubiquity of prediction and if increases in demand eliminate prediction, RQ2 is more specifically concerned with distinct mechanisms enabling prediction and if they individually modulate predictive processing. Thus, the goal is to first establish if predictive language processing does indeed occur despite increased task demand (RQ1) and, if so, to determine if different types of demands lead to similar reductions in prediction (domain general effects) or if these demands may be more contextual.

- RQ2.1: Does a concurrent visuospatial load reduce predictive gaze behavior?
- RQ2.2: Does a concurrent visuospatial load reduce target pre-activation in the absence of a visual array?
- RQ2.3: Are first and second stage predictions differentially affected by increased speech rate?

# CHAPTER 2. LITERATURE REVIEW

## 2.1 Working Memory and Predictive Processing

Working memory is a limited capacity, short-term system responsible for temporarily holding and manipulating relevant information (e.g., Baddeley, 2003). Working memory has been shown to be an important aspect of both L1 (e.g., Daneman & Merikle, 1996) and L2 (e.g., Szmalec et al., 2012) language processing. Prediction, as a mechanism that serves to aid language comprehension, may be an explanatory factor for the involvement of working memory in language comprehension. As discussed by Ryskin et al. (2020), children, older adults, and L2 speakers are all groups that have both reduced working memory capacities and reduced evidence of prediction. Their reduced working memory capacity is one possible explanation as to why these groups show reduced predictive behavior.

There are, however, multiple conflicting findings on the relationship between working memory and predictive processing. Until quite recently, most studies examining the relationship between working memory and prediction were correlational (for a summary of these correlational studies, see Table 1 at the end of this subsection). That is, studies had participants complete a task that measured prediction and complete at least one secondary task that assessed various measures associated with working memory capacity. Subgroups of participants were created based on these measures (e.g., "high" or "low" working memory participants), or a correlational analysis was performed between the prediction measure and the secondary task measure. The results from eye-tracking studies are quite mixed and there seems to be no or quite limited results regarding working memory and spoken word processing in EEG.

For example, a study by Huettig & Janse (2016) had 105 L1 Dutch participants aged 32-77 listen to syntactically predictable sentences. They then measured the participants performance in (among other, non- working memory related tasks) a nonword repetition task, a backwards digit span task, and a visuospatial working memory task (the Corsi block task) and constructed a "Working Memory" measure based on their performance in these three tasks. They found that the participants' Working Memory score was significantly correlated with their predictive gaze behavior, with higher working memory participants showing significantly more predictive looks to a target than lower working memory participants. They found a similar significant pattern with measures for processing speed, with participants with faster processing speed showing significantly more predictive gaze behavior.

On the other hand, a very similar study was performed by Kukona et al. (2016). They had a group of 77 L1 English participants aged 16-24 listen to semantically predictable sentences. These participants completed an extensive battery consisting of 26 measures after the task, including working memory as measured by the reading span task and visuospatial working memory as measured by the Corsi block task. They found no relationship between either working memory measure and predictive gaze behavior and suggest that processing speed and word knowledge differences are two key mechanisms enabling predictions.

More recent studies examining working memory and prediction continue the confusion. A large-scale (N = 487, ages 9-90) study in Hungarian by Hintz et al. (2024) used a reaction time based visual world task and found little role for working memory as measured by the forward digit span and backwards digit span tasks. A Turkish children visual world study by Özkan et al. (2022) only found one specific type of working memory task, a sentence repetition task, to be correlated with prediction. Two visual world studies (Li & Qu, 2024; Angulo-Chavira et al., 2022) measured the performance in different working memory tasks and found evidence for a role of working in the onset of semantic predictions. Specifically, both studies found that higher working memory participants made semantic predictions (but not their phonological predictions) earlier than lower working participants, implying a role for working in the speed of semantic activation. Thus, the current visual world evidence paints a very unclear picture of the role of working memory in predictive processing. Given the large variety of age ranges and tasks measured in each experiment, no two experiments are quite comparable. Even when multiple studies use the working memory measure (e.g., reading span), results are mixed.

Most of the published studies investigating the links between working memory and prediction using EEG and the N400 seem to be reading studies (i.e., not spoken word comprehension). However, the elicitation of the N400 is, generally speaking, considered to be modality independent. That is, spoken language, written language, and signed language all elicit quite similar N400 responses (e.g., Kutas et al., 1987). However, reading is its own complex skill that involves several steps that are not present in spoken word comprehension (e.g., letter and word recognition, phoneme to grapheme conversion). Furthermore, working memory abilities have been found to uniquely contribute to both foundational reading skills and reading comprehension (Peng et al., 2018). Thus, specifically when looking at the role of working memory, we may not necessarily expect the same pattern of effects in the N400 between reading sentences and listening to sentences.

That being said, at least three studies have used EEG to examine the effect of working memory on predictive processing while reading. Each of these studies separated their participants into high and low working memory groups based on their performance in the reading span task. Otten & Van Berkum (2009) and Ness & Meltzer-Asscher (2018) found similar results. Namely, they both found that the N400 response in both the high and low working memory capacity groups was similar and that differences only showed in a later, different neural response. Thus, both authors suggested that working memory does not influence the pre-activation of a word, but rather influences the brains response to this pre-activated word. However, a study by Ding et al. (2023) found a differential N400 response specifically to moderately predictable nouns, with moderately predictable nouns eliciting a smaller N400 (i.e., are more semantically pre-activated) in the higher working memory group than the lower working memory group. This suggests that higher working memory participants either pre-activated more information or were more able to integrate a moderately predictable target than lower working memory participants.

A summary of the previously mentioned studies can be seen in Table 1. Unfortunately, the studies varied in far too many other ways to accurately capture in a table. For example, these studies were conducted in Dutch, English, Hungarian, Chinese, and Spanish. Some looked at predictions enabled by syntax and others by semantics. Özkan et al. (2022) looked at children aged 4-8, Hintz et al. (2024) looked at people aged 9-90, and multiple studies looked at university aged groups. The main finding from these combined results seems to be that, when working memory plays a role in predictive processing, it is likely to be context dependent. For example, Hintz et al. (2024) used a visual array with only two items, which may have minimized the involvement of working memory. The findings from Ding et al. (2023) may suggest that working memory is more useful at moderate levels of predictability, as opposed to the highly predictable or unpredictable sentences that are normally looked at in prediction research. Thus, while there seems to be circumstances in which working memory can be linked to predictive processing, these circumstances remain unclear. The large variance between studies and their findings makes it difficult to draw meaningful conclusions.

Table 1. Summary of correlational findings between predictive processing and working memory.

| Study | Methodology | Working Memory Measures | Significant Effects |
|---|---|---|---|
| Huettig & Janse 2016 | Visual World / Listening | Nonword Repetition,<br>Backwards Digit Span,<br>Corsi Blocks | Yes<br>(combined) |
| Kukona et al., 2016 | Visual World / Listening | Reading span,<br>Corsi Blocks | No<br>No |
| Özkan et al., 2022 | Visual World / Listening | Sentence Repetition,<br>Forward Digit Span,<br>Backward Digit Span | Yes<br>No<br>No |
| Hintz et al., 2024 | Visual World / Listening | Forward Digit Span,<br>Backward Digit Span | No<br>No |
| Li & Qu, 2024 | Visual World / Listening | Reading Span | Yes |
| Angulo-Chavira et al., 2022 | Visual World / Listening | Backwards Digit Span,<br>Corsi Blocks | Yes<br>No |
| Otten & Van Berkum, 2009 | EEG / Reading | Reading Span | No |
| Ness & Meltzer-Asscher, 2018 | EEG / Reading | Reading Span | No |
| Ding et al., 2023 | EEG / Reading | Reading Span | Yes |

## 2.2 Direct Manipulations of Cognitive Demand in Spoken Language Prediction

As outlined above, attempts to indirectly link working memory capacity or general working memory abilities to predictive processing have been, in the best case, inconclusive. Thus, the focus of the research has been on directly manipulating the cognitive demands of language processing during a prediction task. The research presented in this dissertation has focused on two manipulations: the effect of a concurrent visuospatial demand (as a domain general working memory manipulation) and the effect of speech rate. Ongoing research also looks at the effects of a concurrent phonological demand. The following sections will discuss each of these demands, including the previous findings related to these manipulations (insofar as there are previous findings) and how these manipulations were used to address their role in predictive language processing.

### *2.2.1 Directly limiting cognitive resources*

One of the earlier points of evidence for a possible involvement of visuospatial working memory in predictive processing comes from a study by Heyman et al. (2015). However, this study involved reading instead of spoken language processing. Their study looked at the effect of a visuospatial working memory task on semantic priming, which is most commonly explained as an effect that results due to automatic spreading activation (e.g., Collins & Loftus, 1975). This same process of automatic spreading activation is also assumed to play an important role in prediction (e.g., Kukona et al., 2011, Pickering & Gambi 2018). In the study by Heyman (2015), participants were first shown either a simple (low demand) or a complex (high demand) dot matrix that they had to remember. Then, they were presented with a sequence of a prime word (e.g., thunder), a blank screen, and a target word (e.g., lightning) or a nonword. When the target word appeared, participants had to judge whether this was an existing word or not. This sequence was presented five times and after the fifth time, a blank matrix appeared and participants had to reproduce the previously seen dot matrix. The prime and target could have three different types of relationships: a forward association, a backwards association, or a symmetrical association. Priming in forward associated word pairs is thought to be due to prospective processes (which they refer to as "expectancy generation"), backwards associations with retrospective matching processes, and symmetrical associations with both/either. Interestingly, they found that the priming effect for forward associations (and thus the prospective processes most similar to prediction) were not only reduced but completely eliminated with a high concurrent visuospatial working memory demand. The findings of this study suggest that at least some part of the process of activating a word, the spread of activation of that word to a new word, and the resulting activation of that new word is resource dependent.

Some tentative evidence for an effect of cognitive demand on predictive gaze behavior can be seen in a study by Liu et al. (2022). In their study, L1 Dutch participants saw a visual world array while either consecutively interpreting (listen, then interpret) or simultaneously interpreting (interpret while listening) Dutch sentences into English. An unintended finding of their study was that, during the more demanding task of simultaneously interpreting, 25% of their participants did not move their eyes at all. Since they did not move their eyes, there was no possibility of predictive gaze behavior. While it remains unclear if this manipulation affected prediction itself, the study offers evidence that cognitively demanding situations can affect gaze behavior and thus our ability to measure prediction using eye-tracking.

The first study to directly and intentionally manipulate cognitive resource availability in a visual world eye-tracking task was by Ito et al. (2018). In their study, they reduced the availability of cognitive resources in a visual world task by having participants remember a list of five words before completing a visual world trial and recalling this list of words after the visual world trial. They found that when either L1 or L2 English participants had to remember the list of words, they showed significantly less predictive gaze behavior. They suggest that this is evidence that cognitive resources are required for predictive gaze behavior and that predictive gaze behavior is stronger when more resources are available. This finding has also been taken as evidence for a domain-general effect of increased cognitive demand leading to reduced predictive processing.

The domain-general interpretations of this study were an important consideration for the usage of a visuospatial working memory task in the studies included in this dissertation. A verbal working memory task such as the one used by Ito et al. (2018) involves both the semantic network and the language production system, two systems that are assumed to be major contributing mechanisms of predictive processing (e.g., Huettig, 2015; Pickering & Gambi, 2018). As previously mentioned, in the Corsi block tapping tasks, participants are shown an array of squares and are indicated to remember the order and location of some number of them. Using the Corsi block tapping task minimizes the involvement of both the semantic network and the language production system, as there are no specific words associated with the task and the task does not lend itself to verbal rehearsal strategies. Thus, disruptions due to a concurrent visuospatial working memory task would be much stronger evidence for a domain general involvement of working memory resources than evidence from a task that activates the specific mechanisms involved in predictive processing.

Two of the three studies featured in this dissertation focus on the effects of a concurrent visuospatial working memory demand on predictive processing. Allison et al. (2025; Study 1) was the first study to establish that a concurrent visuospatial working memory task reduces predictive gaze behavior. The initial findings of this study were presented at a conference in 2023 and have since inspired multiple others to investigate this effect (Fernandez et al., 2025; Ito, 2025; Liu et al., 2025; Nota et al., 2024; Wang et al., 2025). These studies (except for Liu et al., 2025 who used a dot matrix instead) used the same visuospatial working memory manipulation: a digitized version of the Corsi block tapping task. In this task, participants need to remember the order and location of a specified series of a subset of squares from a larger set (e.g., remember 4 from 9). The task demand increases as the number of squares to be remembered increases. The studies mentioned above have required participants to remember the order and location of between 2 and 5 squares to increase (visuospatial) working memory demand.

The effect of a concurrent visuospatial working memory task on prediction was first established in the context of the visual world paradigm (Allison et al., 2025; Study 1). When L2 English participants perform a concurrent visuospatial working memory task while completing a visual world trial, they exhibit significantly less predictive gaze behavior. This effect was replicated by Wang et al. (2025) and showed the same pattern, only this time with L1 English speakers. In both studies, participants could hear semantically predictable or unpredictable sentences in either trials with a concurrent visuospatial working memory demand or trials without. In the concurrent working memory demand trials, participants were first provided a series of squares to remember, then listened to a predictable or unpredictable sentence, then clicked the picture best matching the sentence, and finally clicked the squares in the order they were indicated at the beginning of the trial. Although both studies found the same pattern of results (i.e., participants showed significantly less predictive gaze behavior during the working memory blocks), these could be (and were) interpreted in two different ways: either as an effect specific to the coordination of auditory and visual input (i.e., an effect contingent on the visual array in the visual world paradigm; Allison et al., 2025) or as a domain-general reduction in available working memory resources leading to a reduction in predictive gaze behavior (Wang et al., 2025). The second study in this collection was designed to help address which of these two conflicting interpretations was more likely.

As a recap, all of the current evidence is in agreement that concurrent working memory demands reduce predictive gaze behavior in visual world eye-tracking, with both visuospatial and phonological demands having been shown to reduce predictive gaze behavior. However, the interpretations of these effects differ: Allison et al. (2025, Study 1 here) suggests that the effects of visuospatial working memory are contingent on language/vision interactions, Ito et al. (2018) and Wang et al. (2025) suggest that the effects of both phonological and visuospatial working memory are domain-general effects of working memory reduction, and Liu et al. (2025) suggest that visuospatial working memory demands and phonological working memory demands affect two distinct mechanisms (and are presumably not domain-general effects, although they do not explicitly mention these effects in this context). Thus, there are still open questions to be answered. Are reductions in predictive gaze behavior due to visuospatial and phonological working memory demands "the same?" In other words, is prediction dependent on cognitive resource availability and do these two manipulations reduce these resources? Are visuospatial and phonological working memory demands distinct and thus affecting predictive gaze behavior through different mechanisms? Are these disruptions in the activation of a target word or features or rather disruptions in gaze behavior?

### *2.2.2 Speech rate*

I've heard speech rate described as the "dirty secret" of prediction research. For much of the research using the visual world paradigm, the default speech rate is one that most typical listeners would consider abnormally, or at the very least noticeably, slow. As noted by Ito & Knoeferle (2022), it is "common to use a relatively slow speech rate" and, as noted by Huettig & Guerra (2019), "experimenters tend to present well-articulated and fairly slow speech rates to participants." Slow speech rate presentations have been the default for predictive language processing, with the idea that people have more time to predict if the sentences are slower. In fact, this type of presentation is such a default that, until recently, speech rate was essentially an afterthought.

For example, Fernandez et al., (2020) reviewed 45 visual world studies published in high-impact psycholinguistic journals between the years 2008 and 2018. Of these 45 studies, only three studies actually reported their speech rates, with rates of 1.3, 2.6 and 3.1 syllables per second. Even the fastest of these reported rates is relatively slow. For context on speed, Wang (2021) analyzed 77 hours and 58 minutes of news monologues, interviews, conversations, and academic lectures in the UK and found an average speech rate of 4.4 syllables per second. Given the sources from these recordings (especially news monologues and academic lectures), it's not unimaginable that many of these utterances are also perhaps on the slower, more annunciated end of normal speech.

Of the remaining 42 studies, 10 reported a "normal" speech rate. Though, as also noted by Fernandez et al. (2020), the definition of what normal is vague, and they report estimates for "normal" speech rates as low as 2.50 and as high as 8.0 syllables per second in the literature. Indeed, following a chain of citations in some recent papers leads to something like Fónagy & Magdics (1960) who report an average Hungarian speech rate of 11.35 sounds per second. This was then interpreted by Wilshire (1999) as being between 2.5 and 4.56 syllables per second, and this interpretation has been cited multiple times. Alternatively, some citations led back to a study by Moos and Trouvain (2007), which simply stated that normal rates are between 3 and 5 syllables per second with no further citation or empirical data. One of the most cited studies on speech rate is from Tauroza & Allison (1990), who analyzed British English conversations, academic lectures, interviews, and radio monologues and found an average speech rate of 240 syllables per minute (or roughly 4 syllables per second). The aforementioned study by Wang (2021) was designed as a replication of this and found a slightly faster speech rate today (265 syllables per minute / roughly 4.4 syllables per second).

The remaining 32 studies surveyed by Fernandez et al. (2020) reported nothing at all about speech rate. However, it seems that speech rate is relatively understudied as a whole and not just taken for granted in the context of visual world eye-tracking research. Given how influential information processing theories are on today's models and ideas of the brain and cognition, this lack of speech rate research seems particularly surprising. Saying the same words more or less quickly alters the information density of the statement; more words in less time means that more information per unit of time is being conveyed. Furthermore, research by Jaeger & Levy (2006) and Jaeger (2010) has shown that speakers are implicitly sensitive to information density and that information density can predict the likelihood of speakers producing certain structures.

The existing research on the effect of speech rate on predictive processing seems to be largely limited to the studies by Huettig & Guerra (2019) and Fernandez et al. (2020), both of which found important roles for speech rate. The study by Huettig & Guerra (2019) manipulated speech rates in Dutch sentences but did not report the actual speed of their items in syllables per second. However, they reported an example sentence "*Kijk naar de afgebeelde fiets*" and the mean utterance duration of their normal speed sentences (1815.58 ms), which comes out to around 4.4 syllables per second (quite well in line with what the literature on English speech rates reports as normal). The same sentence was also recorded in a slow condition, which they report had a mean utterance duration of 4170.05 ms and would thus correspond to approximately 1.9 syllables per second. In their experiment 1, they found that L1 Dutch participants predicted as early as possible in both speech rate conditions when given a four second preview of the visual display. With only a one second preview of the visual display (in their experiment 2), they found evidence for prediction only in the slow speech rate condition. Finally, in their experiment 3, they showed that if participants are explicitly asked to predict, that there are small prediction effects in the normal speech condition. Thus, L1 Dutch participants were less able to use gender as a predictive cue at fast speech rates when there was little time to preview a visual array, even if they were explicitly asked to predict.

The study by Fernandez et al. (2020) looked at the effect of speech rates of 3.5, 4.5, 5.5, and 6.0 syllables per second on younger (18-24) and older (40-75) adults (Experiment 1) and in L1 and

L2 speakers (Experiment 2). Critically, they found that different groups have different optimal speech rates for prediction. Specifically, they suggest that younger L1 speakers have a faster optimal speech rate (5.5 syllables per second) than older L1 speakers (4.5 syllables per second) or L2 younger L2 speakers (also 4.5 syllables per second). Furthermore, they found *weaker* predictive behavior for the younger L1 English group in the 3.5 syllables per second condition. Thus, much of the research on prediction has used auditory stimuli with slow speech rates which may not be representative of true predictive behavior. Given the findings of how our brains are attuned to the regularities of the world, it is not particularly surprising that our brains may be most attuned to the speech rate which we are most commonly exposed to.

Taken together, the current evidence is that speech rate is an underreported aspect of predictive language processing that differentially affects language users. Thus, one of the major goals of this research is to establish the necessity of these abnormally slow speech rates that are often used in prediction research. The research presented here addresses the question of if L2 speakers show robust predictive language processing at normal speech rates. I address this in three specific questions: (Experiment 1) Do L2 speakers show robust predictive gaze behavior at normal speech rates, (Experiment 2) Do L2 speakers show robust pre-activation of target words as evidenced by the N400 at normal speech rates, and (Experiment 3) Do L2 speakers show robust predictive gaze behavior at multiple stages of prediction at a range of speech rates.

# CHAPTER 3. INCLUDED ARTICLES

LANGUAGE, COGNITION AND NEUROSCIENCE
2025, VOL. 40, NO. 9, 1252–1261
https://doi.org/10.1080/23273798.2025.2522272



REGULAR ARTICLE



# Visuospatial working memory load reduces semantic prediction in the visual world

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

[a]Center for Cognitive Science, University of Kaiserslautern-Landau, Kaiserslautern, Germany; [b]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; [c]Faculty of Psychology, University of Lisbon, Lisbon, Portugal

**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.

**ARTICLE HISTORY**
Received 6 October 2024
Accepted 26 May 2025

**KEYWORDS**
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 maximise 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 characterisation 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

**CONTACT** Christopher Allison allison@rptu.de Center for Cognitive Science, University of Kaiserslautern-Landau, Gottlieb-Daimler-Straße 47, 67663 Kaiserslautern, Germany

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, et al., 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 et al., 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, et al. (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, et al. (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 hypothesised 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, et al. (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, et al. (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 300 × 300 pixel greyscale drawings taken from the Multi-Pic database (Duñabeitia et al., 2018).

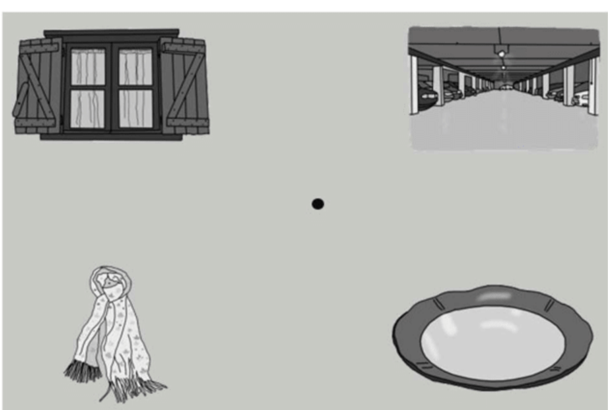


**Figure 1.** Example of a visual array that accompanied the sentence pair The waiter/runner brings/remembers the plate.

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 normalised 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).

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 500 ms 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 minimise any linguistic interference during the Visual World trials.

**Table 1.** Standardized format of the sentences, the normalised 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 |

**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 colour 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.

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 randomised, 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 $1024 \times 768$ 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 2000ms 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 1903ms 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 $300 \times 300$ pixel

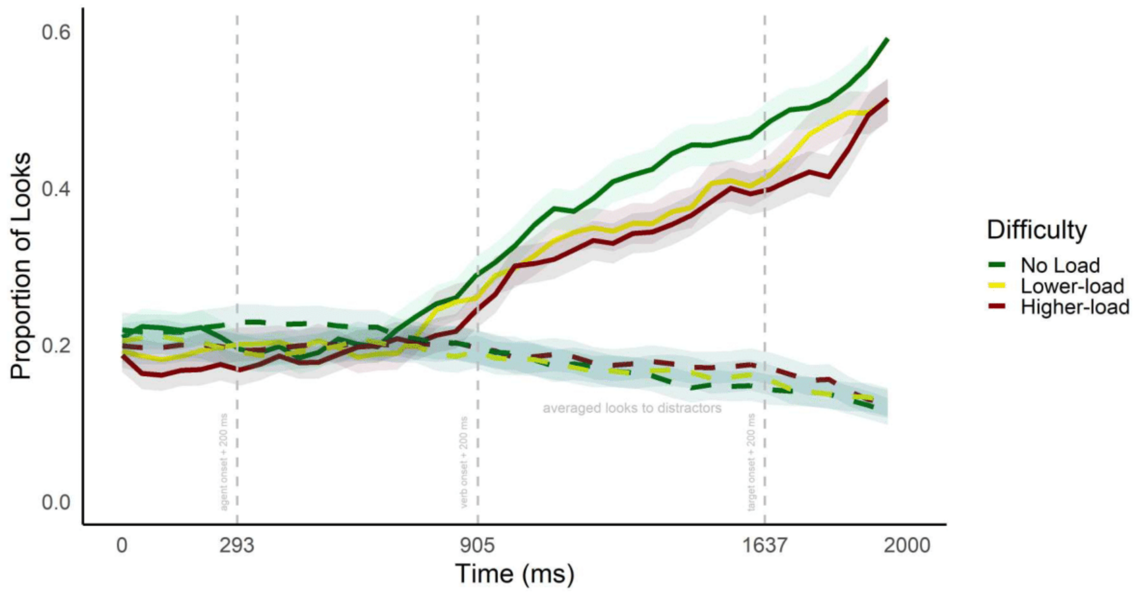



**Figure 3.** Fixation proportion data to the predictable target in the three cognitive load conditions. Looks in the no-load, lower-load, and higher-load condition are represented by green, yellow, and red lines, respectively. The dotted lines represent an averaged look to the distractors. The bands 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.

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.

**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 |

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), visualisation (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.

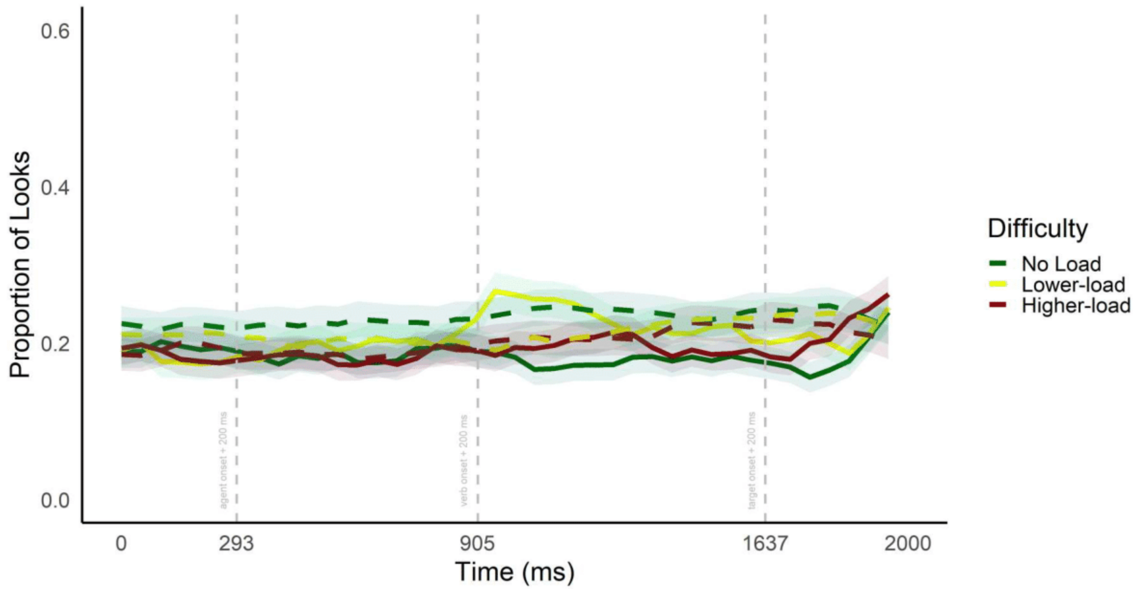



**Figure 4.** Fixation proportion data to the unpredictable target in the three cognitive load conditions. Looks in the no-load, lower-load, and higher-load condition are represented by green, yellow, and red lines, respectively. The dotted lines represent an averaged look to the distractors. The bands 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.

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 ($\beta = -0.43$, $SE = 0.07$, $t = -6.3$) 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.

## 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. Fernandez et al., 2020; Huettig & Guerra, 2019). 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, et al. (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 statement

No potential conflict of interest was reported by the author(s).

## Data availability

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

## 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*, 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., & Chan, A. (2021). Influence of syntactic complexity on second language prediction. In E. Kaan, & T. Grüter (Eds.), *Prediction in Second Language Processing and Learning* (pp. 69–89). 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*. McGill University.

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*, *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*, 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*. NAACL '01: Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies.

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* (pp. 61–72). Indiana University Press.

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.

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*, 92–101. https://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/s0140525(12001495

Porretta, V., Kyröläinen, A. J., van Rij, J., & Järvikivi, J. (2016). VWPre: tools for preprocessing visual world data. https://cran.r-project.org/package=VWPre

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation 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

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

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

Wagner, A., Pals, C., de Blecourt, C. M., Sarampalis, A., & Başkent, D. (2016). Does signal degradation affect top–down processing of speech?. In P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, & C. Lanting (Eds.), *Physiology, psychoacoustics and cognition in normal and impaired hearing* (pp. 297–306). Springer.

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

# Are Visuospatial Working Memory Influences on Prediction in Speech Contingent on the Co-presence of a Relevant Visual World? An ERP study

Christopher Allison[a]*, Falk Huettig[a,b, c], Ann-Kathrin Beck[a], Shanley Allen[a], and Thomas Lachmann[a,d]

[a]*Center for Cognitive Science, RPTU University Kaiserslautern-Landau, Kaiserslautern, Germany;* [b]*Psychology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands;* [c]*Faculty of Psychology, University of Lisbon, Lisbon, Portugal;* [d]*Nebrija Research Center in Cognition, Nebrija University, Madrid, Spain*

Corresponding author:

Christopher Allison

**allison@rptu.de**

## Are Visuospatial Working Memory Influences on Prediction in Speech Contingent on the Co-presence of a Relevant Visual World? An ERP study

A concurrent visuospatial working memory (vWM) task has been shown to reduce predictive gaze behaviour in several visual world eye-tracking experiments. This reduction could be due to general aspects of prediction or an effect contingent on the additional demand of visually co-present objects. In an N400 event-related potentials study, L2 English participants heard the same sentences and completed the same vWM task from a study which showed a reduction in predictive gaze behaviour, but without visually co-present objects. We observed robust effects of predictability but evidence against an influence of the vWM task. These results suggest that vWM does not interfere with the pre-activation of a target word without a co-present visual array. We conclude that previously observed effects of vWM on prediction are instead likely specific to gaze behaviour and situations in which spoken language refers to a co-present relevant visual context.

Keywords: prediction; working memory; N400, language processing

### Introduction

Prediction has become a major framework for understanding the human mind and brain (see Clark 2015; Hohwy, 2013; Huettig, 2025; for comprehensive discussions). Robust evidence for this can be seen, for example, in the study of predictive language processing (for recent reviews, see Pickering & Gambi, 2018; Ryskin & Nieuwland, 2023). These studies have found that populations vary in their predictive capability and that a variety of underlying cognitive mechanisms, including working memory, may support predictive processing (e.g., Huettig et al., 2016; Kukona et al., 2016; Hintz et al., 2024). As discussed by Ryskin et al. (2020), reductions in prediction in children, older adults, and second language speakers (L2) have implicated the involvement of executive functioning during forward-looking behaviour. One core aspect of executive functioning is working memory (Diamond, 2013), but its role in prediction is particularly unclear. While some studies have provided evidence for the involvement of working memory in predictive processing (e.g., Huettig et al., 2016), others have not (e.g., Kukona et al., 2016). A deeper understanding of the factors that facilitate and hinder predictive processing is an important step to determining the role of working memory in predictive language processing and better understanding the underlying mechanisms thereof.

Two methodologies have provided the bulk of the evidence for predictive processing in spoken language comprehension: (1) visual world eye-tracking (Cooper, 1974; Tanenhaus et al., 1995) and, in particular, preferential gaze behaviour towards a target object before the corresponding target word is encountered (for an overview see Huettig, Rommers, et al., 2011) and (2) event related potentials (ERPs) and, in particular, the N400 amplitude (Kutas & Hillyard, 1980) in response to a target word (for an overview, see Kutas & Federmeier, 2011). While both of these methods offer insight into the precise time course of language comprehension and findings between them are typically comparable, there are crucial differences between the two methods. Visual world eye-tracking can provide strong evidence for predictive processing by measuring anticipatory eye gaze *before* a listener hears a target word, but it is crucially dependent on a co-present visual array. This interfacing between language and vision may shape the cognitive requirements of prediction (e.g., Huettig, Olivers, et al., 2011). ERP evidence, on the other hand, can examine prediction in the absence of visually co-present objects. However, the N400 is typically measured *after* a listener hears a target word and, while it has been shown to reflect predictive processing, it has also been shown to additionally reflect the ease of integrating a target word in its current context (Baggio & Hagoort, 2011; Nieuwland et al., 2020; Mantegna et al., 2019).

### *Visuospatial working memory and prediction*

There is a relatively limited body of research directly testing the effects of visuospatial working memory (vWM) on predictive spoken language processing. Most relevant are several recent visual world eye-tracking studies that have shown that a concurrent non-language working memory manipulation (specifically, a visuospatial manipulation) can reduce predictive gaze behaviour (Allison et al., 2025; Liu et al., 2025; Wang et al., 2025).

The first evidence establishing a direct link between vWM capacity and predictive gaze behaviour was shown in a study by Allison et al. (2025). This study used a 2x3 within-subjects design, where L2 English participants saw either predictable or unpredictable sentences with either no concurrent vWM demand, a lower vWM demand, or a higher vWM demand. The vWM task was a modified Corsi block-tapping task (Corsi, 1972) requiring participants to remember the location of 0, 2, or 4 squares indicated from an array of 9. If participants had to remember a sequence of squares at the start of a trial, they had to recall the sequence in the same order and location after completing a visual world trial. Participants heard 'normal' speed (for a discussion on speech rate in the visual world paradigm, see Fernandez et al., 2020), semantically predictable or unpredictable sentences (e.g., *The waiter brings the plate* or *The runner remembers the plate*). Especially when completing the more demanding vWM manipulation of remembering the location of 4 squares, L2 participants showed significantly reduced predictive gaze behaviour.

A methodologically quite similar study was conducted by Wang et al. (2025). This study used a 2x2 within-subjects design, where L1 English participants saw either predictable or unpredictable sentences with or without a concurrent vWM demand.

They also used a similar modified Corsi task, requiring participants to remember and later recall the location of 5 squares indicated from an array of 12. Similar to Allison et al. (2025), they found that participants performing a simultaneous vWM task showed reduced predictive gaze behaviour. Additionally, they found that increases in vWM did not affect the onset of predictive gaze behaviour. So, despite an overall reduction in predictive gaze behaviour, participants showed evidence of prediction starting at similar times.

The final study we know of looking at the role of vWM in predictive spoken language processing was conducted by Liu et al. (2025). Specifically, they compared the effects of increased visuospatial working memory demand to the effects of increased vWM demand. In this study, native Mandarin Chinese speakers heard sentences that could be predictable based on the tonal alignment of a numeral and a classifier in Chinese. Half of their participants completed control trials and concurrent vWM trials and the other half of their participants completed control trials and concurrent verbal working memory trials. Again, they found that participants performing a concurrent vWM task showed reduced predictive gaze behaviour. Similarly to Wang et al. (2025), they found that the onset of predictions was similar between control trials and trials with a concurrent vWM despite the overall reduction in predictive gaze behaviour. Interestingly, they found that the onset of prediction in the verbal working memory manipulation trials was much later than for the control or vWM trials, suggesting that vWM and verbal working memory manipulations may be affecting two distinct mechanisms.

In summary, three studies that we know of have examined the role of vWM on predictive processing through visual world eye-tracking. Each of these studies have found a significant reduction in predictive gaze behaviour when there is an extra vWM task. However, it remains unclear if these reductions are contingent on the presence of a concurrent visual array. For example, vWM could be disrupting mechanisms involved in coordinating incoming language with visually copresent objects (connecting language with a current visual environment) or there could be a shared, limited pool of available cognitive resources from which prediction draws and with which vWM interferes. To the best of our knowledge, there have been no EEG studies on the effect of vWM on spoken language predictive processing and, accordingly, no studies that have examined the role of vWM in the absence of concurrent visual objects.

### *Current Study*

The goal of this study was to examine prediction in spoken language processing after decoupling the auditory sentence processing and the visuospatial processing inherent to visual world eye-tracking studies. To achieve this, we used the N400 ERP component to examine whether an increase in vWM demand in the absence of a visually co-present environment interferes with the semantic pre-activation of a target word. We used the same auditory stimuli and the same vWM manipulation previously shown to reduce predictive gaze behaviour in a visual world eye-tracking study (Allison et al., 2025). Additionally, a similar population of diverse L2 English speakers as in Allison et al. (2025) performed the task. This population has already shown sensitivity to the manipulation, gaze behaviour in L1 (Wang et al., 2025) and L2 (Allison et al., 2025) speakers has been shown to be similarly impacted by a concurrent vWM demand and, generally, we expect L1 and L2 English speakers to engage similar predictive mechanisms when processing simple, semantically predictable sentences (e.g., Kaan, 2014). Furthermore, by having participants listen in an L2, we hope to increase the inherent working memory demand of the task and thus allow more of a role for working memory capacity (e.g., Hintz, 2024).

## Materials and Method

### *Participants*

Thirty-three participants from the University of Kaiserslautern-Landau (RPTU; 10 females) between ages 18 and 30 ($M$ = 25.18, $SD$ = 2.57) participated in this experiment. All participants were proficient L2 English speakers (mean age of English acquisition = 7.84, $SD$ = 4.21) with an average LexTALE proficiency score of 74.66 ($SD$ = 11.23). The native languages of the participants were as follows: German (7), Persian (4), Turkish (4), Tamil (3), Urdu (3), Hindi (2), Kannada (2), Malayalam (2), Spanish (2), Arabic (1), Balochi (1), Gujarati (1) and Marathi (1). All participants reported being right-handed and having normal or corrected-to-normal vision and hearing. All participants were recruited from within or around the University of Kaiserslautern-Landau (RPTU) at the Kaiserslautern campus, provided written informed consent, and were compensated either with either money or participant hours. The experiment was approved by the ethics board of the RPTU University Kaiserslautern-Landau social sciences department.

### *Materials*

We used the same stimuli as Allison et al. (2025) and Fernandez et al. (2024). Specifically, we used 80 sentence pairs and their corresponding visual arrays. These sentences were recorded by a male native speaker of Scottish English. In each sentence pair, the predictability of the target word was manipulated by changing the agent and the verb of the sentence. Each sentence pair consisted of a predictable sentence (e.g., *The waiter brings the* ***plate***) and an unpredictable sentence (e.g., *The runner remembers the* ***plate***). Thus, both sentences in the pair share a target word (***plate***) that may be semantically predictable or unpredictable. Each sentence consisted of five words (i.e., The Agent Verb The Target), was normalised in duration to 1903.07 ms, and had an average speech rate of 3.47 syllables per second ($SD$ = 0.77).

Each sentence pair had a corresponding visual array consisting of four objects that appeared after participants heard the sentence (Figure 1). Arrays corresponding to predictable sentences contained one target object that was semantically predictable (e.g., plate), one distractor object that was plausible but not predictable (e.g., scarf), and two distractor objects that were neither plausible nor predictable (e.g., garage, window). The unpredictable sentence in the pair had the same corresponding array and thus had one target object that was not predictable (e.g., plate) and three objects that were unpredictable distractors (e.g., scarf, garage, window). All objects were 300x300 pixel greyscale drawings taken from the MultiPic database (Duñabeitia et al., 2018).

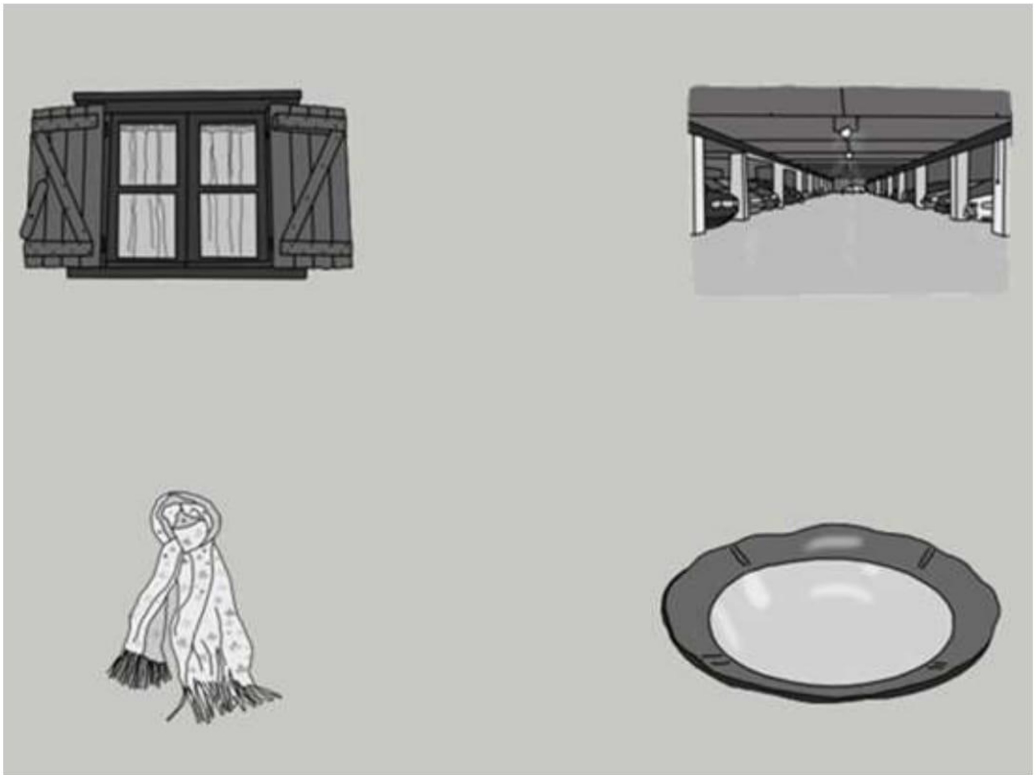


Figure 1. Example of a visual array containing a target and three distractors. This array corresponds to the predictable sentence *The waiter brings the plate* and the unpredictable sentence *The runner remembers the plate*.

Visuospatial working memory demand was manipulated using a modified version of the Corsi block tapping task (Corsi, 1972). Each trial began with the presentation of 9 randomly located white squares (Figure 2). In the no load condition, participants saw these 9 squares for 2000 ms. After 2000 ms, participants saw a screen containing only a fixation dot while they listened to the sentence. In the load condition, a total of four squares were indicated by a color change from white to dark grey. One square was indicated at a time for 500 ms with a 500 ms interval between indications.

Figure 2. Example of the visuospatial working memory task. Participants were presented with a random array of nine squares such as this at the start of each trial. Randomly and in half of the trials, four of these squares would sequentially change to dark grey as an indicator to remember their order and location.

Each participant saw both the predictable and unpredictable sentence of the sentence pair, leading to a total of 160 trials. For each participant, these sentences were randomly assigned to either the load condition or the no load condition. This resulted in 4 conditions consisting of 40 trials each: predictable – no load, unpredictable – no load, predictable – load, and unpredictable – load.

***Procedure***

Participants were individually tested in an electrically shielded room. Each participant completed two practice trials followed by 160 experimental trials. Participants could take breaks between each trial. Each trial started by the participant looking at the fixation dot and pressing the space bar on a keyboard to initiate the trial. Upon initiating the trial, participants saw 9 white squares on a light grey background (Figure 2).

Randomly in half the trials, some of these squares would be indicated and the participant had to remember the order and location of these squares. In the other half of the trials, participants proceeded from this screen after 2000 ms. After presentation of the squares, the screen was replaced with a screen containing only a fixation dot while either a predictable or unpredictable sentence was played through a set of speakers. After the presentation of the sentence, four pictures appeared on the screen (Figure 1) and participants had to select the picture that best matched the sentence they just heard. If any of the squares were indicated at the start of the trial (i.e., if participants were completing a trial in the load condition), the 9 white squares reappeared and participants had to click the squares in the order and location that was previously indicated. If participants were completing a trial in the no-load condition, the trial ended after participants selected the most corresponding picture. The experimental session lasted approximately one hour.

***EEG Recording***

Brain activity was recorded with 26 Ag/AgCl cap-mounted electrodes (EasyCap GmbH; Gilching, Germany) with the BrainVision EEG-System (BrainProducts GmbH; Gilching, Germany). The electrodes were positioned on an extended 10–20 system (Jasper, 1958; Figure 3), plus two electrodes placed at the mastoids. Additionally, four electrodes were positioned around the eyes to record blinks and eye-movements: one electrode above and below the right eye and one electrode at the outer canthus of each eye. The ground electrode was placed at the electrode site AFz and the online reference was located at the electrode site FCz. The signal was recorded with electrode impedances lower than 2 KΩ for the reference and ground and lower than 10 KΩ for all other electrodes. The sampling frequency was 1000 Hz.

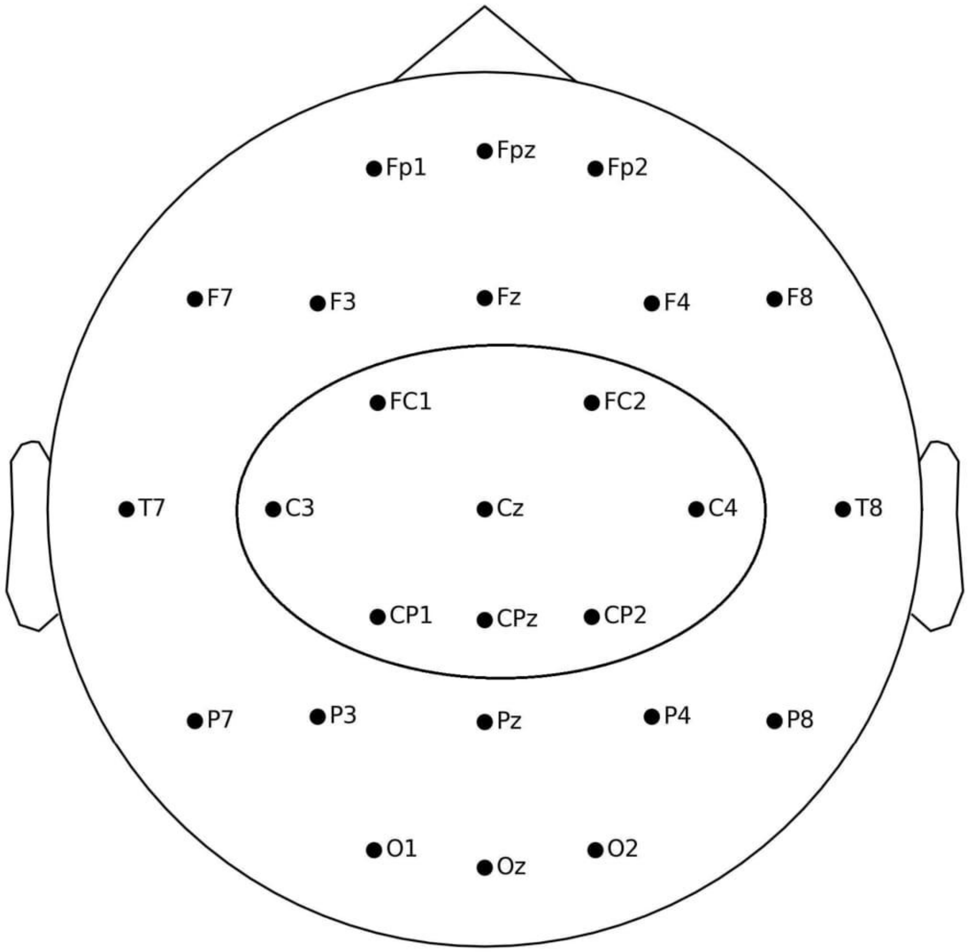



Figure 3. Position of the recorded electrodes. The circled electrodes were included in the N400 analysis.

***EEG Processing***

The data were analysed using the MNE-Python (Gramfort et al., 2014). The data were re-referenced offline to the average of the left and right mastoid. Eye movements were corrected using an independent component analysis (ICA) with the 'fastica' algorithm (Hyvärinen & Oja, 2000). The 'AutoReject' algorithm (Jas et al., 2016) was used to automatically detect and remove bad epochs and channels. Any removed channels were corrected through interpolation of the surrounding channels. One participant was removed from further analysis due to artifact rejection leading to the removal of more than 50% of trials. From the remaining participants, a total of 11.2% of the trials were removed due to artifacts. This resulted in the following average number of trials per condition being analysed: 35.1 for predictable – no load, 35.8 for unpredictable – no load, 35.4 for predictable – load, and 35.8 for unpredictable – load.

***Statistical analysis***

Our critical condition is the predictable – load condition. If an increase in vWM demand interferes with the strength of pre-activation of a target word, we expect to see some evidence of an N400 effect in this condition (i.e., a negative deflection centred around 400 ms after target word onset). If this is not the case, we expect no difference between the predictable – load condition and the predictable – no load condition; the pre-activation of the target word would remain the same regardless of vWM. Since we do not expect vWM demand to have an effect on the pre-activation of a target word, we analyse the data using a Bayesian repeated measures ANOVA. Frequentist approaches are less appropriate when trying to provide evidence for a null effect, as a non-significant finding is not necessarily informative. Using a Bayesian approach and by reporting Bayes Factors (BF), we can quantify if our data are in support of an effect ($BF_{10} > 1$ or $BF_{01} < 1$), if there is an absence of evidence ($BF_{10} \sim 1$ or $BF_{01} \sim 1$), or if there is evidence of the absence of an effect ($BF_{10} < 1$ or $BF_{01} > 1$; Keysers et al., 2020).

We conducted a 2x2 Bayesian repeated-measures ANOVA testing the effect of *predictability* (predictable or unpredictable) and *concurrent vWM demand* (with or without) using uniform model priors as implemented in JASP version 0.95 (JASP Team, 2024). We report evidence in support of the alternative hypothesis as $BF_{10}$, with values > 3 providing substantial evidence. For evidence supporting the null hypothesis, we report $BF_{01}$, again with values >3 providing substantial evidence in favour of the null hypothesis. We also report frequentist ANOVA p-values and $\eta^2_p$ from a 2x2 repeated measures ANOVA. Our dependent variable is the amplitude averaged over a specified timespan and selection of electrodes. For the N400 analysis, amplitudes are averaged between 350 and 550 ms after target word onset and across electrodes Fz, FC1, FCz, FC2, C3, Cz, C4, CP1, CPz, and CP2.

## Results

### *Behavioural data*

Participants correctly chose the target picture after sentence presentation in 95% (*SD* = 0.04) of trials, with participant accuracies ranging between 85% and 99% and condition accuracies ranging from 94.4% (unpredictable – load) to 95.8% (predictable – load). Participants correctly recalled the visuospatial cognitive load manipulation in 77% (*SD* = 0.14) of the load trials, with average participant accuracies ranging between 48% and 100%.

### *ERP analysis: N400*

Figure 4 shows the ERP waveforms (negative values going down) for each of the four conditions across each of the analysed electrode sites. Visual inspections of the graphs show clearly larger negative deflections at around 400 ms in the unpredictable conditions. Both the Bayesian and frequentist ANOVA show evidence for an effect of predictability, with unpredictable stimuli eliciting more negative amplitudes across the spatiotemporal interest area ($BF_{10} = 4.6$; $F_{1,31} = 8.35$, $p < .01$, $\eta^2_p$=0.21). The frequentist ANOVA shows no evidence for an effect of cognitive load ($F_{1,31} = 0.23$, $p = .63$, $\eta^2_p$ =0.008) and no evidence of an interaction between predictability and cognitive load

($F_{1,31} = 0.09$, $p = .77$, $\eta^2_p = 0.003$). The Bayesian model provides substantial evidence against an effect of visuospatial cognitive load ($BF_{01} = 3.4$) and against an interaction between predictability and visuospatial cognitive load ($BF_{01} = 3.9$). Specifically, the observed data for visuospatial cognitive load and the interaction between predictability and cognitive load are, respectively, 3.4 and 3.9 times more likely under the null hypothesis than under the alternative hypothesis. The results of this analysis are available online at

https://osf.io/xhc35/overview?view_only=803aeb4030a944df961eca062c7f3f37

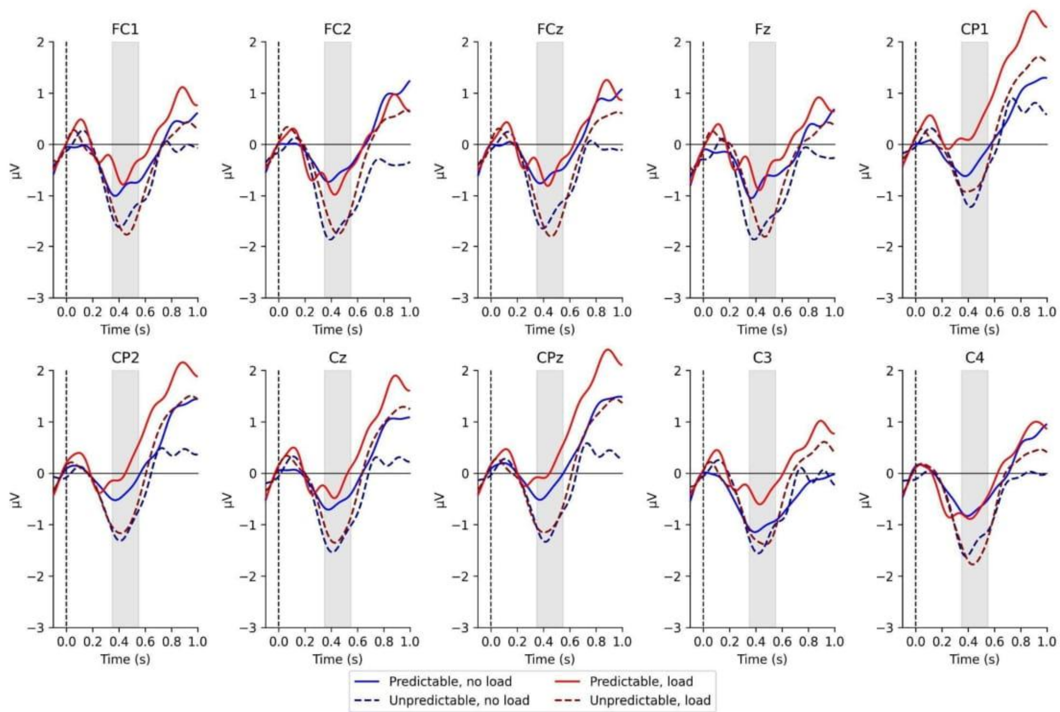



Figure 4. Grand average ERP results (negatives plotted down) for each of the four conditions at each region of interest. Solid blue lines represent the predictable – no load condition, dashed blue lines represent the unpredictable – no load condition, solid red lines represent the predictable – load condition, and dashed red lines represent the unpredictable – load condition. The grey shaded box represents the analysed N400 time window (350 to 550 ms after target onset). ERPs presented here were filtered using a 15 Hz low-pass filter for visualization purposes.

## Discussion

In this experiment, participants listened to predictable and unpredictable sentences either with or without a concurrent vWM task. Randomly and in half of the trials, participants were required to remember the visuospatial location of 4 indicated squares before listening to a sentence and recall the order and location of these squares afterwards. The N400 ERP component, 350-550 ms after target word onset, was measured. We found substantial evidence for an effect of predictability on N400 amplitude and substantial evidence for the absence of an effect of vWM and for the absence of an interaction between predictability and vWM. The present findings confirm that L2 English speakers are capable of pre-activating target words even when completing a demanding secondary task. Importantly, the present study suggests that increases in vWM may not interfere with the pre-activation of a target word or the features associated with a target word in the absence of a relevant co-present visual world that the speech is referring to.

### *Working memory and prediction in visual world studies*

One interpretation of the present findings thus is that at least some of the effects of working memory on prediction are inherent to predictive *gaze* behaviour specifically, especially in the context of the visual world paradigm. Intrinsic to the visual world paradigm is the interplay of language, vision, memory, and attention (Huettig, Olivers, et al., 2011), and working memory plays a role in each of these. Indeed, the model of language-vision interaction proposed by Huettig, Olivers, et al. (2011) suggests that working memory is the key component through which visual, linguistic, and location information are bound (cf. Altmann & Mirkovic, 2009). While visual and linguistic representations for an object are stored in long term memory, in people's interaction in their visual surroundings and especially in the visual world paradigm, this representational knowledge is mediated by the objects' temporary and arbitrary current location. In this model, a crucial step during language-vision interaction is the binding of (1) at least some degree of object representation (with progressively more detailed information encoded with increasing time or object familiarity) with (2) the object's current arbitrary location in space. This binding results in a temporary nexus of linguistic and visuospatial information which is held in working memory. Incoming spoken language activates stored representational knowledge and the overlap between this activated knowledge and the representations in the temporary nexus of information can then guide gaze behaviour.

This nexus may then be serving as the source for disruptions due to the vWM manipulations shown in previous visual world eye-tracking research (Allison et al., 2025; Liu et al., 2025; Wang et al., 2025). Specifically, increases in vWM may lead to the formation of a representational nexus with relatively weaker links or 'incomplete' links between stored information and its current visuospatial index. Weaker or absent linkings are then less able to guide fixations and, importantly, this entire process is contingent on a current visual environment. Thus, while predictive gaze behaviour in visual world eye-tracking studies may require vWM resources, such resources may not be required during spoken language prediction in the absence of co-present objects. Further research is required to confirm this conclusion.

***Domain general vs domain specific working memory***

Our findings, while not providing strong evidence in this regard, may also merit some discussion of domain-general and domain-specific involvements of working memory in predictive language processing. The evidence against an effect of vWM may suggest a lack of an influence of domain general working memory resources in predictive language processing. While the role of working memory in predictive processing is still debated, Ryskin et al. (2020) propose the following possibilities: (1a) executive resources are needed to maintain the *context* in working memory, (1b) executive resources are needed to generate/maintain *predictions* in working memory), and (1c) linguistic prediction is implemented in domain-general inhibitory and selection mechanisms. Alternatively, they describe a hypothesis (2) in which linguistic prediction does not recruit executive resources and that differences can be explained through more usage-based mechanisms.

Note that strong tests of the hypotheses put forward by Ryskin and colleagues require strong manipulations of amount and type of cognitive demands. The relevance of the present findings to their hypotheses hence is somewhat limited. In regard to the type of language-vision interactions tapped by the visual world paradigm, however, their hypothesis 1a, arguably, is *not* well-supported. If limiting (domain-general) executive resources resulted in a poorer representation of the current context, one would also expect a reduction in an unpredictable condition with a concurrent vWM demand. However, both Allison et al. (2025) and Wang et al. (2025) found that disruptions were specific to the predictable condition. This suggests that maintaining the current context (at least with simple visual displays) does not require resources shared with vWM.

Given the results of the present study, hypothesis 1b also seems unlikely. If (domain-general) executive resources were required for the generation or maintenance of predictions, the N400, as a measure thought to index the degree to which the features of a target word are currently pre-activated, should be sensitive to this. As we found no modulation of the N400 due to vWM, this suggests that generating or maintaining predictions in working memory does not require resources shared with vWM (in the absence of contextually relevant visual input).

Ryskin's hypothesis 1c proposes that (domain-general) executive resource availability controls how well low-probability continuations are inhibited and how well high-probability continuations are selected. In limited resource situations, they suggest that the selection and inhibition processes may be impacted and lead to less specific or less accurate predictions. The results of the present study also seem to make this interpretation unlikely, as this hypothesis is also contingent on feature activation. If reducing the availability of executive resources reduces the likelihood of selecting high probability continuations and thus activating the features associated with this continuation, we would also expect to see this reflected in the N400 component.

Taken together, the findings from the previous eye-tracking research and this study are somewhat more in support of Ryskin's hypothesis 2, that linguistic predictions are rooted in language experience and may not recruit executive resources (at least with the simple type of language usually used in visual world studies). Specifically, it seems less likely that prediction of spoken language uses cognitive resources that are shared with vWM. Instead, we suggest that the previously reported disruptions in predictive gaze behaviour due to vWM are contingent on the processing of concurrent visual objects during language-vision interactions. In the absence of relevant visual information, there seems to be little role for vWM during prediction in spoken language processing. Even in the presence of a visual array in visual world eye-tracking studies, participants with a concurrent vWM demand are able to start predicting at times comparable to control conditions (Liu et al., 2025; Wang et al., 2025). This suggests that the reductions seen in predictive gaze behaviour are more likely due to issues linking an object's activated representation with its current location (cf. the blank screen studies, Altmann, 2004; Richardson & Spivey, 2000; Spivey & Geng, 2001).

### *Conclusion*

The main purpose of this study was to use EEG to probe prediction in spoken language processing by decoupling sentence processing and the visuospatial processing inherent to visual world eye-tracking studies. To achieve this, we measured the N400 ERP response to a set of auditory stimuli that have previously shown disruptions in predictive gaze behaviour in a visual world eye-tracking task due to the same vWM task in a comparable group of L2 English speakers (Allison et al., 2025). Thus, we assume that the vWM task had comparable effects on spoken sentence processing in the visual world paradigm study and the present ERP study. Whereas the previous eye-tracking study found disruptions in predictive gaze behaviour due to the vWM task, we found in the present study that L2 English speaking participants equally pre-activated semantically predictable target words with or without a concurrent vWM demand. This finding is most compatible with the conclusion that reductions in the evidence of prediction due to increases in vsWM demand are contingent on a visual array and that, while such increases may reduce predictive gaze behaviour towards semantically predictable targets, this reduction is specific to situations in which prediction in language occurs within a relevant co-present visual environment.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Data availability**

The data that support the findings of this study are openly available on the Open

Science Framework at

https://osf.io/xhc35/overview?view_only=803aeb4030a944df961eca062c7f3f37

## References


Allison, C., Huettig, F., Fernandez, L., & Lachmann, T. (2025). Visuospatial working memory load reduces semantic prediction in the visual world. *Language, Cognition and Neuroscience*, *40*(9), 1252–1261. https://doi.org/10.1080/23273798.2025.2522272

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm.' *Cognition*, *93*(2), B79–B87. https://doi.org/10.1016/j.cognition.2004.02.005

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

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, *26*(9), 1338–1367. https://doi.org/10.1080/01690965.2010.542671

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

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

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

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*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Fernandez, L. B., Hadley, L. V., Koç, A., Gamboa, J. C., & Allen, S. E. (2025). Is there a cost when predictions are not met? A VWP study investigating L1 and L2 speakers. *Quarterly Journal of Experimental Psychology*, *78*(7), 1237–1259. https://doi.org/10.1177/17470218241270200

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing

MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Hintz, F., Voeten, C. C., Dobó, D., Lukics, K. S., & Lukács, Á. (2024). The role of general cognitive skills in integrating visual and linguistic information during sentence comprehension: Individual differences across the lifespan. *Scientific Reports*, *14*(1), 17797. https://doi.org/10.1038/s41598-024-68674-3

Hohwy, J. (2013). *The Predictive Mind*. OUP Oxford.

Huettig, F. (n.d.). *Looking Ahead*. Cambridge University Press & Assessment. Retrieved November 3, 2025, from https://www.cambridge.org/universitypress/subjects/psychology/cognition/looking-ahead-new-science-predictive-mind

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

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

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

JASP Team (2025). JASP (Version 0.95.3) [Computer software].

Jasper, H. (1958). Ten-twenty electrode system of the international federation. *Electroencephalography Clinical Neurophysiology*, *10*, 371–375.

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

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Liu, S., Hu, W., & Liu, X. (2025). Different effects of verbal and visual working memory loads on Language prediction. *Scientific Reports*, *15*. https://doi.org/10.1038/s41598-025-03556-w

Mantegna, F., Hintz, F., Ostarek, M., Alday, P. M., & Huettig, F. (2019). Distinguishing integration and prediction accounts of ERP N400 modulations in language processing through experimental design. *Neuropsychologia*, *134*, 107199. https://doi.org/10.1016/j.neuropsychologia.2019.107199

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

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

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295. https://doi.org/10.1016/S0010-0277(00)00084-6

Ryskin, R., Levy, R. P., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*, 107258. https://doi.org/10.1016/j.neuropsychologia.2019.107258

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: What is next? *Trends in Cognitive Sciences*, *27*(11), 1032–1052. https://doi.org/10.1016/j.tics.2023.08.003

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

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

Wang, Z., Hui, C. K. E., & Ito, A. (2025, March 27-29). *Domain-general working memory modulates semantic prediction in native speakers* [Poster Presentation]. Annual Conference of the Society for Human Sentence Processing. College Park, Maryland, United States. https://www.researchgate.net/publication/390760945_Domain-general_working_memory_modulates_semantic_prediction_in_native_speakers

*Bilingualism: Language and Cognition*

**cambridge.org/bil**

**Research Article**

**Cite this article:** Fernandez, L.B., Hadley, L.V., Gamboa, J.C.B., Allison, C. and Allen, S.E.M. (2025). The impact of speech rate on first- and second-stage prediction in L1 and L2 speakers. *Bilingualism: Language and Cognition* 1–13. https://doi.org/10.1017/S1366728925100515

Received: 29 January 2025
Revised: 5 August 2025
Accepted: 26 August 2025

**Keywords:**
prediction; speech-rate; bilingualism; Visual World Paradigm; eye-tracking

**Corresponding author:**
Lauren V. Hadley;
Email: lauren.hadley1@nottingham.ac.uk

This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.



CAMBRIDGE UNIVERSITY PRESS

# The impact of speech rate on first- and second-stage prediction in L1 and L2 speakers

Leigh B. Fernandez[1], Lauren V. Hadley[2], John C.B. Gamboa[1], Christopher Allison[1] and Shanley E.M. Allen[1]

[1]Center for Cognitive Science, RPTU University of Kaiserslautern-Landau, Kaiserslautern, Germany and [2]Hearing Sciences – Scottish Section, School of Medicine, University of Nottingham, Glasgow, UK

## Abstract

We investigate timing and eye-movement behavior during semantic prediction in L1 and L2 speakers of English using the Visual World Paradigm, additionally exploring speech rate. We differentiate first-stage predictions, considered to be automatic and relatively cost-free, from second-stage predictions, which are non-automatic and more cognitively demanding, with differences between L1 and L2 speakers believed to arise in second-stage predictions. We found no differences in the divergence of looks to the target in first- or second-stage predictions across groups. However, speech rate played an important role. Both L1 and L2 speakers showed similar first-stage predictions at slower speech rates, but L1 speakers showed earlier predictions as the speech rate increased. L2 speakers showed reduced and more variable second-stage predictions, suggesting they were impacted during the more demanding second-stage prediction. This may indicate a wait-and-see strategy to help reduce costs associated with second-stage prediction.

## Highlights

- L1 and L2 speakers showed evidence of early- and late-stage predictions
- No timing differences between predictions in L1 and L2 speakers
- L1 and L2 speakers are differentially impacted by speech rate
- Increased speech rate specifically impacted L2 late-stage predictions
- Supports research that late-stage prediction is more demanding

## 1. Introduction

Processing a spoken utterance requires multiple complex mechanisms working together. Sounds must be perceived, meaning must be assigned to these sounds and these meanings must be combined to create something comprehensible. Statistical regularities in each of these steps lead to the ability to predict likely continuations of the input. An ever-increasing amount of research shows that first language speakers (L1) can make use of multiple sources of information to predict upcoming linguistic information before encountering it (for reviews, see Ferreira & Chantavarin, 2018; Huettig, 2015; Huettig & Mani, 2016; Kamide, 2008; Staub, 2015).

When it comes to second language speakers (L2), early research did not find evidence for prediction (e.g., Grüter et al., 2012; Lew-Williams & Fernald, 2010; Martin et al., 2013). However, more recent research has found that L2 speakers do indeed predict (for reviews see Hopp, 2022; Kaan & Grüter, 2021). This has led to the general consensus that L2 speakers can make predictions, and that the mechanisms involved in prediction are the same for L1 and L2 speakers (e.g., Kaan, 2014). When differences arise between L1 and L2 speakers, they may be explained by other factors such as individual differences (e.g., proficiency, quality of lexical representation, cognitive resources, processing strategies, etc.; Kaan, 2014), methodological factors (e.g., speech rate; Fernandez et al., 2020), or a combination of both. In the current study, we test whether L1 and L2 speakers of English show differences in the timing and pattern of eye-movement behavior during semantic prediction, particularly at different stages of prediction, while exploring a methodological factor of speech rate.

### 1.1. Prediction stages

While it is possible that predictions could occur at one stage, Pickering and Gambi (2018) have argued that prediction involves two stages. They claim that the first stage is an early, rapid, relatively resource-free automatic stage driven by spreading activation that characterizes semantic priming, while the second is a later, slower, more integrative and resource-demanding non-automatic stage that makes use of more real-world information. Additionally, this later, more resource-demanding stage has been argued to be the likely source of differences in L1 and L2 predictive capabilities (Ito & Pickering, 2021). In the next sections, we provide evidence for two-stage prediction in L1 speakers, followed by evidence in L2 speakers. This is followed by a section outlining prediction in relation to our exploratory factor, speech rate.

#### 1.1.1. Evidence for two-stage prediction in L1 speakers

As just mentioned, the first stage of prediction is theorized to be a fast, automatized and thus relatively cost-free form of prediction largely rooted in spreading activation. While this type of prediction is rapid, it is not necessarily accurate, given that lexical activation will include associatively related concepts which may not fit contextually within a sentence. This type of first-stage prediction can be seen using the Visual World Paradigm (VWP). For example, Kukona et al. (2011) found that L1 English participants look predictively toward pictures of a *crook* and a *policeman* after hearing *Toby will arrest*, suggesting that these predictive eye movements stem from the verb *arrest* rather than from sentence context. However, unlike Kukona et al. (2011), Gambi et al. (2016) found that after hearing *Pingu will ride*, L1 children and adults looked only to the appropriate target (*horse*), and not at the agent-related (*cowboy*), suggesting that participants did not make associatively related (and ultimately incorrect) predictions. The authors argued that this different pattern of results may be due to the auditory recordings being quite slow (given that they were used with children), which may have led to unrepresentative predictive eye movements. Therefore, we will explore the contribution of speech rate to both stages of prediction.

The second stage of prediction, on the other hand, is theorized to be a slower, non-automatized and thus a resource-demanding form of prediction that involves integrating real-world knowledge or contextual information together with the linguistic input. While this type of prediction is slower, it allows predictions to be tailored to a given context and, therefore, potentially more accurate than those based on the first stage alone. For example, Corps et al. (2022) conducted a study that found evidence for both stages of prediction using the VWP. In their study, young adult participants heard sentences such as *I would like to wear the nice* spoken by either a male or female while viewing two verb-related objects (wearable, e.g., *tie* or *dress*) and two non-verb related objects (non-wearable, e.g., *drill* or *hairdryer*). These objects were selected such that one of each type was stereotypically masculine (e.g., *drill* and *tie*) and one was stereotypically feminine (e.g., *hairdryer* and *dress*). They found that listeners quickly predicted based on semantic association, evidenced by an equal increase in proportion of looks toward the two wearable objects (over the non-wearable objects), approximately 500 ms after the onset of *wear*. However, they also found that listeners made a later prediction based on real-world (gender stereotyping) knowledge evidenced by looks approximately 640 ms after the onset of *wear* toward the wearable object that stereotypically matched the gender of the speaker (over the wearable objects not stereotypically worn by the gender of the speaker; i.e., the male saying *wear* led to increased looks to *tie* over *dress* and the female saying *wear* led to increased looks to *dress* over *tie*). This suggests that L1 speakers make first-stage automatic predictions based on spreading activation (*wear* leading to an equal increase in looks to *tie* and *dress*) and second-stage non-automatic predictions based on real-world knowledge (male speakers leading to an increase of looks to *tie* and female speakers leading to an increase of looks to *dress*).

Further evidence for two stages of prediction comes from a VWP study comparing prediction in younger adults (mean age = 20.35 years) and older adults (mean age = 68.87 years) (Fernandez et al., 2025). In the stimuli, two objects were initially relevant, but could be narrowed down in a second step. Specifically, participants listened to sentences like *The singer played the guitar* while viewing a target (*guitar*), an agent-related object (*microphone*), a verb-related object (*cards*) and a distractor object (*strawberry*). While *singer* is related to both *guitar* and *microphone*, upon hearing the verb *played*, the prediction could be narrowed to *guitar*. The authors found that both groups showed increased looks to the target *guitar* and agent-related *microphone* after hearing the agent *singer*; though younger adults were significantly earlier (during the presentation of the agent *singer*) than older adults (soon after the onset of the verb *plays*). Both groups then showed later prioritization of *guitar* over the agent-related *microphone*, though interestingly, the older adults looked at *guitar* significantly earlier (within the latter part of the presentation of the verb *plays*) than the younger adults (within the latter part of the presentation of the article prior to the target *the*). The authors argued that the increased looks to *guitar* after hearing *singer* constituted first-stage prediction based on spreading activation from the agent, and the prioritization of *guitar* over *microphone* constituted second-stage prediction based on tailoring predictions following the verb (i.e., actively ruling out the agent-related object *microphone* based on combining both the agent and verb; *singer* + *plays*). The authors further argued that the delayed first-stage prediction by older adults relative to younger adults stems from age-related decreases in semantic network efficiency and lexical access speed (e.g., Cosgrove et al., 2021, 2023) and that the quicker second-stage prediction was related to age-related increases in real-world knowledge, efficiency of inhibiting irrelevant information and efficiency in shifting resources between targets (Veríssimo et al., 2022; but see, for example, Buckner, 2004; Fjell et al., 2017; Hasher & Zacks, 1988, for contrary evidence).

Thus, there is increasing recognition that there may be different stages and mechanisms involved in prediction, and that these stages may be differently affected by individual difference measures.

#### 1.1.2. Evidence for two-stage prediction in L2 speakers

As previously mentioned, when it comes to prediction in L2 speakers, the general consensus is that L1 and L2 prediction mechanisms are the same, and that differences that arise between these groups are most likely due to individual differences and/or methodological factors. Given that processing an L2 is more cognitively demanding (e.g., Segalowitz & Hulstijn, 2009), and that only second-stage prediction is hypothesized to involve substantial cognitive resources, it stands to reason that it may be second-stage predictions that are particularly impacted in L2 speakers (e.g., Ito & Pickering, 2021).

Recently, Corps et al. (2023) tested L2 speakers of English using the same items from their L1 study (Corps et al., 2022) and directly compared the timing of the two stages of prediction across groups. Similar to L1 speakers, Corps et al. (2023) found evidence for two-stage prediction in L2 speakers. Furthermore, when the authors compared the two groups, they found no differences in the timing of first-stage prediction, but found that L2 speakers were delayed in second-stage prediction relative to L1 speakers.

Additional potential evidence for the stages of prediction in L2 speakers comes from Peters et al. (2018) who investigated prediction with lower skilled participants (i.e., participants who scored low on a vocabulary test or self-identified as an L2 speaker) and higher skilled participants (i.e., participants who scored high on a vocabulary test or self-identified as an L1 speaker) using the VWP. In their first experiment, Peters et al. used stimuli similar to Fernandez et al. (2025) in that the visual array for the predictable items contained two objects related to the agent, which were narrowed down by the verb. For example, participants heard a sentence like *The pirate hides the treasure* while viewing four objects: a target (*treasure*), an agent-related object (*ship*), a verb-related object (*bone*) and an unrelated distractor (*cat*). Peters et al. found that both groups made predictions after hearing *The pirate* in terms of looks to the target (*treasure*) and the agent-related object (*ship*), followed by increased looks toward *treasure* (and decreased looks toward *ship*) after hearing *hides.* While Peters et al. do not operationalize their findings in terms of stages of prediction, we argue that this finding is evidence of second-stage prediction. Interestingly, and unlike Corps et al., Peters et al. did not find differences at this potential second stage of prediction across groups, but it is important to note that they could not accurately estimate (or compare) the time at which the groups looked toward the target relative to the agent-related object. Visual inspection of their fixation graphs seems to anecdotally suggest that participants who identified as L1 speakers looked earlier to the target than participants who did not identify as L1 speakers. In addition, Peters et al. do not directly test first-stage prediction (i.e., they do not compare the target object to the verb-related or distractor item), though visual inspection again provides anecdotal evidence of an early divergence between the target and the verb-related and distractor item across groups.

While we hesitate to draw too many parallels between Corps et al. and Peters et al., we believe that both studies provide evidence that prediction involves two stages in L2 speakers: a resource-free first stage based on spreading activation and a more costly second stage involving tailoring based on real-world knowledge given the situation, which may be additionally impacted in a cognitively demanding L2. However, research investigating both stages is limited. In the next section, we discuss how the two stages may be influenced by differences based on the stimuli (i.e., speech rate).

### 1.2. Exploratory factor

As previously outlined, the prediction mechanisms employed by L1 and L2 speakers are believed to be qualitatively the same, and when quantitative differences arise across these groups, they are believed to stem from individual or methodological factors (e.g., Kaan, 2014). Thus, we expect both L1 and L2 speakers to generate first- and second-stage predictions, but expect that the timing of their predictions may differ due to the difference in cognitive resources required for L1 versus L2 language processing. Since speech rate also affects the resource requirements for language processing (Müller et al., 2019), it is possible that speech rate drives some of the quantitative differences that have been reported between groups. Therefore, in the current study, we directly explore how speech rate impacts these first- and second-stage predictions in both L1 and L2 speakers.

Given that the VWP focuses on spoken stimuli, it comes as no surprise that the rate at which the spoken stimuli are presented can impact a listener's ability to make predictions. Unfortunately, there seems to be no consensus on what constitutes a "normal" speech rate (being reported as anywhere between 2.5 and 8.0 syllables per second; see Fernandez et al., 2020). Additionally, research using the VWP does not consistently report speech rate, and when speech rate is reported, it is often reported as "normal" (which is a potentially broad range) or it tends to be on the slower side (e.g., 3 syllables per second or less; Fernandez et al., 2020). For example, Kukona (2023) estimated that the speech rate used in Peters et al. (2018) study investigating the narrowing of prediction based on the verb (e.g., *The pirate hides the treasure*) was approximately 3.0 syllables per second. Using the VWP, Huettig and Guerra (2019) found that prediction only occurred for L1 speakers when speech rate was "slow" relative to "normal" (though the authors do not report the speech rate, Kukona estimated the rate to be 2 and 4.5 syllables per second, respectively). In a study that more directly manipulated speech rate in the VWP, Fernandez et al. (2020) tested predictive eye movements with sentences presented at 3.5, 4.5, 5.5 and 6 syllables per second. Fernandez et al. found that both college-aged L1 and L2 speakers showed an inverse-U pattern with predictive eye movements increasing from slower speech rates up to an "optimal" speech rate (where prediction was highest) and decreasing at faster speech rates. The authors found that L1 speakers showed the highest level of prediction at 5.5 syllables per second, while L2 speakers showed the most efficient prediction at speech rates slightly faster than 3.5 syllables per second (but slower than 4.5 syllables per second). Together, these studies, along with the Gambi et al. (2016) study previously mentioned (which found no evidence of agent-related activation at slow speech rates), suggest that speech rate plays an important role in the VWP and that different speech rates may lead to different patterns of predictive behavior. Therefore, in this study, we include speech rate as a continuous variable in our models (and its interaction with language (L1/L2), and with language across time).

## 2. Current study

Research suggests that both L1 and L2 speakers make first-stage predictions similarly, while L2 speakers may be delayed in terms of more costly second-stage predictions. Research also suggests that methodological factors, such as speech rate, may impact the first and/or second stage of prediction. However, research on the stages of prediction is rather limited and often contradictory. Therefore, in the current study, we test first- and second-stage prediction with L1 and L2 speakers of English while investigating the role of speech rate. In addition, we use statistical approaches uniquely suited to investigate these stages of prediction: we use regressions that allow us to model the potential non-linearity of eye-movement patterns and model speech rate as a continuous variable (General Additive Mixed Models (GAMM); e.g., Porretta et al., 2018; Wieling, 2018; Wood, 2017), as well as a non-parametric bootstrapping approach that allows us to compare divergence times across groups (Divergence Point Analysis, DPA; Stone et al., 2021).

In terms of first-stage predictions, it could be the case that L2 speakers are delayed relative to L1 speakers due to weaker semantic network efficiency. This would be similar to previous findings showing older adults having a delayed onset of first-stage predictions (Fernandez et al., 2025). However, we believe that this effect is (at least to some degree) age-specific. When comparing groups of younger adults, semantic predictions generally have been shown to be quite similar between L1 and L2 speakers and Corps et al. (2022, 2023) specifically found similar onset of first-stage predictions in L1 and L2 speakers. Thus, we hypothesize that both L1 and L2 speakers will make first-stage predictions at similar times. In terms of second-stage prediction, we hypothesize, again in line with Corps et al., that L2 speakers will show delayed second-stage predictions relative to L1 speakers, given the higher resource intensiveness of L2 language processing. In relation to speech rate, we hypothesize an inverse-U pattern. Specifically, we hypothesize that prediction (in the form of looks to the target relative to a competitor) will increase up until an "optimal" rate and then decrease at speech rates higher than the "optimal" rate (Fernandez et al., 2020). Additionally, we believe that this "optimal rate" (or the peak of the inverse-U) will be different for L1 and L2 speakers. For first-stage predictions, we hypothesize that looking at the target may increase until speech rates slightly faster

**Table 1.** Participant information

| N | L1 | Female/ Male | Mean age (sd) | Mean age (sd) of English acquisition | Mean score (sd) on English proficiency test – Oxford Placement Test (OPT – scale from 0–100) |
|---|---|---|---|---|---|
| 46 | English | 36/10 | 20.28 (3.54) | NA | 91.17 (5.74) |
| 45 | German | 28/17 | 25.75 (4.86) | 9.55 (1.84) | 77.96 (10.59) |

than 3.5 syllables per second for L2 speakers, and up to 5.5 syllables per second (though in the current study, the fastest speech rate is 4.6 syllables per second) for L1 speakers (as found in Fernandez et al., 2020). For second-stage predictions, we hypothesize that changes in speech rate may impact both groups and that higher speech rates may be particularly impactful in second-stage predictions for L2 speakers, given the increased cognitive demands of this stage. For example, the optimal rate may shift for L2 speakers, such that looks to the target start decreasing at rates slower than approximately 3.5 syllables per second.

## 3. Methods

Sample size was based on a recent webcam-based eye-tracking study investigating verb-based prediction, which found that 40 participants (and 16 items per condition) obtained approximately 90% power (Prystauka et al., 2024).

### *3.1. Participants*

#### 3.1.1. L1 English

Fifty L1 English speakers were recruited from the University of Alberta (Canada) for course credit. Four participants were excluded since they were raised in a bilingual environment. Therefore, 46 monolingually raised L1 English speakers were included in the study, none of whom were exposed to an L2 before the age of 5. All participants reported normal hearing and normal or corrected-to-normal vision. See Table 1 for additional information.

#### 3.1.2. L2 English

Forty-five L2 English speakers (L1 German) were recruited from the University of Kaiserslautern-Landau (Germany), for course credit or payment (8 Euro). All participants reported being monolingually raised L1 German speakers, who had not been exposed to an L2 before the age of 5. All participants reported normal hearing and normal or corrected-to-normal vision. See Table 1 for additional information.

### *3.2. Materials*

The sentence stimuli were taken from Holt et al. (2021)[1] and were identical to those used in Fernandez et al. (2025). All sentences followed the same structure (see Table 2): The [agent] [verb] the [critical word]. There were 32 pairs of sentences, each consisting of one predictable and one unpredictable sentence. The sentence

**Table 2.** Example stimuli from Holt sentences and item information (range in parentheses)

| | Example stimuli | Overall item information | |
|---|---|---|---|
| | Sentence pair | Mean (range) Zipf frequency by agent, verb, & critical word | Mean (range) syllable count by agent, verb, & critical word |
| Predictable | The singer plays the guitar. | Agent = 4.17 (2.96–5.29)<br>Verb = 3.88 (2.44–5.99)<br>Critical word = 4.32 (3.25–4.95) | Agent = 1.41 (1–3)<br>Verb = 1.78 (1–2)<br>Critical word = 1.41 (1–3) |
| Unpredictable | The cousin forgets the guitar. | Agent = 4.75 (2.92–5.86)<br>Verb = 4.16 (2.57–6.51)<br>Critical word = 4.34 (3.25–4.95) | Agent = 1.78 (1–3)<br>Verb = 1.62 (1–4)<br>Critical word = 1.41 (1–3) |

pairs were manipulated such that the critical word was the same across both sentence types, but was either predictable or unpredictable based on the preceding information. Based on the SUBTLEX-UK database of Zipf frequencies (van Heuven et al., 2014), all agents, verbs and critical words were medium to high frequency. See Table 2 for example items, the Zipf frequency of the agent, verb and critical word, and the mean syllable count per word.

Each sentence pair had a corresponding visual array consisting of four objects[2]. For the example in Table 2, this included: a target object (*guitar*), an agent-related object for the predictable sentence (related to the agent but not the verb; *microphone*), a verb-related object for the predictable sentence (related to the verb but not the agent; *cards*) and a distractor object (not related to the agent, verb or target; *strawberry*). For the unpredictable sentences, all objects were unrelated to the agent and verb. Similarly, to the sentence words, the object names' frequencies ranged from medium to high based on the SUBTLEX-UK database (M = 4.14, SD = 0.61). Additionally, to avoid phonological overlap, none of the words for the objects shared initial phonemes. All objects were grayscale 300x300 pixel jpeg objects and the placement of the objects was rotated and counterbalanced across items (objects provided by Holt et al., 2021).

Predictability was established with an online pre-test (with L1 speakers of English), in which participants were provided with truncated sentences (the experimental sentences with the target word absent) and the four images in the array. Participants were then asked to choose the image that they believed would likely come next. Participants chose the target image over the other images 98.08% (SD 4.39) of the time for the predictable items and 25.26% (SD 16.54) of the time for the unpredictable items (for additional pre-testing information, see Fernandez et al., 2025).

#### 3.2.1. Auditory information

All items were recorded by a male L1 Scottish-English speaker with a Blue Yeti USB microphone at a sampling rate of 48,000 Hz using Audacity® recording software (Audacity Team, 2021). For ease of post hoc acoustic manipulation, a click track (90 beats per minute) was used to ensure all constituents were spoken within a similar time frame. All constituents were normalized to the mean of that

[1]This study investigated prediction with children with hearing loss. Holt et al. did not analyze or report second stage predictions.

[2]For copyright reasons we do not include the objects here.

https://doi.org/10.1017/S1366728925100515 Published online by Cambridge University Press

**Table 3.** Mean normalized constituent durations (ms)

| | 1st The | Agent | Verb | 2nd The | Object |
|---|---|---|---|---|---|
| Mean (ms) | 93.58 | 612.72 | 602.05 | 130.27 | 464.45 |

constituent across all items using Praat (Boersma & Weenink, 2021), making all corresponding constituents the same length across items (see Table 3 for mean normalized constituent duration). While all the items were same duration, they naturally differed in speech rate. Speech rate for the items was determined by dividing the number of syllables in each item by the duration of the item, and the items ranged from 2.5 to 4.6 syllables per second (see OSF for a visualization of the distribution; https://osf.io/4ke82/).

### 3.3. Apparatus

For both participant groups, eye movements were recorded using an Eyelink 1000 eye-tracker sampling at 1000 Hz, with their head stabilized using a chin rest. Only the right eye was recorded (viewing was binocular). Monitors had a 60 Hz refresh rate and a 1024x768 resolution, and participants listened through Philips Bass + on-ear headphones. The L1 English participants sat approximately 50 cm away from a 20' Dell monitor (model 2009 W1), while the L2 English participants sat approximately 85 cm away from a 19' Dell monitor flat screen cathode ray tube (model P1130).

### 3.4. Procedure

The same procedure was followed for both L1 and L2 speakers in the following order: participants provided their informed consent, took part in the eye tracking tasks, completed a language background questionnaire (LSBQ; Anderson et al., 2018) and finally completed a proficiency test (Oxford Placement Test – Part A; OPT). The study lasted for approximately 45 minutes. The eye-tracking task consisted of 107 trials, three of which were practice, 32 of which were critical items for this experiment (16 predictable/16 unpredictable), and 72 of which served as fillers (experimental items for a study not reported here). The 104 experimental items were placed into two blocks (52 items each) with a break in between.

The eye-tracking task started with a standard 9-point calibration. During the experiment, participants were told there was no time limit and that they were able to pace the study and take breaks as needed (if a participant took a break, they were recalibrated). There was an additional mandatory break between the two blocks, where all participants were recalibrated. Instructions were provided in written form on the screen and in verbal form by the experimenter. Participants were instructed that they would hear a short sentence while viewing an array of objects. Their task was to select the object that they believed best matched the sentence using the mouse after the auditory stimulus. Participants were further instructed that they could not click on the object until after the sentence played in its entirety and a green border appeared around the display.

All trials began with a drift correction in the center of the screen. In order to start the trial, participants were required to look at the drift correct and simultaneously press the space bar. Each trial started with the four objects being displayed on the screen for 2000 ms, after which the auditory stimulus would be played. After the auditory stimulus presentation ended, the objects remained on the screen for an additional 2000 ms and then a green border encompassed the display and the mouse icon appeared. Participants could then use the mouse to click on one of the four objects (participants had to click within the object for their choice to be recorded), which would end the trial.

### 3.5. Analysis

To investigate the pattern of looks to the objects and the contribution of speech rate, we used generalized additive mixed models (GAMMs; e.g., Porretta et al., 2018; Wieling, 2018; Wood, 2017), and to investigate the timing of looks to different objects we used diverge point analysis (DPA; Stone et al., 2021), both are ideal for VWP analyses. GAMMs model both linear and non-linear relationships and can deal with the autocorrelation that is inherent to eye movements across short time spans (or consecutive time bins). However, GAMMs cannot compare effect onsets across groups. Therefore, we additionally investigate the timing of looks using DPA. DPA can also deal with autocorrelation (since the bootstrapping procedure does not assume a distribution) and Type 1 error that comes from running many tests across a time window.

Similar to Fernandez et al. (2025), analyses included only the predictable items (with the unpredictable items serving as a baseline and to ensure that participants were not employing any particular strategy that might overestimate prediction). To investigate first-stage prediction, we ran a GAMM and a DPA comparing looks to the target (*guitar*) versus looks to the verb-related object (*cards*). Unlike Fernandez et al. (2025), we chose to compare the target to the verb-related object and not to the distractor object, because we believe that this is a more conservative approach, given that *cards* should be activated and ruled out after hearing the verb, while the distractor should not be activated at all. Evidence for first-stage prediction should be seen in the form of rapid spreading activation from *singer*, which would activate *guitar* (and *microphone*), but lead *cards* to be disregarded. In terms of eye movements, this should be evidenced by an increase in looks to the *guitar* (and *microphone*) relative to *cards* soon after *singer* is processed. To investigate second-stage prediction, we ran a GAMM and a DPA comparing looks to the target (*guitar*) versus looks to the agent-related object (*microphone*) as done in Fernandez et al. (2025) and Peters et al. (2018). As second-stage prediction involves the narrowing of prediction based on additional information, evidence for second-stage prediction should be seen in the form of the selection of *guitar* and the inhibition of *microphone* when processing *singer* + *plays* in combination. In terms of eye movements, this should be evidenced by an increase in looks to the *guitar* relative to *microphone* soon after the *singer* + *plays* is processed.

#### 3.5.1. General Additive Mixed Models (GAMM)

For the GAMM analyses, comparisons were made during the predictive time window (from the onset of the agent to the onset of the target). First-stage prediction compared looks to the target (*guitar*) versus looks to the verb-related object (*cards*) in the predictable items. Second-stage prediction compares looks to the target (*guitar*) versus looks to the agent-related object (*microphone*) in the predictable items. Our main dependent variable was the empirical logit of fixation counts, which is the log-odds ratio of looking toward the target relative to looking at another specific object (e.g., *microphone*) in the array (Barr, 2008). Data were grouped into 20 ms bins and were weighted to control for eye-movement-based dependencies (Barr, 2008). The empirical logit was submitted to a GAMM, which allows us to model non-linear time course data (e.g., Porretta et al., 2018; Wieling, 2018; Wood, 2017) using R (R core team, 2018). We included a parametric fixed effect of language (L1/L2) as an ordered factor. We included ordered factor difference smooth interactions of time x language, and speech rate x language. Additionally, we included a three-way tensor product ordered factor difference smooth interaction of time x speech rate x language. The models included random smooths over time by participant, and random reference/difference smooths of time by item grouped by language (this reduces type 1 error rate and increases power, Sóskuthy, 2021). Model residuals were checked for autocorrelation with the *itsadug* package (van Rij et al., 2022), and autocorrelation was present in both models. Therefore, we fit the same parameters as an autoaggressive error model, in which the model estimates parameters with the assumption that neighboring observations are correlated (this is achieved by including an estimate of autocorrelation into the model; see Wieling, 2018). After adjusting this, autocorrelation at lag 1 was <0.1 (see OSF for visualizations). Additionally, the effective degrees of freedom were checked for oversmoothing, and if the value of the basis function (*k* value) was significant ($p < 0.05$; suggesting that the basis dimension was too restricted), it was doubled (Wieling, 2018; see OSF; https://osf.io/4ke82/). If a $p < 0.05$ could not be reached, the model with the highest p-value was chosen. Significance testing was done by checking the model output of both the parametric and smooth factors in the model output and applying a Bonferroni correction to deal with the increased likelihood of Type 1 error from multiple comparisons (4 comparisons were made for each model: .05/4 yields a p-value of .013; see Sóskuthy, 2021). This approach to significance testing is comparable to model fitting using AIC comparisons and does not require likelihood ratio tests (see Sóskuthy, 2021).

### 3.5.2. Divergence Point Analysis (DPA)

To investigate the timing of looks across the L1 and L2 speakers, we conducted a DPA (Stone et al., 2021). DPA is a non-parametric bootstrapping method that allows comparisons between groups using confidence intervals (CI). The DPA uses t-tests to compare fixations on two objects until 10 consecutive 20 ms bins (or 200 ms) are significantly different. By resampling the original data (via boostrapping), 2000 new datasets and their respective divergence points are generated, and then the mean and 95% CI of these divergence points are calculated. Similar to the GAMMs, we made two comparisons using DPA, though the DPA was run over the whole time window. The first analysis investigated the timing of first-stage prediction and compared looks to the target (*guitar*) relative to looks to the verb-related object (*cards*) in the predictable items. The second analysis investigated the timing of second-stage prediction and compared looks to the target (*guitar*) versus looks to the agent-related object (*microphone*) in the predictable items.

## 4. Results

### *4.1. Accuracy*

Incorrectly answered items were removed before analysis. The target was chosen in 94.2% of the items (thus, 5.8% of the items were removed). Of the 5.8% of items that were removed, 3.8% were due to selection of the agent-related object, 1.5% to selection of the verb-related object and 0.5% to selection of the distractor. See Figure 1 for a visualization of the fixation proportion for the correctly answered trials.

#### 4.1.1. First-stage prediction

**4.1.1.1. GAMM.** To investigate first-stage prediction, we tested the empirical logit of looks to the target (*guitar*) versus looks to the verb-related object (*cards*) in the predicable items from the onset of the agent to onset of the target): see Figure S1 in the Supplementary material for a visualization.

The results from the GAMM can be seen in Table 4.

The only significant parameter was the tensor product ordered factor difference smooth interaction between time and speech rate ($F = 16.04$, $p < .0001$). Contour plots were used to visualize this interaction; see Figure 2. The top left plot displays the speech rate x time interaction for the L2 speakers, and the top right plot displays the speech rate x time interaction for the L1 speakers. In these two plots, pink indicates greater empirical logits (greater looks to the target relative to the verb-related object) and green indicates less empirical logits (fewer looks to the target relative to the verb-related object). The bottom left plot displays the difference between the L2 and L1 speakers. The positive values indicate that the L2 speakers have greater empirical logits (greater looks to the target relative to the verb-related object) than the L1 speakers (with the difference increasing the more pink it becomes), while the smaller values (approaching 0) indicate that the L1 speakers have greater empirical logits than the L2 speakers (with the difference increasing the more green it becomes).

What is clear in these graphs is that both L1 and L2 speakers are showing a negative or 0 empirical logit (indicating participants are not looking toward the target) until around 700-800 ms (around the offset of the agent). At this point, the empirical logit becomes positive (indicating participants are looking at the target), suggesting they are making first-stage predictions after hearing the agent. To further aid in the interpretation of the impact of speech rate, we visualized the L2-L1 difference across time at three speeds, 2.6, 3.6 and 4.6 syllables per second (see Figure S2 in the Supplementary material). These values were arbitrarily selected to cover the range of speech rates used in the experiment.

These visualizations suggest that both L1 and L2 speakers are making similar looks to the target across time at the slower (2.6 syllables per second) and middle (3.6 syllables per second) speech rates. However, at the faster (4.6 syllables per second) speech rates L1 speakers are making more looks to the target as soon as they hear the agent (as evidence by the larger empirical logit between 430 and 730 ms) and L2 speakers show an increase of looks to the target later (starting at approximately 1300 ms–to the end of the window) in the difference graph. This suggests that at the faster speech rates, L1 speakers look to the target earlier (during the agent) until the end of the time window, while L2 speakers look to the target later (at the end of the verb), with their looks to the target continuing to increase as the time window unfolds. Overall, both groups show similar patterns of looks to the target at slower and middle speech rates, and L1 speakers make more and earlier looks to the target at faster speech rates, with L2 speakers seeming to "catch up" by showing later and greater looks to the target at the end of the time window.

**4.1.1.2. DPA.** To test the first-stage prediction, we compared looks to the target (*guitar*) versus looks to the verb-related object (*cards*) in the predictable items. The DPA revealed that looks to these objects diverged at 709.25 ms [CI: 620,920] ms for L1 speakers and 766.97 ms [CI: 720,920] for L2 speakers. The difference between the two groups is 57.72 [CI: −140, 240], and given that the CI crosses 0, we do not conclude that L1 and L2 speakers differ in the timing of looks to the target relative to the verb-related object, see Figure 3.

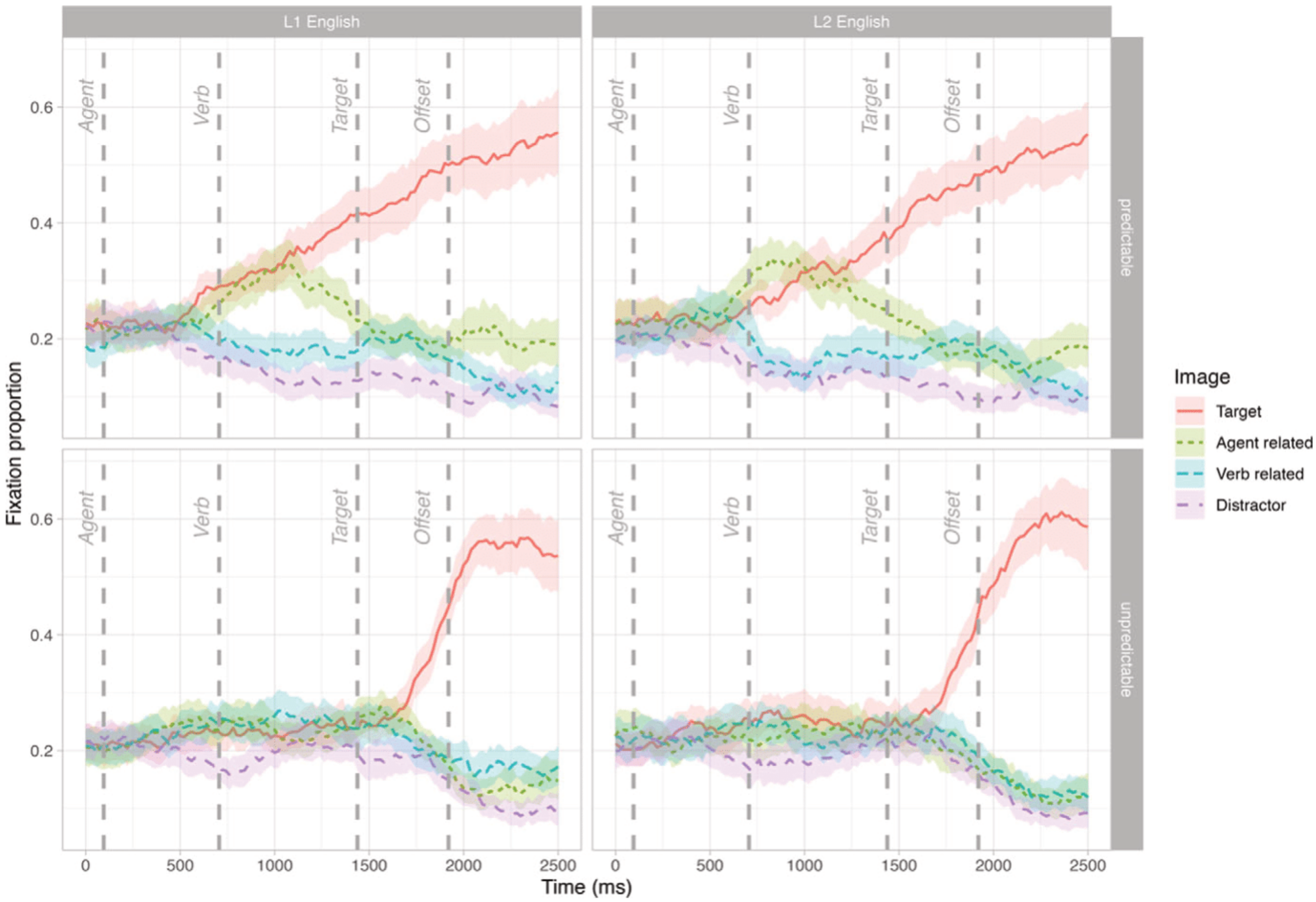



**Figure 1.** Mean fixation proportion to all objects across language and sentence type (correctly answered items only).

**Table 4.** First-stage prediction GAMM output for the empirical logit during the prediction time window. Part A reports the parametric coefficients and Part B reports the smooth terms.

| A. Parametric coefficients | Estimate | Std. error | t-value | p-value |
|---|---|---|---|---|
| Intercept | 0.18 | 0.04 | 4.76 | < 0.0001 |
| Language (L2) | −0.06 | 0.05 | −1.18 | 0.86 |
| **B. Smooth terms** | **EDF** | **Ref.df** | **F-value** | **p-value** |
| s(speech rate) | 1.61 | 1.97 | 19.46 | < 0.0001 |
| s(time) | 1.00 | 1.00 | 0.09 | 0.76 |
| s(time):Language (L2) | 1.00 | 1.00 | 0.33 | 0.57 |
| s(speech rate): Language (L2) | 1.00 | 1.00 | 0.46 | 0.50 |
| ti(time, speech rate): Language (L2) | 8.14 | 8.81 | 16.04 | < 0.0001 |
| s(time, participant) | 683.98 | 817.00 | 10.67 | < 0.0001 |
| s(time, item) | 68.84 | 74.91 | 4.09 | < 0.0001 |
| s(time, item): Language (L2) | 22.02 | 31.89 | 0.95 | 0.56 |

#### 4.1.2. Second-stage prediction

**4.1.2.1. GAMM.** To investigate second-stage prediction, we tested the empirical logit of looks to the target (*guitar*) relative to the agent-related object (*microphone*) in the predictable items from the onset of the verb to the onset of the target, see Figure S3 in the Supplementary material for visualization.

The results from the GAMM can be seen in Table 5.

The ordered factor difference smooth interaction of speech rate by language was significant ($F$ = 48.75, $p$ < .0001). For visualization of this interaction, see Figure S4 in the Supplementary material. The tensor product ordered factor difference smooth interaction between time, speech rate and language was significant ($F$ = 12.04, $p$ < .0001). Contour plots were used to visualize this interaction; see Figure 4. The top left plot displays the speech rate x time interaction for the L2 speakers, and the top right plot displays the speech rate x time interaction for the L1 speakers. In these two plots, pink indicates greater empirical logits (greater looks to the target) and green indicates less empirical logits (less looks to the target). The bottom left plot displays the difference between the L2 and L1 speakers. The positive values indicate that the L2 speakers have greater empirical logits than the L1 speakers, while the smaller values (approaching 0) indicate that the L1 speakers have greater empirical logits than the L2 speakers.

What is clear in these graphs is that L1 and L2 speakers show differences in second-stage prediction. L2 speakers show very little evidence of an empirical logit over 0, with looks to the target increasing over looks to the agent-related object only at the very start and very end of the time window, particularly at the fastest speech rates. Meanwhile, L1 speakers show a general increase in looks to the target as speech rate and time increase. To further aid in

https://doi.org/10.1017/S1366728925100515 Published online by Cambridge University Press

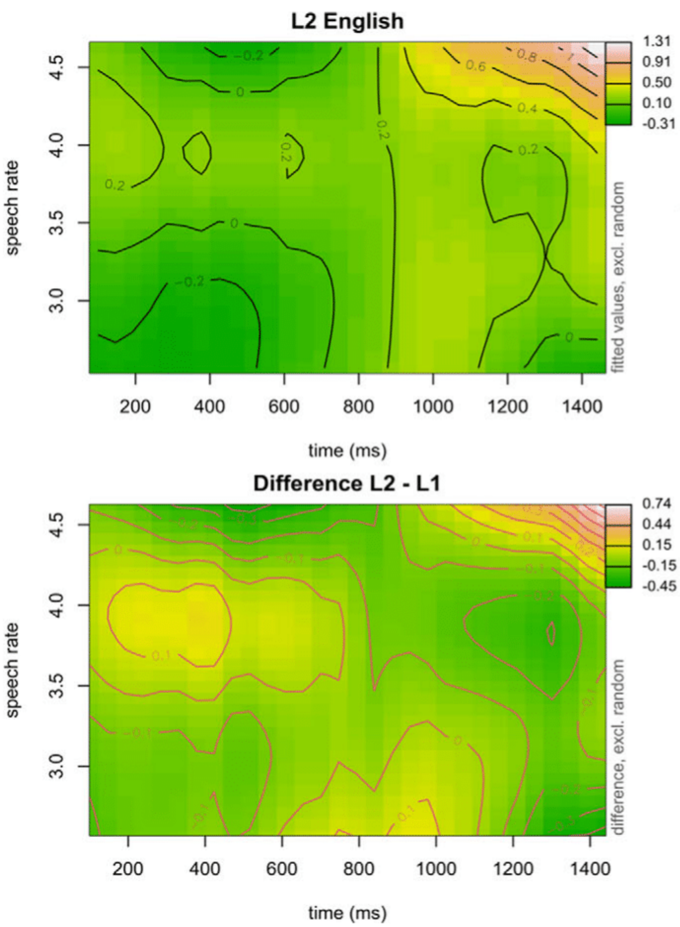



**Figure 2.** Contour plots for the empirical logit of the three-way interaction between time, speech rate and language group. *Top left panel.* Contour plot for L2 speakers. *Top right panel.* Contour plot for L1 speakers. In both of the *top panels*, negative values (i.e., green) indicate lower empirical logits (looks to the target) and positive values (i.e., pink) indicate greater empirical logits. *Bottom left panel.* The difference between the L2 and L1 speakers. In the *bottom panel*, the negative values indicate greater empirical logits (looks to the target) by L1 speakers relative to L2, and positive values indicate greater empirical logits by L2.

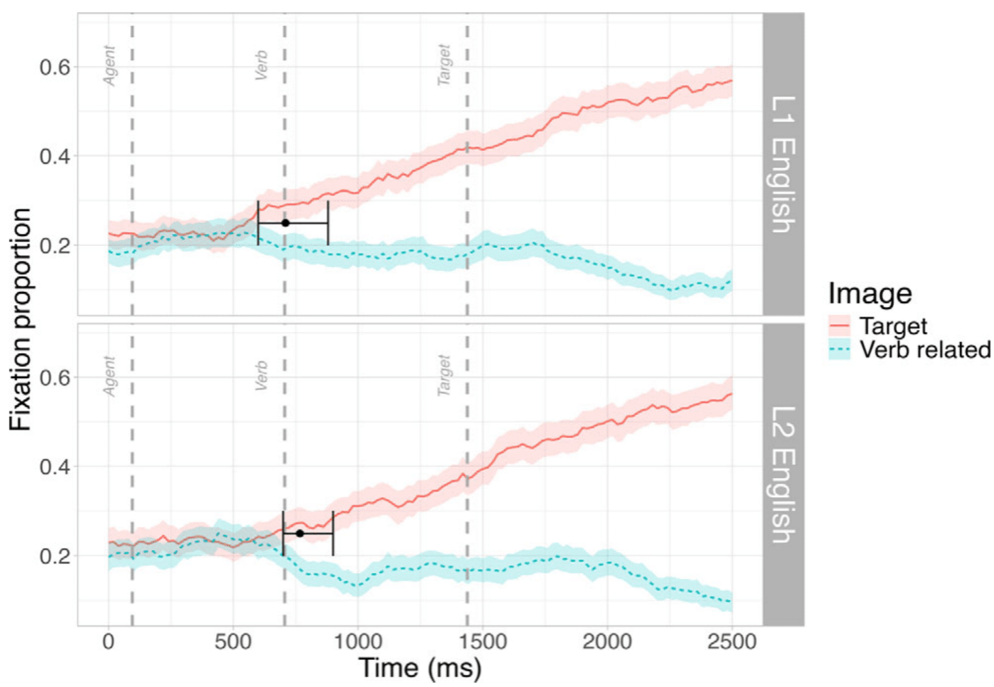



**Figure 3.** Divergence point and 95% confidence intervals superimposed on the fixation proportion of looks to the target and verb-related object.

https://doi.org/10.1017/S1366728925100515 Published online by Cambridge University Press

**Table 5.** Second-stage prediction GAMM output for the empirical logit during the prediction time window. Part A reports the parametric coefficients and Part B reports the smooth terms.

| A. Parametric coefficients | Estimate | Std. error | t-value | p-value |
|---|---|---|---|---|
| Intercept | 0.07 | 0.03 | 2.06 | < 0.05 |
| Language (L2) | −0.07 | 0.05 | −1.53 | 0.13 |
| **B. Smooth terms** | **EDF** | **Ref.df** | **F-value** | **p-value** |
| s(speech rate) | 1.00 | 1.00 | 177.25 | < 0.0001 |
| s(time) | 1.00 | 1.00 | 0.49 | 0.48 |
| s(time):Language (L2) | 1.00 | 1.00 | 0.53 | 0.47 |
| s(speech rate): Language (L2) | 1.00 | 1.00 | 48.75 | < 0.0001 |
| ti(time, speech rate): Language (L2) | 7.15 | 8.35 | 12.04 | < 0.0001 |
| s(time, participant) | 673.00 | 817.00 | 9.47 | < 0.0001 |
| s(time, item) | 19.25 | 27.86 | 0.99 | 0.49 |
| s(time, item): Language (L2) | 23.09 | 32.88 | 0.92 | 0.61 |

interpretation, we visualized the L2-L1 difference across time at three speeds, 2.6, 3.6 and 4.6 syllables per second (see Figure S5 in the Supplementary material).

These visualizations suggest that at the slower speech rates, and earlier in the time window, both groups are showing similar patterns of looks, with L2 speakers making slightly more looks to the target toward the end of the window, as evidenced by the larger values (indicating greater looks by the L2 group) in the bottom right corner of the difference graph. As speech rate increases, L1 speakers make more and earlier looks to the target (at both 3.6 syllables and 4.6 syllables per second) as evidenced by the smaller values (indicating greater looks by the L1 group) in the top right corner of the difference graph. Overall, this suggests that at faster speech rates, L1 speakers look to the target while hearing the verb more than L2 speakers, who show greater and earlier looks to the target at the slower speech rates.

**4.1.2.2. DPA.** To test the second-stage prediction, we compared looks to the target- versus agent-related object in the predictable items. The DPA revealed that looks to these objects diverge at 1299 [CI: 1240, 1420] ms for L1 speakers and 1396 [CI:1320, 1520] ms for L2 speakers. The difference between the two groups is 122.61 [CI: −40, 240]. Given that the CI crosses 0, we do not conclude that L1 and L2 speakers differ in the timing of looks to the verb-related relative to the distractor object; see Figure 5.

## 5. Discussion

It is well established that L1 speakers make predictions while listening to speech (e.g., see Ferreira & Chantavarin, 2018; Huettig,

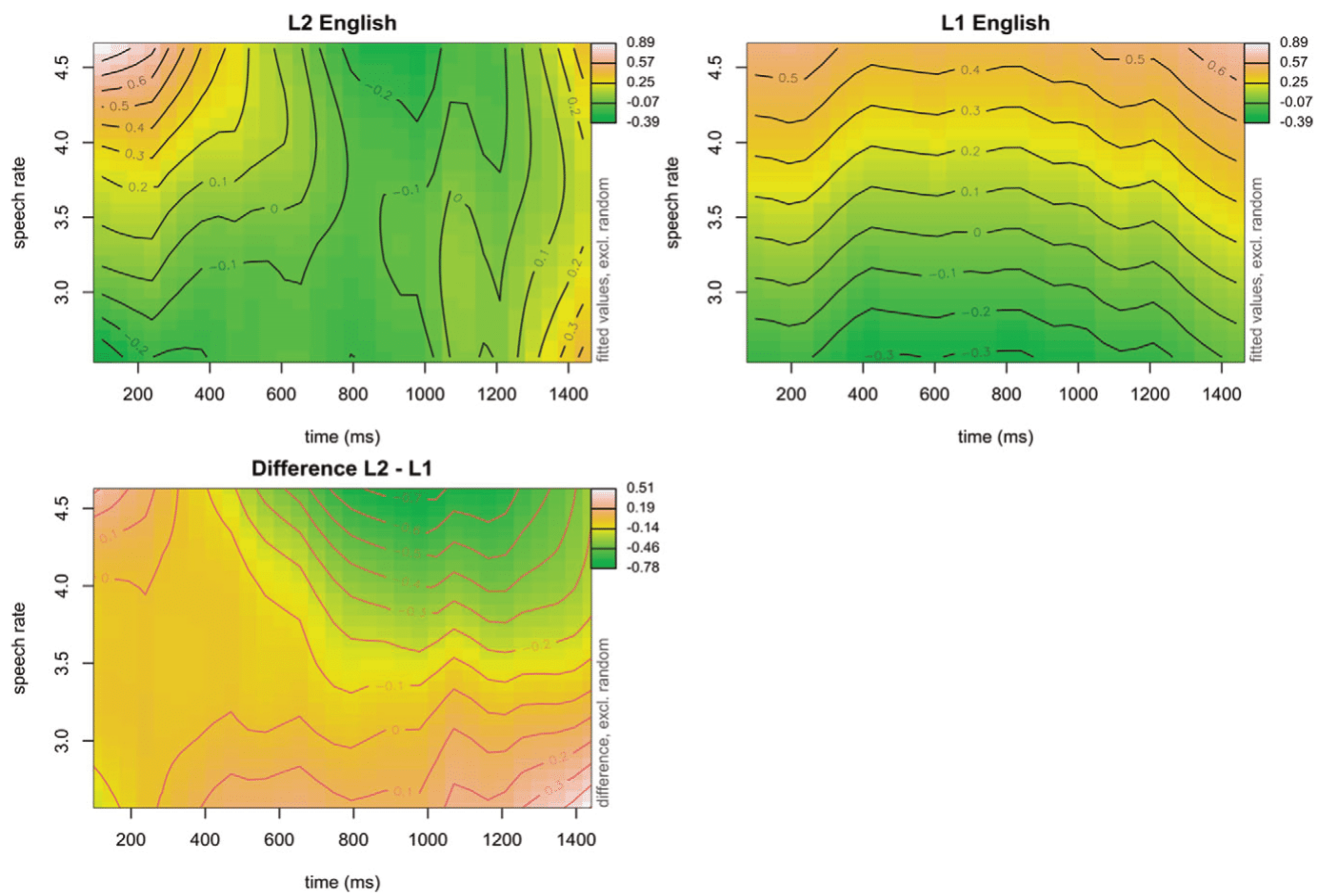



**Figure 4.** Contour plots for the empirical logit of the three-way interaction between time, speech rate and language group. *Top left panel.* Contour plot for L2 speakers. *Top right panel.* Contour plot for L1 speakers. In both of the *top panels*, negative values (i.e., green) indicate lower empirical logits (looks to the target) and positive values (i.e., pink) indicate greater empirical logits. *Bottom left panel.* The difference between the L2 and L1 speakers. In the *bottom panel*, the negative values indicate greater empirical logits (looks to the target) by L1 speakers relative to L2, and positive values indicate greater empirical logits by L2.

https://doi.org/10.1017/S1366728925100515 Published online by Cambridge University Press

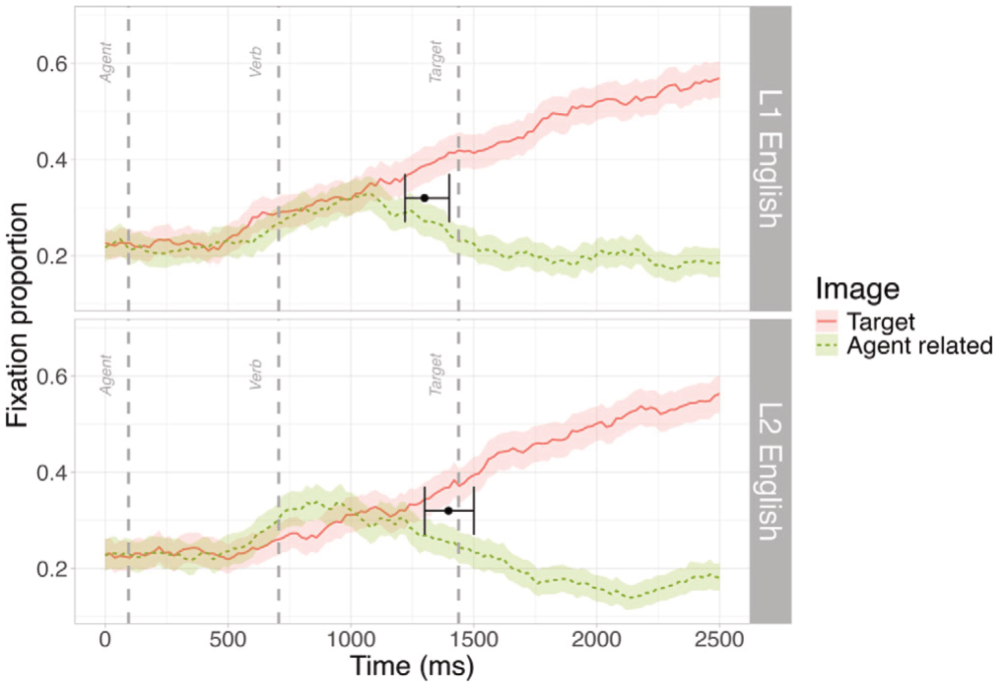



**Figure 5.** Divergence point and 95% confidence intervals superimposed on the fixation proportion of looks to the target and agent-related object.

2015; Huettig & Mani, 2016; Kamide, 2008; Staub, 2015). The general consensus is that L2 speakers are able to make predictions in the same way as L1 speakers, with differences between the groups stemming from individual differences and/or methodological factors (e.g., Hopp, 2022; Kaan & Grüter, 2021). It has also been argued that prediction occurs in two stages: an automatic, relatively cost-free first stage, and a non-automatic, more costly second stage (e.g., Ito & Pickering, 2021; Pickering & Gambi, 2018). Evidence suggests that it might be the late, more costly stage, where L2 speakers show differences relative to L1 speakers (e.g., Corps et al., 2023). However, this may be modulated by factors such as speech rate (Gambi et al., 2016). In the current study, we therefore investigated both first- and second-stage prediction in L1 and L2 speakers of English while taking into account the role of speech rate. We hypothesized that L1 and L2 speakers would make first-stage predictions at similar times, but that L2 speakers would be delayed in making second-stage predictions. In terms of speech rate, we hypothesized that both groups would show an inverse-U pattern of predictions across speech rates, with "optimal" speech rates being slower for L2 speakers than L1 speakers. Further, we hypothesized that the "optimal" rate for L2 speakers in the more costly second stage would be lower than that of the first stage and start decreasing at slower rates.

To test first-stage prediction, we compared looks to the target versus looks to the verb-related object while listening to predictable sentences. We found that both groups quickly looked toward the target around the start of the verb, with the DPA showing no difference in the timing of looks between the two groups. The GAMM revealed that both groups were making first-stage predictions around the verb onset between 700 and 800 ms (consistent with the DPA timings). Both groups showed similar patterns of looks to the target at speech rates between approximately 2.5–4.0 syllables per second. As the speech rate increased above approximately 4.0 syllables per second, both groups continued to show evidence of prediction. However, L1 speakers made first-stage predictions while processing the agent, while L2 speakers made first-stage predictions later, only at the verb offset. This suggests that, for both groups, as speech rate increases, so do looks to the target, which supports previous findings that predictive eye movements increase from slower speech rates (Fernandez et al., 2020). However, in the current study, the "optimal" speech rate for L2 speakers was up to 4.5 syllables per second (unlike Fernandez et al., who found the optimal rate to be around 3.5 syllables per second). This may be due to the type of syntactic prediction tested in Fernandez et al. (2020) being more costly than the first-stage prediction in the current study, thus leading to a faster "optimal" speech rate.

To test second-stage prediction, we compared looks to the target versus looks to the agent-related object while listening to predictable sentences. Given that the second stage is more cognitively demanding and subject to more conscious control, it may be that when processing a more cognitively demanding L2, there are fewer resources to invest in making second-stage predictions. We found that both groups showed competition between the target and agent-related object and that, soon after hearing the verb (and before the target is spoken), both groups predictively look to the appropriate target. Surprisingly, while the divergence point of the two groups was numerically different (1300 ms for the L1 speakers, corresponding to approximately the onset of the article, versus 1400 ms for the L2 speakers, corresponding to approximately the end of the article prior to the target), the DPA did not reveal the groups to differ. In terms of speech rate, both groups interestingly show an early preference for the target at the fastest speech rates, which indicates looks to the target during the presentation of the agent, at least when it is presented quickly. The L1 speakers showed a very clear increase in prediction as speech rate increased, with looks to target incrementally increasing with speech rate. Starting at approximately 4 syllables per

https://doi.org/10.1017/S1366728925100515 Published online by Cambridge University Press

second, the L1 group consistently (and increasingly) made predictions, which may indicate that their optimal rate is 4.6 syllables per second or higher. This finding supports Fernandez et al. (2020), that when speech rates are too slow, prediction is less efficient. The L2 group showed very little early evidence of disambiguating the target from the agent-related image, and did not show an empirical logit above zero (indicating a preference for the target) until approximately 1300 ms across all speech rates. Overall, L2 speakers did not show a clear pattern with any optimal speech rate. While the DPA showed that L2 speakers looked to the target before it was spoken, it may be that L2 speakers adopted a "wait and see" strategy (e.g., McMurray et al., 2017; Van Petten & Luka, 2012) in which they wait for more of the sentence to unfold, thus accruing more information, before making a prediction. This strategy may reduce processing costs by limiting the set of active alternative interpretations that then need to be inhibited. However, it is important to note that we did not actively manipulate speech rate, so future research designed to more directly test the impact of speech rate could help elucidate this potential further.

We believe that these data support previous research demonstrating that both L1 and L2 speakers make first- and second-stage predictions (e.g., Corps et al., 2022, 2023; Ito & Pickering, 2021; Pickering & Gambi, 2018). When it comes to first-stage predictions, both L1 and L2 speakers showed similar patterns of looks to the target until approximately 4 syllables per second. At faster speeds, both groups continued to show increased looks to the target; however, L1 made showed earlier looks to the target (during the agent) than L2 speakers (after the verb). When it comes to second-stage prediction, there were no timing differences between the groups, but the impact of speech rate on these stages shows the more nuanced difficulties that L2 speakers face in more cognitively demanding linguistic situations. Particularly, L2 speakers showed reduced and more variable predictions, unlike L1 speakers, who showed increased prediction as speech rate increased. That is, when prediction is more cognitively demanding, L1 speakers are able to predict more consistently, even as speech rate increases, while L2 speakers, who have less cognitive resources available, show less, later and more variable prediction. This may reflect a wait-and-see strategy in which L2 speakers wait for more information before committing to a prediction to potentially reduce the costs that accompany second-stage prediction.

As mentioned, and unlike Corps et al. (2023), we did not see overall DPA differences in second-stage prediction between L1 and L2 speakers. We believe there are at least three reasons why the timing differences did not arise. First, the second-stage prediction in our items relied on combining verb and agent information, while Corps et al. (2022, 2023) relied on stereotypy. It is possible that the latter requires more cognitive resources, and thus, there is more opportunity to see these second-stage differences. Second, there is also a temporal gap between activating competitors and ruling them out in the current study, while in the Corps et al. study, there is no such temporal gap and both activation of relevant objects and narrowing down of those objects can occur simultaneously at the verb. It may be that the simultaneous activating and narrowing is more cognitively demanding, again making second-stage differences more apparent. Third, Corps et al. (2023) recruited a heterogeneous group of L2 speakers, while the current study recruited only L1 speakers of German. It is possible that similarities between English and German made the second-stage prediction less costly.

This raises another aspect we believe is worth discussing: the potential influence of proficiency. The impact of proficiency on prediction is not entirely clear, with some researchers finding increased L2 prediction abilities with higher proficiency (e.g., Chambers & Cooke, 2009; Dussias et al., 2013; Hopp, 2013) while other research finds no relationship between proficiency and predictive abilities (Hopp, 2015; Ito et al., 2018; Kaan & Grüter, 2021; Kim & Grüter, 2021; Perdomo & Kaan, 2021). While we collected a measure of proficiency in the current study, we did not include it in our models because (1) it was highly correlated with speaker group and (2) we compared a model with and without proficiency score and the inclusion of proficiency did not improve model fit (see OSF for correlation and model comparison). While the exact contributions of proficiency aren't clear, what is clear is that speakers with the proficiency level in the current study (i.e., intermediate to high proficiency) are capable of making semantic predictions at multiple speech rates.

## 6. Conclusions

In this study, we investigated first- and second-stage prediction in L1 and L2 speakers of English while exploring the role of speech rate. We found evidence that both groups make first- and second-stage predictions at similar times. However, speech rate played an important role in both stages. During the first stage, both L1 and L2 speakers showed similar patterns of looks to the target at slower speech rates. At faster speech rates, however, L1 speakers showed earlier predictions than L2 speakers. This suggests that L2 speakers may make slightly later first-stage predictions (relative to L1 speakers) at faster speech rates. Additionally, L2 speakers seem to have an "optimal" speech rate of up to 4.5 syllables per second (faster than found by Fernandez et al., 2020) in this stage. During the second stage, L1 speakers showed a clear increase in prediction as speech rate increased, while L2 speakers showed a later, reduced and more variable pattern relative to L1 speakers. This may reflect a wait-and-see approach. Particularly, when prediction is more costly, L2 speakers may wait for more information before committing to a prediction. This supports the literature that second-stage prediction is more costly and has a greater impact on speakers with less available cognitive resources, and highlights the importance of choosing an appropriate speech rate in VWP research.

**Supplementary material.** The supplementary material for this article can be found at http://doi.org/10.1017/S1366728925100515.

**Data availability statement.** The data and statistical code that support the findings of this study are openly available in OSF at https://osf.io/4ke82/.

## References

**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.

**Audacity Team**. (2021). Audacity (Version 3.0) [Computer software]. https://www.audacityteam.org/.

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

**Boersma, P.**, & **Weenink, D.** (2021). *Praat: Doing phonetics by computer [Computer program]* (Version 6.1.44) [Computer software]. Retrieved from https://www.praat.org

**Buckner, R. L.** (2004). Memory and Executive Function in Aging and AD: Multiple Factors that Cause Decline and Reserve Factors that Compensate. *Neuron*, **44**(1), 195–208. https://doi.org/10.1016/j.neuron.2004.09.006

**Chambers, C. G.**, & **Cooke, H.** (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **35**(4), 1029–1040. https://doi.org/10.1037/a0015901.

**Corps, R. E.**, **Brooke, C.**, & **Pickering, M. J.** (2022). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language*, **122**. https://doi.org/10.1016/j.jml.2021.104298.

**Corps, R. E.**, **Liao, M.**, & **Pickering, M. J.** (2023). Evidence for two stages of prediction in non-native speakers: A visual-world eye-tracking study. *Bilingualism: Language and Cognition*, **26**(1), 231–243. https://doi.org/10.1017/S1366728922000499.

**Cosgrove, A. L.**, **Beaty, R. E.**, **Diaz, M. T.**, & **Kenett, Y. N.** (2023). Age differences in semantic network structure: Acquiring knowledge shapes semantic memory. *Psychology and Aging*, **38**(2), 87–102. https://doi.org/10.1037/pag0000721.

**Cosgrove, A. L.**, **Kenett, Y. N.**, **Beaty, R. E.**, & **Diaz, M. T.** (2021). Quantifying flexibility in thought: The resiliency of semantic networks differs across the lifespan. *Cognition*, **211**, 104631. https://doi.org/10.1016/j.cognition.2021.104631.

**Dussias, P. E.**, **Valdeś, J. R. K.**, **Guzzardo, R. E. T.**, & **Gerfen, C.** (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, **35**, 353–387. https://doi.org/10.1017/S0272263112000915.

**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.**, **Shehzad, M.**, & **Hadley, L. V.** (2025). Younger adults may be faster at making semantic predictions, but older adults are more efficient. *Psychology and Aging*, **40**(3), 318–375. https://doi.org/10.1037/pag0000872.

**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.

**Fjell, A. M.**, **Sneve, M. H.**, **Grydeland, H.**, **Storsve, A. B.**, & **Walhovd, K. B.** (2017). The Disconnected Brain and Executive Function Decline in Aging. *Cerebral Cortex*, **27**(3), 2303–2317. https://doi.org/10.1093/cercor/bhw082

**Gambi, C.**, **Pickering, M. J.**, & **Rabagliati, H.** (2016). Beyond associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, **157**, 340–351. https://doi.org/10.1016/j.cognition.2016.10.003.

**Grüter, T.**, **Lew-Williams, C.**, & **Fernald, A.** (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, **28**(2), 191–215. https://doi.org/10.1177/0267658312437990.

**Hasher, L.**, & **Zacks, R. T.** (1988). Working Memory, Comprehension, and Aging: A Review and a New View. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. **22**, pp. 193–225). Academic Press. https://doi.org/10.1016/S0079-7421(08)60041-9

**Holt, R.**, **Bruggeman, L.**, & **Demuth, K.** (2021). Children with hearing loss can predict during sentence processing. *Cognition*, **212**, 104684. https://doi.org/10.1016/j.cognition.2021.104684

**Hopp, H.** (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, **29**, 33–56. https://doi.org/10.1177/0267658312461803.

**Hopp, H.** (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, **53**(3). 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.**, & **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.**, & **Mani, N.** (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, **31**(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223.

**Ito, A.**, & **Pickering, M. J.** (2021). Automaticity and prediction in non-native language comprehension. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 25–46). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.02ito.

**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.

**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.

**Kaan, E.**, & **Grüter, T.** (2021). Prediction in second language processing and learning: Advances and directions. In E. Kaan & T. Grüter (Eds.), *Prediction in second-language processing and learning* (pp. 1–24). John Benjamins Publishing Company.

**Kamide, Y.** (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, **2**(4), 647–670. https://doi.org/10.1111/j.1749-818X.2008.00072.x.

**Kim, H.**, & **Grüter, T.** (2021). Predictive processing of implicit causality in a second language: A visual-world eye-tracking study. *Studies in Second Language Acquisition*, **43**(1), 133–154. https://doi.org/10.1017/S0272263120000443.

**Kukona, A.** (2023). Predictive sentence processing at speed: Evidence from online mouse cursor tracking. *Cognitive Science*, **47**(4). https://doi.org/10.1111/cogs.13285.

**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. https://doi.org/10.1016/j.cognition.2010.12.002.

**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. https://doi.org/10.1016/j.jml.2010.07.003.

**Martin, C. D.**, **Thierry, G.**, **Kuipers, J.-R.**, **Boutonnet, B.**, **Foucart, A.**, & **Costa, A.** (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, **69**(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001.

**McMurray, B.**, **Farris-Trimble, A.**, & **Rigler, H.** (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition*, **169**, 147–164. https://doi.org/10.1016/j.cognition.2017.08.013.

**Müller, J. A.**, **Wendt, D.**, **Kollmeier, B.**, **Debener, S.**, & **Brand, T.** (2019). Effect of speech rate on neural tracking of speech. *Frontiers in Psychology*, **10**, 449.

**Perdomo, M.**, & **Kaan, E.** (2021). Prosodic cues in second-language speech processing: A visual world eye-tracking study. *Second Language Research*, **37**(2), 349–375. https://doi.org/10.1177/0267658319879196.

**Peters, R. E.**, **Grüter, T.**, & **Borovsky, A.** (2018). Vocabulary size and native speaker self-identification influence flexibility in linguistic prediction among adult bilinguals. *Applied PsychoLinguistics*, **39**(6), 1439–1469. https://doi.org/10.1017/S0142716418000383.

**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.

**Porretta, V.**, **Kyröläinen, A.-J.**, **van Rij, J.**, & **Järvikivi, J.** (2018). Visual world paradigm data: From Preprocessing to nonlinear time-course analysis. In Czarnowski, R. J. Howlett, & L. C. Jain (Eds.), *Intelligent decision technologies 2017* (pp. 268–277). Springer International Publishing.

**Prystauka, Y.**, **Altmann, G. T. M.**, & **Rothman, J.** (2024). Online eye tracking and real-time sentence processing: On opportunities and efficacy for capturing psycholinguistic effects of different magnitudes and diversity. *Behavior Research Methods*, **56**(4), 3504–3522. https://doi.org/10.3758/s13428-023-02176-4.

**R Core Team**. (2018). R: A language environment for statistical computing *[computer software]*. R foundation for statistical computing. https://www.R-project.org/

**Segalowitz, N.**, & **Hulstijn, J. H.** (2009). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University Press.

**Sóskuthy, M.** (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, **84**, 101017. https://doi.org/10.1016/j.wocn.2020.101017

**Staub, A.** (2015). The effect of lexical predictability on eye movements in Reading. *Critical Review and Theoretical Interpretation. Language and Linguistics Compass*, **9**(8), 311–327. https://doi.org/10.1111/lnc3.12151.

**Stone, K.**, **Lago, S.**, & **Schad, D. J.** (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, **24**(5), 833–841Placeholder Text. https://doi.org/10.1017/S1366728920000607.

**van Heuven, W. J. B.**, **Mandera, P.**, **Keuleers, E.**, & **Brysbaert, M.** (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, **67**(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521.

**Van Petten, C.**, & **Luka, B. J.** (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *Predictive information processing in the brain: Principles. Neural Mechanisms and Models*, **83**(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015.

**van Rij, J.**, **Wieling, M.**, **Baayen, H.**, & **van Rijn, H.** (2022). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* (Version 2.4.1) [Computer software]. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=itsadug

**Veríssimo, J.**, **Verhaeghen, P.**, **Goldman, N.**, **Weinstein, M.**, & **Ullman, M. T.** (2022). Evidence that ageing yields improvements as well as declines across attention and executive functions. *Nature Human Behaviour*, **6**(1), 97–110. https://doi.org/10.1038/s41562-021-01169-7.

**Wieling, M.** (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, **70**, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002.

**Wood, S. N.** (2017). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.

# CHAPTER 4. DISCUSSION

The goal of this body of work was to examine the effects of increasing cognitive demand on spoken language predictive processing. Testing the limits of predictive processing can help to inform the theory behind the biological architecture and cognitive mechanisms of prediction. The papers included in this dissertation focused on addressing two existing gaps in the literature: (1) the direct manipulation of cognitive demand during a prediction task as opposed to correlating performance on a prediction task and a secondary task and (2) the role of speech rate, which is often un(der)reported and atypically slow. The studies presented here address the following research questions, each of which will be discussed in turn.

RQ1. Do L2 speakers predict despite a variety of challenging situations?

- RQ1.1 Do L2 speakers need a "slow" speech rate to predict?
- RQ1.2 Do L2 speakers show consistent evidence of prediction with increased demands across stages of prediction and different measurements of prediction?

RQ2. Does increasing the cognitive demand of a prediction task reduce prediction?

- RQ2.1: Does a concurrent visuospatial load reduce predictive gaze behavior?
- RQ2.2: Does a concurrent visuospatial load reduce target pre-activation in the absence of a visual array?
- RQ2.3: Are first and second stage predictions differentially affected by increased speech rate?

## Research Question 1

RQ1. Do L2 speakers predict despite a variety of challenging situations?

- RQ1.1 Do L2 speakers need a "slow" speech rate to predict?
- RQ1.2 Do L2 speakers show consistent evidence of prediction with increased demands across stages of prediction and different measurements of prediction?

The broad answer to RQ1 is resoundingly "yes." Across all three studies, which comprised two different measurement methodologies, concurrent visuospatial working memory demands, L2 language processing, and normal or fast speech rates, participants showed strong evidence of predictive processing. Such robust findings of predictive processing, in my opinion, suggest a rather ubiquitous role of prediction in spoken language processing. Some theoretical frameworks for prediction suggest that prediction is an obligatory aspect of processing (e.g., Dell & Chang, 2014), while some frameworks suggest that at least some aspects of prediction are optional (e.g., Huettig & Mani, 2016; Pickering & Gambi, 2018). One limitation of this interpretation was the use of highly predictable stimuli across the included studies. That is, the predictable sentences were quite predictable and such predictable sentences are relatively uncommon in naturalistic language processing. Thus, predictions could theoretically only be ubiquitous when the sentences are highly constraining. However, given the evidence that there is no particular cost associated with generating a prediction that is not satisfied (e.g., Fernandez et al., 2025, Frisson et al., 2017), it seems likely that at least feature based semantic pre-activation is a pervasive aspect of spoken language processing.

Regarding RQ1.1 (Do L2 speakers need a "slow" speech rate to predict?), the answer is quite clearly "no." In fact, these results suggest that slower speech rates might even be a hinderance (and especially so for L1 speakers). Study 3 found that L2 speakers had, for at least one stage of prediction, an optimal speech rate nearing 4.5 syllables per second. As mentioned in section 2.2.2, this is nearly double the speed of many commonly reported speech rates (if they are at all reported).

Across all three studies, L2 participants showed robust predictions at normal speech rates. Intuitively, given the theoretical framework of the predictive brain, this makes sense. If language exposure is the basis upon which our statistical regularities are formed and thus the basis for predictions, then it makes sense that predictions are optimal at the most commonly encountered speech rate: normal. Changing the speech rate, especially to one that sounds abnormally slow, changes the context of language processing and may thus influence predictive behavior. On the other hand, increasing the speech rate, especially above certain thresholds, increases extra demands on predictive processing. In faster speech conditions, participants are limited in both processing (there is simply less time) and working memory (there is more information per unit of time) and both of these factors can play important roles in predictive processing.

Regarding RQ1.2 (Do L2 speakers show consistent evidence of prediction with increased demands across stages of prediction and different measurements of prediction?) the answer is also "yes." While several manipulations reduced predictive behavior (a concurrent visuospatial working memory demand in Study 1, faster speech rates in study 3), none of these manipulations eliminated predictions. This finding was consistent across methodologies (eye-tracking and EEG) and across stages of prediction (stage 1 and stage 2).

While previous research has shown that L2 speakers can reliably use semantic information as a predictive cue (e.g., Chambers & Cooke, 2009; Hopp 2015), this has only been examined with a concurrent demand in one previous study (Ito et al., 2018). Furthermore, the manipulation used in the Ito et al. (2018) study nearly eliminated L2 gaze behavior, with significant evidence for prediction only occurring after target word onset. Thus, these are the first studies to show strong evidence that L2 speakers can predict with extra demands.

Similarly, these studies provide some of the first evidence for prediction despite normal speech rates. One of the few studies to manipulate speech rate was by Huettig & Guerra (2019) and they found that, with 1 second previews of the visual array in a visual world eye-tracking study, L1 participants did not show evidence of predictive gaze behavior. Instead, participants only showed evidence for prediction with explicit instruction (i.e., being told to predict the target of the sentence) or with a long preview time (4 seconds). Each of the studies here used a 2 second preview time for the visual array and each showed evidence for L2 predictive gaze behavior at normal speech rates.

Taken together, these studies show that semantically guided predictive gaze behavior is robust despite a variety of additional demands, including L2 processing, concurrent visuospatial working memory demands, and normal speech rates. Whereas previous research (Huettig & Guerra, 2019; Ito et al., 2018) suggests that normal speech rate and concurrent working memory demands more or less eliminate predictive (gaze) behavior, these studies show that prediction is in fact possible. Thus, it is important to keep in mind the complexity of prediction; there are multiple mechanisms involved, multiple pathways to prediction, and multiple interactions between them. Disruptions in prediction could happen at any number of these levels. It is further important to consider the limitations of gaze behavior as a dependent variable or the effect of concurrent language/vision interactions. The combined results of Study 1 and 2 suggest that some effects may be contingent on either the measurement of gaze behavior or the combination of language and vision.

## Research Question 2

RQ2. Does increasing the cognitive demand of a prediction task reduce prediction?

- RQ2.1: Does a concurrent visuospatial load reduce predictive gaze behavior?

- RQ2.2: Does a concurrent visuospatial load reduce target pre-activation in the absence of a visual array?
- RQ2.3: Are first and second stage predictions differentially affected by increased speech rate?

The broad answer to RQ2 (Does increasing the cognitive demand of a prediction task reduce prediction?) is more "it depends." The relationship between cognitive demand and prediction seems to depend both on the cognitively demanding task and on the type of prediction. Again, this highlights the complexity of prediction and the fact that multiple interacting mechanisms are involved.

Regarding RQ2.1 (Does a concurrent visuospatial load reduce predictive gaze behavior?), the answer is "yes." Performing a visual world prediction task with a concurrent visuospatial working memory demand does reduce predictive gaze behavior. This can (and has, see Wang et al., 2025) been interpreted as an influence of domain general working memory on prediction and thus as evidence for the involvement of domain general mechanisms in predictive language processing.

However, this finding is also interpretable as a specific interference unique to language-vision interactions (like in the VWP). In this case, the reduction in predictive gaze behavior may instead be due to disruptions in gaze behavior and not necessarily in prediction. For example, the current arbitrary location of an object in a visual array is temporarily encoded and may be linked with the preexisting representational object knowledge that is activated by currently incoming auditory input. Thus, increases in visuospatial working memory may disrupt this short-term nexus of information formed between the visuospatial index of an object and the currently activated stored representational knowledge (see Huettig, Olivers, et al. 2011).

The two possible interpretations of this finding was the main motivation of study 2 of this dissertation. Study 2 is essentially a replication of study 1 without a concurrent visual array. By presenting similar participants with the same stimuli as used in study 1 and by measuring a neural response instead of tracking gaze behavior, we were able to observe evidence for prediction free from the confound of extra visual information.

RQ2.2 (Does a concurrent visuospatial load reduce target pre-activation in the absence of a visual array?) was aimed at providing direction towards the effect of visuospatial working memory being domain-general or contingent on language/vision interactions. The answer to this question was "no." While we did find evidence that L2 speakers pre-activate semantic features of a target word, this effect was not modulated by a concurrent visuospatial working memory task. We found evidence against the involvement of the visuospatial working memory task. That is, participants equally pre-activated the semantic features of a predictable target word both with and without a concurrent visuospatial working memory demand. This finding suggests that the pre-activation of a target word or features is resilient to the effects of a concurrent visuospatial working memory task.

Crucially, this finding has strong implications for study 1, which found that a concurrent visuospatial working memory task *did* reduce predictive gaze behavior. These combined findings suggest that the effects of visuospatial working memory on prediction are limited to language-vision contexts. Some previous evidence from Liu et al. (2022) has already shown that a considerable number of participants (25%) performing a challenging task during a visual world trial (in this case, simultaneous interpreting) showed no gaze behavior towards any displayed items in a visual world trial. The authors question whether this subset of participants who did not move their eyes also did not predict. I suggest, given the findings here, that semantic pre-activation of the target word is likely still occurring even if the eyes do not move. Increased working memory demands may interfere with the linking between pre-activated representational knowledge with the current arbitrary location of the object which leads to a lack of guidance for eye movements.

Regarding RQ2.3 (Are first- and second- stage predictions differentially affected by increased speech rate?) the answer is “yes.” For first stage predictions (i.e., for predictions that could be driven purely by spreading activation/associative mechanisms), L1 and L2 speakers showed strong and similar predictive gaze behavior patterns from 2.5 to 4.0 syllables per second. In this speech rate, predictive looks *increased* as speech rate increased, again indicating that atypically slow speech rates may lead to unrepresentative eye movements. Above 4.0 syllables per second, L2 speakers still showed predictive behavior, but it was slightly delayed in comparison to L1 speakers. For second stage predictions (i.e., a prediction that required the combination of the properties of both the agent and the verb), L1 speakers continued to show a similar pattern of increasing predictive looks as speech rate increased, up to around 4.6 syllables per second. L2 speakers, however, showed no clear pattern or an optimal speech rate. Thus, while they still showed evidence of second stage predictions, increased speech rate was uniquely disruptive to this more complex, demanding stage of prediction.

Overall, the answers to RQ2 show that predictions that can be driven mostly by spreading activation are resilient to concurrent visuospatial working memory demands and to increased speech rates in non-native language processing. However, predictions are driven by multiple mechanisms. Over these three studies, I have provided evidence for three possible mechanisms involved in providing evidence for predictive language processing: (1) a mechanism that may be driven primarily by spreading activation that is resilient to a concurrent visuospatial working memory and increases in speech rate, (2) a mechanism required for language-vision interactions that is specifically affected by increased visuospatial working memory demand, and (3) a combinatory mechanism that is specifically affected by increased speech rate. It is also possible that (2) and (3) are both a single combinatory mechanism and that combining language/vision representations and combining multiple semantic feature representations are mechanistically similar. If further research shows that these two are impacted by similar limitations, that would provide evidence that there may be a single mechanism responsible for combining multiple streams of information and for combining multiple active representations.

## Implications

The studies presented in this dissertation have several important implications for predictive language processing.

Methodologically, these studies show that speech rates used in the presentation of auditory stimuli need to be considered more carefully. Not only are L2 English speakers capable of semantic prediction at normal speech rates (Study 1, Study 2), but too slow of speech rates actually lead to lower and less representative predictive gaze behaviors (Study 3). Studies on L2 semantic prediction are still presenting participants with slow speech rates (e.g., Ito et al., 2025) and findings from such studies may be less generalizable or less comparable to studies with normal speech rates. If we as a field wish to make claims that prediction is a ubiquitous aspect of language processing, efforts should be made to show the effects in increasingly more normal or naturalistic language processing scenarios.

The findings of studies 1 and 2 also show some of the methodological limitations of the visual world paradigm and the strength of comparing similar stimuli across methodologies. Specifically, the VWP requires eye movements as evidence for prediction and either eye-movements may be disrupted or the formation of the prediction may be disrupted. Either disruption leads to a similar pattern of gaze behavior that can be difficult to disentangle. In study 2 (EEG/N400), we found that the same stimuli used in study 1 were equally predictable with or without a concurrent visuospatial working memory task. This suggests that, in study 1 (eye-tracking), at least the same set of features were pre-activated (and possibly even more strongly, given that representations may also be activated by seeing the visual array). By examining prediction in the same stimuli during both an eye-tracking and an EEG study, we can see that gaze behavior is being driven by more than just the extent to which a word is pre-activated; there must be a specific mechanism beyond the semantic pre-activation of a word (as evidenced by the N400) that is affected by visuospatial working memory. Linking hypotheses for connecting VWP measures to theoretical and neurobiological models of language processing (see e.g. Magnuson, 2019) must consider the nature of this extra mechanism and further research could uncover more of the nature of this mechanism. While I have previously discussed that this mechanism may be one that temporarily stores a link between the current arbitrary location of a word with its active mental representations, this is not the only possibility. For example, mental model theories of language processing (e.g., Altmann & Ekves, 2019) suggest that an event representation is encoded with an inherent aspect of space ("spatial relations between protagonists, and between events in the situation model"). Increases in visuospatial working memory could thus interfere with the mental representation of "space" that is used to guide prediction. It could also be the case that, due to the nature of the Corsi block task used in this study, participants simply run into a visuospatial working memory limitation. In VWP experiments, participants inherently activate some spatially indexed visuo-semantic features of the objects in a display (e.g., Apfelbaum et al., 2021). These four spatially indexed objects plus the four squares they were tasked to remember in the Corsi task leads to participants potentially encoding (or attempting to encode) the location of 8 objects, which participants may not be able to reliably do (e.g., Miller, 1956). Further elucidation of exactly what mechanism a concurrent visuospatial working memory demand affects can meaningfully inform theories of predictive language processing.

## Conclusion

The three studies presented in this dissertation are among the few to directly test the effects of increased demand on predictive language processing. These studies provide evidence that semantic predictions are robust even with the concurrent demands of L2 processing, normal speech rates, and concurrent visuospatial demands. While a concurrent visuospatial working memory demand or faster speech rates may disrupt semantic predictions, these effects seemingly only occur during more complex, combinatory processes such as language/vision interaction or the narrowing down of multiple active representations. Further research could usefully address the exact nature of the combinatory mechanism(s) involved in predictive language processing.

## REFERENCES FOR FRAME TEXT


Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817–840. https://doi.org/10.1037/rev0000154

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

Angulo-Chavira, A. Q., Castellón-Flores, A. M., López-Santillán, H., & Arias-Trejo, N. (n.d.). *Phono-semantic prediction during language comprehension: Effects of working memory*.

Apfelbaum, K. S., Klein-Packard, J., & McMurray, B. (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm. *Journal of Memory and Language*, *121*, 104279. https://doi.org/10.1016/j.jml.2021.104279

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Bargh, J. A. (1994). The four horsemen of automaticity; Awareness, intention, efficiency, and control in social cognition. *Handbook of Social Cognition*, *1*, 1–40.

Bever, T. G. (1970). The cognitive basis for linguistic structures. *Cognition and the Development of Language.*

Cai, Z., Haslett, D., Duan, X., Wang, S., & Pickering, M. (2023). *Does ChatGPT resemble humans in language use?* https://doi.org/10.31234/osf.io/s49qv

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1029–1040. https://doi.org/10.1037/a0015901

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. https://doi.org/10.1037/11571-006

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Corps, R. E., Brooke, C., & Pickering, M. J. (2021). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language*, *122*, 104298. https://doi.org/10.1016/j.jml.2021.104298

Corps, R. E., Liao, M., & Pickering, M. J. (2022). Evidence for two stages of prediction in non-native speakers: A visual-world eye-tracking study. *Bilingualism: Language and Cognition*, *26*(1), 231–243. https://doi.org/10.1017/S1366728922000499

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

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. https://doi.org/10.3758/BF03214546

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 B: Biological Sciences*, *369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

DeLong, K. A., Urbach, T. P., & Kutas, M. (2017). Is there a replication crisis? Perhaps. Is this an example? No: a commentary on Ito, Martin, and Nieuwland (2016). Language, Cognition and Neuroscience, 32(8), 966–973. https://doi.org/10.1080/23273798.2017.1279339

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930. https://doi.org/10.1017/S1366728916000547

Ding, J., Zhang, Y., Liang, P., & Li, X. (2023). Modulation of working memory capacity on predictive processing during language comprehension. *Language, Cognition and Neuroscience*, *38*(8), 1133–1152. https://doi.org/10.1080/23273798.2023.2212819

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Federmeier, K. D., & Kutas, M. (2019). What's "left"? Hemispheric sensitivity to predictability and congruity during sentence reading by older adults. *Neuropsychologia*, *133*, 107173. https://doi.org/10.1016/j.neuropsychologia.2019.107173

Fernandez, L. B., Shehzad, M., & Hadley, L. V. (2025, September 4-6). *Does prediction require executive resources?* [Poster Presentation]. Architectures and Mechanisms for Language Processing. Prague, Czech Republic.

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. (2025). Is there a cost when predictions are not met? A VWP study investigating L1 and L2 speakers. *Quarterly Journal of Experimental Psychology*, *78*(7), 1237–1259. https://doi.org/10.1177/17470218241270200

Fernandez, L. B., Shehzad, M., & Hadley, L. V. (2025). Effects of Hearing Loss on Semantic Prediction: Delayed Prediction for Intelligible Speech When Listening Is Demanding. *Ear and Hearing*, *46*(6), 1440. https://doi.org/10.1097/AUD.0000000000001679

Fónagy, I., & Magdics, K. (1960). Speed of utterance in phrases of different lengths. *Language and Speech*, *3*(4), 179–192.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, *95*, 200–214. https://doi.org/10.1016/j.jml.2017.04.007

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of Experimental Child Psychology*, *173*, 351–370. https://doi.org/10.1016/j.jecp.2018.04.012

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119. https://doi.org/10.1073/pnas.2201968119

Heyman, T., Van Rensbergen, B., Storms, G., Hutchison, K. A., & De Deyne, S. (2015). The influence of working memory load on semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 911–920. https://doi.org/10.1037/xlm0000050

Hintz, F., Voeten, C. C., Dobó, D., Lukics, K. S., & Lukács, Á. (2024). The role of general cognitive skills in integrating visual and linguistic information during sentence comprehension: Individual differences across the lifespan. *Scientific Reports*, *14*(1), 17797. https://doi.org/10.1038/s41598-024-68674-3

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. https://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. https://doi.org/10.1177/0267658315624960

Huettig, F. (n.d.). *Looking Ahead*. Cambridge University Press & Assessment. Retrieved November 3, 2025, from https://www.cambridge.org/universitypress/subjects/psychology/cognition/looking-ahead-new-science-predictive-mind

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., & 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., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223

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

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

Ito, A. (2025, March 27-29). *Are working memory resources required for linguistic prediction during reading?* [Poster Presentation]. Annual Conference of the Society for Human Sentence Processing. College Park, Maryland, United States.

Ito, A., Bautista, A., & Martin, C. (2025). Cognate facilitation effect on verb-based semantic prediction in L2 is modulated by L2 proficiency. *Bilingualism: Language and Cognition*, *28*(3), 684–696. https://doi.org/10.1017/S1366728924000968

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., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017a). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. Language, Cognition and Neuroscience, 32(8), 954–965. https://doi.org/10.1080/23273798.2016.1242761

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017b). Why the A/AN prediction effect may be hard to replicate: A rebuttal to Delong, Urbach, and Kutas (2017). Language, Cognition and Neuroscience, 32(8), 974–983. https://doi.org/10.1080/23273798.2017.1323112

Ito, A., & Pickering, M. J. (2021). Chapter 2. Automaticity and prediction in non-native language comprehension. In E. Kaan & T. Grüter (Eds.), *Prediction in Second Language Processing and Learning* (pp. 25–46). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.02ito

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Jaeger, T., & Levy, R. (2006). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems*, *19*. https://proceedings.neurips.cc/paper_files/paper/2006/hash/c6a01432c8138d46ba39957a8250e027-Abstract.html

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

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

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. https://doi.org/10.1016/j.cognition.2010.12.002

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalography and Clinical Neurophysiology. Supplement*, *39*, 325–330.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Li, K., Hopkins, A. K., Bau, D., Viégas, F., Pfister, H., & Wattenberg, M. (2024). *Emergent World Representations: Exploring a Sequence Model Trained on a Synthetic Task* (No. arXiv:2210.13382). arXiv. https://doi.org/10.48550/arXiv.2210.13382

Li, X., & Qu, Q. (2024). Verbal working memory capacity modulates semantic and phonological prediction in spoken comprehension. *Psychonomic Bulletin & Review*, *31*(1), 249–258. https://doi.org/10.3758/s13423-023-02348-5

Liu, Y., Hintz, F., Liang, J., & Huettig, F. (2022). Prediction in challenging situations: Most bilinguals can predict upcoming semantically-related words in their L1 source language when interpreting. *Bilingualism: Language and Cognition*, *25*(5), 801–815. https://doi.org/10.1017/S1366728922000232

Liu, S., Hu, W., & Liu, X. (2025). Different effects of verbal and visual working memory loads on Language prediction. Scientific Reports, 15(1), 22432. https://doi.org/10.1038/s41598-025-03556-w

Lupyan, G., & Clark, A. (2015). Words and the World: Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

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

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

Mantegna, F., Hintz, F., Ostarek, M., Alday, P. M., & Huettig, F. (2019). Distinguishing integration and prediction accounts of ERP N400 modulations in language processing through experimental design. *Neuropsychologia*, *134*, 107199. https://doi.org/10.1016/j.neuropsychologia.2019.107199

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

Miller, G. A. (1956). *The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information[1]*.

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability—Evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*(1). https://doi.org/10.16910/jemr.5.1.3

Moos, A., & Trouvain, J. (2007). Comprehension of ultra-fast speech–blind vs.“normally hearing” persons. *In Proceedings of the 16th International Congress of Phonetic Sciences (Vol. 1, pp. 677-680).*

Ness, T., & Meltzer-Asscher, A. (2018). Predictive Pre-updating and Working Memory Capacity: Evidence from Event-related Potentials. *Journal of Cognitive Neuroscience*, *30*(12), 1916–1938. https://doi.org/10.1162/jocn_a_01322

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The Influence of Language Proficiency on Lexical Semantic Processing in Native and Late Learners of English. *Journal of Cognitive Neuroscience*, *24*(5), 1205–1223. https://doi.org/10.1162/jocn_a_00143

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. eLife, 7, e33468. https://doi.org/10.7554/eLife.33468

Nota, N., Corps, R. E., Hadley, L. V., Naylor, G., & Pickering, M. J. (2024, September 5-7). *The effect of cognitive load on different stages of language prediction: a visual-world eye-tracking study*. [Poster Presentation]. Architectures and Mechanisms for Language Processing. Edinburgh, Scotland, United Kingdom.

Nour Eddine, S., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (2024). A predictive coding model of the N400. *Cognition*, *246*, 105755. https://doi.org/10.1016/j.cognition.2024.105755

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

Özkan, D., Küntay, A. C., & Brouwer, S. (2022). The role of verbal and working memory skills in Turkish-speaking children's morphosyntactic prediction. *Applied Psycholinguistics*, *43*(6), 1305–1328. https://doi.org/10.1017/S0142716422000388

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76. https://doi.org/10.1037/bul0000124

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

Rayner, K., Reichle, E., Stroud, M., Williams, C., & Pollatsek, A. (2006). The Effect of Word Frequency, Word Predictability, and Font Difficulty on the Eye Movements of Young and Older Readers. *Psychology and Aging*, *21*, 448–465. https://doi.org/10.1037/0882-7974.21.3.448

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(2), 514–528. https://doi.org/10.1037/a0020990

Ryskin, R., Levy, R. P., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*, 107258. https://doi.org/10.1016/j.neuropsychologia.2019.107258

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: What is next? *Trends in Cognitive Sciences*, *27*(11), 1032–1052. https://doi.org/10.1016/j.tics.2023.08.003

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. Annual Review of Psychology, 69(Volume 69, 2018), 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, *112*(1), 1–36. https://doi.org/10.1037/0096-3445.112.1.1

Szmalec, A., Brysbaert, M., & Duyck, W. (2012). Working memory and (second) language processing. In J. Altarriba & L. Isurin (Eds.), *Memory, Language, and Bilingualism: Theoretical and applied approaches* (1st ed., pp. 74–94). Cambridge University Press. https://doi.org/10.1017/CBO9781139035279.004

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

Tauroza, S., & Allison, D. (1990). Speech Rates in British English. *Applied Linguistics*, *11*(1), 90–105. https://doi.org/10.1093/applin/11.1.90

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of Face Perception. Annual Review of Neuroscience, 31(Volume 31, 2008), 411–437. https://doi.org/10.1146/annurev.neuro.30.051606.094238

Wang, L. (2021). British English-Speaking Speed 2020. *Academic Journal of Humanities & Social Sciences*, *4*(5). https://doi.org/10.25236/AJHSS.2021.040517

Wang, Z., Hui, C. K. E., & Ito, A. (2025, March 27-29). *Domain-general working memory modulates semantic prediction in native speakers* [Poster Presentation]. Annual Conference of the Society for Human Sentence Processing. College Park, Maryland, United States. https://www.researchgate.net/publication/390760945_Domain-general_working_memory_modulates_semantic_prediction_in_native_speakers

Wilshire, C. E. (1999). The “Tongue Twister” Paradigm as a Technique for Studying Phonological Encoding. *Language and Speech*, *42*(1), 57–82. https://doi.org/10.1177/00238309990420010301

Yan, S., Kuperberg, G. R., & Jaeger, T. F. (2017). Prediction (or not) during language processing. A commentary on Nieuwland et al. (2017) and DeLong et al. (2005). Neuroscience. https://doi.org/10.1101/143750

## Curriculum Vitae

Name: Christopher Todd Allison

Anschrift: Erwin-Schrödinger-Straße, Gebäude 57-534, 67663 Kaiserslautern

E-Mail: allison@rptu.de

## Akademischer Werdegang

04/2022 –

Promotion (Dr. rer. Nat)

FB Sozialwissenschaften

Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau

10/2018 – 03/2022:

Master-Studium Cognitive Science (M. Sc.)

Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau

09/2010 – 03/2014:

Bachelor-Studium Biologie (B. Sc.)

University of South Carolina Upstate

## Publikationen

**Allison, C.**, Huettig, F., Fernandez, L., & Lachmann, T. (2025). Visuospatial working memory load reduces semantic prediction in the visual world. *Language, Cognition, and Neuroscience*, 1-10. DOI: 10.1080/23273798.2025.2522272

Fernandez L.B., Hadley L.V., Gamboa J.C.B., **Allison C.,** & Allen S.E.M. (2025). The impact of speech rate on first- and second-stage prediction in L1 and L2 speakers. *Bilingualism: Language and Cognition*, 1-13. DOI: 10.1017/S1366728925100515

**Allison, C.**, Huettig, F., Lachmann, T., Allen, SEM, & Beck, AK. (in review). Are Visuospatial Working Memory Influences on Prediction in Speech Contingent on the Co-presence of a Relevant Visual World? An ERP study. *Language, Cognition, and Neuroscience.*

**Allison, C.**, Huettig, F., Fernandez, LB., Lachmann, T. (submitted). Do Visuospatial and Verbal Working Memory Demands Differentially Affect Predictive Eye Gaze in the Visual World? *Cognitive Science*.

Gamboa, JCB., Alshaibani, S., **Allison, C.**, Fernandez, LB., Allen, SEM (in preparation). Divergence Point Analysis: some limitations of the bootstrapping method of the divergence point.