Edinburgh Research Explorer

A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension

A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension*

Aine Ito[1,2], Martin Corley[1] & Martin J. Pickering[1]

[1] Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom
[2] Faculty of Linguistics, Philology & Phonetics, University of Oxford, United Kingdom

Address for correspondence:
Aine Ito
Faculty of Linguistics, Philology & Phonetics
University of Oxford
Clarendon Institute, Walton Street
Oxford, OX1 2HG
United Kingdom
Email: aine.ito@ling-phil.ox.ac.uk

Abstract

We used the visual world eye-tracking paradigm to investigate the effects of cognitive load on predictive eye movements in L1 (Experiment 1) and L2 (Experiment 2) speakers. Participants listened to sentences whose verb was predictive or non-predictive towards one of four objects they were viewing. They then clicked on a mentioned object. Half the participants additionally performed a working memory task of remembering words. Both L1 and L2 speakers looked more at the target object predictively in predictable- than in non-predictable sentences when they performed the listen-and-click task only. However, this predictability effect was delayed in those who performed the concurrent memory task. This pattern of results was similar in L1 and L2 speakers. L1 and L2 speakers make predictions, but cognitive resources are required for making predictive eye movements. The findings are compatible with the claim that L2 speakers use the same mechanisms as L1 speakers to make predictions.

Keywords: prediction, working memory, bilingualism, comprehension, visual world, eye-tracking

1 Introduction

During comprehension, people construct representations that help them predict what may be mentioned next (Kuperberg & Jaeger, 2016). Are these predictions resource-intensive? To explore how predictions are affected by factors that can reduce available cognitive resources, we used the visual world eye-tracking paradigm to compare predictions between people who performed a concurrent working memory task and those who did not. In addition, we considered predictions in native (L1) speakers and non-native (L2) speakers, as they may differ in their resources and therefore in the extent to which they may be affected by the working memory load manipulation.

People can make linguistic predictions based on various types of information, including semantic information from a preceding verb (Altmann & Kamide, 1999), and more complex, message-level information based on the combination of different aspects of the context (Kamide, Altmann, & Haywood, 2003; Otten & Van Berkum, 2008). Generating predictions based on various sources of information is likely to be resource-demanding. In comparison, predictions driven by verb meaning may rely more on lexical priming and may consume fewer resources (Huettig, 2015). Consider activation from a verb to its typical agents and patients (which occurs independently of sentence contexts; Ferretti, McRae, & Hatherell, 2001). Kukona, Fang, Aicher, Chen, and Magnuson (2011) had participants hear sentences that predicted a verb's patient (e.g., *"Toby arrests the crook."*). They found that verbs (e.g., *arrests*) can lead to predictive eye movements to both their typical agents (e.g., *policeman*) and patients (e.g., *crook*). This finding suggests that predictions can be (at least partly) driven by semantic associations between verbs and nouns.

Although such lexically-driven prediction might be perceived to be relatively easy, two factors might influence a comprehender's ability to predict. The first of these is language

proficiency: L2 speakers of a language may be less proficient at predicting in that language than their L1-speaking peers. The second is the availability of cognitive resources: To the extent that prediction requires such resources, lower availability (for example as a result of a concurrent load task) may compromise the ability to predict. Because L2 comprehension may already demand greater cognitive resources than comprehension in L1 (Segalowitz & Hulstijn, 2009), we might expect that the effects of load would be especially detrimental to prediction in L2. Below we consider the evidence for predictive processing during L1 and L2 comprehension and effects of cognitive load on predictive processing in turn.

1.1   *Prediction during L1 and L2 comprehension*

Visual world eye-tracking experiments studying prediction have found that participants make use of earlier sentence information to direct their eyes to objects that are likely to be mentioned (Altmann & Kamide, 1999; Chambers, Tanenhaus, & Magnuson, 2004; Kamide, Scheepers, & Altmann, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Kukona et al., 2011). For instance, Altmann and Kamide presented L1 English participants with sentences such as "*The boy will eat the cake*" together with a scene depicting a cake and some inedible objects. The participants were more likely to fixate the cake before it was mentioned than when they heard "*The boy will move the cake*." The predictive eye movements suggest that people process sentences incrementally, and integrate information extracted from each word to build predictions about upcoming words.

Chambers and Cooke (2009) found predictive eye movements in L2 speakers in a similar experiment to Altmann and Kamide (1999). Late English-French bilinguals who had relatively high French proficiency listened to French sentences, such as "*Marie va nourrir la poule*" (*Marie will feed the chicken*) or "*Marie va décrire la poule*" (*Marie will describe the chicken*), while viewing a scene where all the depicted objects could plausibly be described

but only the chicken could plausibly be fed. Participants were more likely to look at the chicken when they heard the verb *nourrir* (*feed*) relative to when they heard *décrire* (*describe*) (and before hearing *poule*, chicken). Using a similar experimental design, Dijkgraaf, Hartsuiker, and Duyck (2016) found that both L1 and L2 speakers made similarly predictive eye movements. These findings suggest that L2 speakers are able to predict meaning of an upcoming word based on the meaning of the preceding verb.

However, some studies found evidence for prediction in L1 speakers but not in L2 speakers (Ito, Martin, & Nieuwland, 2016; Martin et al., 2013; Mitsugi & MacWhinney, 2016). Kamide, Altmann, et al. (2003; Experiment 3) showed that L1 Japanese speakers were able to utilise a case marker to predict an upcoming word. Their participants heard sentences with a dative structure (e.g., "*weitoresu-ga kyaku-ni tanosigeni hanbaagaa-o hakobu*"; order-matched English equivalent: waitress-nominative customer-dative merrily hamburger-accusative bring, meaning "*the waitress will merrily bring the hamburger to the customer*"). The presence of both nominative- and dative-marked nouns supported prediction of the occurrence of the direct object *hamburger*. After hearing these nouns, participants tended to fixate a predictable object (a hamburger) before hearing the predictable word. Mitsugi and MacWhinney (2016) replicated their findings with L1 Japanese speakers, but not with intermediate L2 Japanese speakers, despite the fact that their L2 participants exhibited good knowledge of Japanese case markers in an offline grammar test. The authors proposed that L2 speakers' grammatical knowledge might not be readily accessible for use in prediction during online comprehension, as their stimuli required participants to rely on rather complex cues (combinations of semantic and syntactic information) to make predictions. These findings might indicate that intermediate L2 speakers predict less well than L1 speakers, and that they do not predict when predictions are to be made via relatively complex linguistic computation.

Although L2 speakers may not always predict to the same extent as L1 speakers, Kaan (2014) claimed that the fundamental mechanisms of predictive processing do not differ between L1 and L2 speakers. In her framework, whether L2 speakers predict or not may depend on several factors, but the same factors may also affect whether L1 speakers predict or not, and differences in the extent of prediction depend on factors that mediate linguistic processing in general. These include stored lexical information (e.g., frequency, lexical associations) and exposure to the target language; factors that mediate not only prediction, but also general comprehension. According to this account, it is possible that predictions in L1 and L2 are similar when predictive processing does not involve complex linguistic computations, or when the sentence is made up of high-frequency words with a simple syntactic structure, to which L2 speakers have rich exposure.

1.2   *Effects of cognitive load on predictive processing*

Predictions can be made through integration of information encountered in the on-going sentence, information from the non-linguistic (visual world) environment, and information in the comprehender's memory (Slevc & Novick, 2013). Given that this integrative mechanism requires a memory retrieval process, these predictions are likely to be affected by working memory load. Consistent with this hypothesis, Huettig and Janse (2016) found a positive correlation between people's working memory capacities and their predictive eye movements in the visual world paradigm. People with greater working memory capacities made more predictive eye movements, using grammatical gender information conveyed by Dutch articles. Huettig and Janse's findings suggest that some of the cognitive resources that are used for making predictive eye movements are also used for performing a working memory task. But to be confident of this, it is of course necessary to show a causal relationship – that a high cognitive load interferes with predictive eye movements.

Such an interference effect of cognitive load on prediction may be particularly strong during L2 processing. As we have noted, L2 speakers sometimes fail to use complex cues for prediction (e.g., Mitsugi & MacWhinney, 2016), perhaps because cognitive resources are likely to be reduced during L2 processing relative to L1 processing (Segalowitz & Hulstijn, 2009). Therefore, if cognitive resources that are used for L2 comprehension are shared by working memory resources that are used for remembering words, an additional cognitive load during L2 comprehension may increase the effects of cognitive load.

However, predictions that do not rely on complex linguistic computations may not consume many resources. Such predictions may be unaffected by cognitive load. Moreover, if L1 and L2 speakers do not differ in the mechanisms of prediction, as Kaan (2014) claimed, then they may be similarly affected by cognitive load.

1.3   The current study

We investigated whether straightforward linguistic predictions in simple sentences are affected by cognitive load. We used sentences with the Subject-Verb-Object structure (cf. Altmann & Kamide, 1999), where predictions could be made based on the verb meaning. We recruited L1 English speakers (Experiment 1) and advanced L2 speakers of English (Experiment 2) in order to examine effects of language proficiency on predictive eye movements. In both experiments, we recorded participants' eye movements as they listened to sentences containing a predictive verb (e.g., *fold*) that was compatible with one of four depicted objects (target object; e.g., *scarf*) or a non-predictive verb (e.g., *find*) that was compatible with all the depicted objects. We expected participants who made predictions to be more likely to fixate a target object following a verb which predicted it than following a non-predictive verb. Importantly, we expected this difference to emerge before the name of the target object could be processed.

To ensure that none of the words would be new to L2 speakers, target object names had relatively high frequency and low Age of Acquisition (AoA) (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). To examine whether semantic features of the predictable word were pre-activated, we included an object which was semantically related to the predictable target object as one of the distractors. If participants pre-activated semantic features, we expected that semantically related objects would attract more fixations than unrelated distractors (semantic competitor effect, cf. Yee & Sedivy, 2006).

We manipulated cognitive load using a word-remembering task (cf. Gordon et al., 2002). Half the participants had to remember a list of words while listening to the sentences. If predictions hinged on available cognitive resources, we expected predictive eye movements to be delayed or eliminated when participants were under the added memory load. If the cognitive load had a greater effect on L2 speakers' predictive performances, L2 speakers would be less likely to predict than L1 speakers under the cognitive load. If semantic competitor effects were shown to occur under predictive sentences, these might also be reduced under the cognitive load. This would in turn suggest that pre-activating semantic information requires cognitive resources.

2 Experiment 1

2.1 Method

Participants

Forty-eight native English speakers studying at the University of Edinburgh participated in the experiment. All of the participants had normal vision, and none reported any language disorder.

Stimuli

The auditory stimuli comprised 16 sentence pairs, each of which had two conditions, differing only at the critical verb (see Appendix for a full set of items). In the *predictable* condition, the target object was the only appropriate patient of the verb among four depicted objects (e.g., "*The lady will fold the scarf.*"). In the *unpredictable* condition, any of the depicted objects could plausibly be the patient of the verb (e.g., "*The lady will find the scarf.*"). The sentences were recorded by a female native British English speaker, and sampled at 48 kHz with a format of 32-bit float. The speaker read the sentences at a rate of 1.3 syllables per second with some pauses between phrases (following Altmann & Kamide, 1999). The mean durations for the critical verbs and target nouns were 870 ms and 1098 ms respectively. The relatively slow speech was intended to create optimal conditions for predictive eye movements, such that any effects of load or population would be easy to observe.

Figure 1. An example picture (for the sentences "The lady will fold/ find the scarf.").

The visual stimuli were 16 experimental displays, each with four objects, one depicted in each quadrant (Figure 1). All the objects in a given display were presented in one of seven colours (grey, pink, purple, yellow, green, blue, and brown) in either a dark or a light shade.[1] Each of the target objects (e.g., scarf) was matched with a semantic competitor that was in the same semantic category (e.g., high heels) (according to Van Overschelde, Rawson, & Dunlosky, 2004). The other two objects were distractors, and they were also in the same category as each other (e.g., violin, piano) but from a different category to the target and the semantic competitor, to prevent participants determining that the target object would be one that had a category coordinate in the array. We assessed semantic relatedness using pairwise Latent Semantic Analysis (LSA) between target words and semantic competitor/distractor words (Landauer & Dumais, 1997). The LSA value was higher for the semantic competitor words ($M = .36$, $SD = .19$) than for the distractor words ($M = .10$, $SD = .07$), $p < .001$. The names of target, semantic competitor, and distractor objects did not differ in CELEX frequency (Baayen, Piepenbrock, & Gulikers, 1995), $F(2, 56) = .36$, $p = .70$, or AoA, $F(2, 54) = 1.7$, $p = .19$. The mean frequency (per 17.9 million) and AoA for object names were 1094 ($SD = 1633$) and 5.0 years ($SD = 1.4$ years), respectively. The positions of all object types were counterbalanced across items.

We tested the experimental items for predictability. Twenty-one native English speakers were presented with the coloured pictures and sentences without target words, but with two-word shade and colour modifiers (e.g., "*The lady will fold/find the dark brown _____.*"), and told to give the name of one of the depicted objects to complete the sentence. After excluding unclear answers (1%), which could refer to more than one object, participants selected the target object 92% of the time in the predictable condition and 26% of the time in the unpredictable condition (with the other responses split among the semantic competitor and distractors). Thus, target objects were generally considered to be the most

plausible continuation in the predictable condition, and no more (or less) plausible than the other three objects in the unpredictable condition.

The experimental items additionally included 16 fillers. The filler sentences were similar in length and sentence structure to the critical sentences. Accompanying pictures depicted four objects in one of the colours and shades used in the critical items, and between one and four of the objects could serve as a plausible patient of the verb. The pictures also comprised two pairs of semantically related objects.

For the working memory task, 160 mid-frequency words were selected from low-concreteness (concreteness < 3, on the scale of 1-5) words in the corpus of Brysbaert, Warriner, and Kuperman (2014). All the words had the maximum of three syllables. Each picture-sentence pair was matched with a set of 5 words. The words were unrelated to the picture or to the sentence. The words in each set were semantically unrelated to one another and did not share onset or offset syllables.

Procedure

There were two conditions regarding the cognitive load manipulation (no-load, and load), and participants were sequentially assigned to one of the conditions. Participants were seated in front of a computer screen and tested individually in a quiet room. They were instructed that they would hear a sentence and see a picture at the same time, and were asked to click on a mentioned object. The presentation order was randomised, and every participant saw items in a different order. No participant saw more than two items in the same condition successively. Eye movements were recorded using an EyeLink 1000 tower mount eye-tracker sampling at 500 Hz. Participants placed their chin on a chin rest, and the eye-tracker was calibrated using a nine-point calibration grid. The pictures were presented on a computer

monitor at a resolution of 1024×768 pixels. Before every trial, drift correction was performed. The pictures disappeared when the participants clicked on an object.

Participants first clicked the mouse when they were ready. Participants in the load condition then saw five words (presented together) on the screen for eight seconds. All participants then saw a 500 ms blank screen followed by the pictures. Pictures were presented 1000 ms before the sentence onset in order to give participants a preview (cf. Huettig, Rommers, & Meyer, 2011). Participants then clicked on the picture that they judged to correspond to the final word. Participants in the load condition then attempted to list the words in any order within eight seconds.

No feedback was given during the experiment. The position of the mouse pointer was corrected to the centre of the screen after every trial. The experiment started with two practice trials, and lasted for about 15-25 minutes.

2.2   Results

Behavioural task accuracy

Because of a software error, the mouse-clicking responses were not recorded for participants who were assigned to the no-load condition. The accuracy for the target clicking task in the load condition was 100%. The mean number of correctly recalled words for the working memory task was 3.6 ($SD = .71$; range = 2.3-4.8).

Eye-tracking data analyses

The eye-tracking data were analysed using linear mixed-effects models with the lme4 package (Bates, Maechler, & Dai, 2008) in R (R Development Core Team, 2015). The proportion of time spent fixating on target and semantic competitor objects was calculated separately for each 50 ms bin relative to the target noun onset (following Altmann & Kamide,

1999). We then transformed fixation probability in every time bin into log odds, using the

empirical logit function (Barr, 2008). Blinks were not considered as part of fixations. We

constructed two linear mixed-effects models, which evaluated the log-transformed fixation

probabilities on target objects and on semantic competitor objects as predicted by

Predictability (predictable vs. unpredictable), Load (no-load vs. load), and the interaction of

Predictability by Load. The models included random intercepts by participants and by items.

The by-participant random slope for Load was not included because random slopes are not

appropriate for between-subject factors. The random slopes for Predictability and for the

Predictability by Load interaction were not included because the model with it did not

converge in several bins. The variable Predictability was numerically coded and centred.

Because we were interested in the time-course of prediction, the models described above

were run for every 50 ms bin from 1500 ms before to 500 ms after the target word onset

(Borovsky, Elman, & Fernald, 2012; Ellis, Borovsky, Elman, & Evans, 2015).[2] This way of

analysing the time-course increases the likelihood of Type I errors, but we note that the

differences reported below show consistently reliable effects over multiple bins. The effect of

Predictability was evaluated by assessing whether the absolute *t*-value exceeded 2 (Baayen,

Davidson, & Bates, 2008).

One of the participants in the load condition failed to complete two trials because of a

technical problem. These trials were treated as missing in the eye-tracking analyses.

Effects of prediction and load

Figure 2 shows fixation probabilities on target objects and mean fixation probabilities

on distractor objects in the predictable and unpredictable conditions, separately for the load

condition and for the no-load condition. The time was synchronised to target noun onset, with

verb onset and offset being the means of all the critical items. The graphs show the time window from 2000 ms before to 500 ms after the target noun onset.

Visual inspection of the graphs suggests that differences in the fixation proportions on target objects in the predictable versus the unpredictable condition began to emerge later in the load condition than in the no-load condition. In support of this, the linear mixed-effects model showed an interaction of Predictability by Load ($|t|$s > 2) in every 50 ms window from 1000 ms before the target noun onset until 500 ms after the target noun onset. The interactions indicate that predictive eye movements were delayed by load.

To understand the interactions in more detail, we ran another model for no-load and load conditions separately. The model evaluated the fixation probability on target objects as predicted by Predictability (predictable vs. unpredictable), including random intercepts by participants and by items. As the upper panel of Figure 2 indicates, participants in the no-load condition were more likely to look at target objects in the predictable condition than in the unpredictable condition from 1050 ms before the noun onset onwards (shown as ● in Figure 2). This corresponded almost exactly to mean target verb offset. The result suggests that participants in the no-load condition predicted upcoming objects that were predictable. As the lower panel of Figure 2 shows, participants in the load conditions were also were also more likely to look at target objects in the predictable condition than in the unpredictable condition, but this effect did not emerge until 250 ms before the target noun onset. To sum up, the analyses show that predictive eye movements occurred in both conditions, but that they began about 800 ms earlier in the no-load than the load condition.
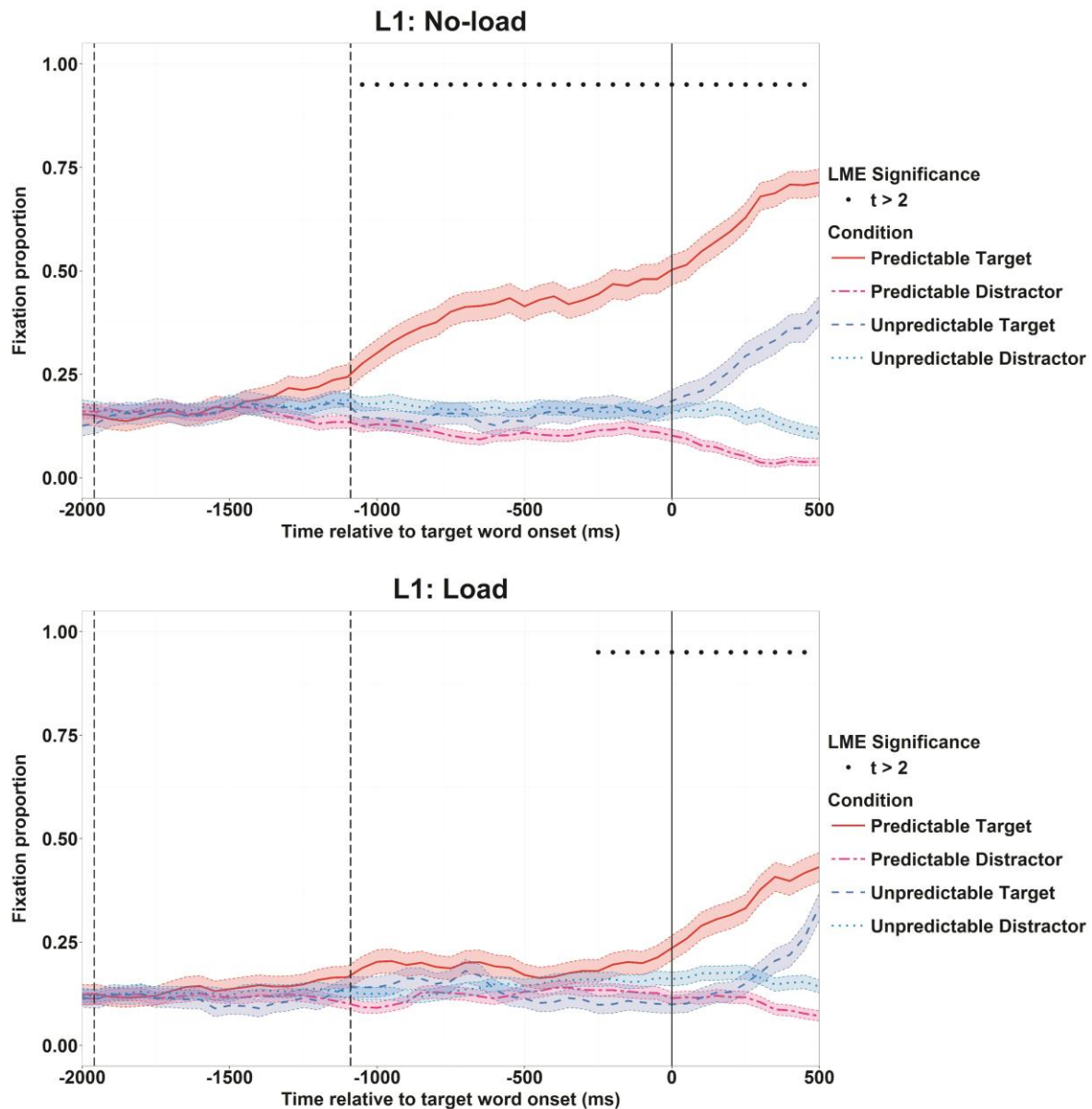
Figure 2. Fixation probabilities on target objects and mean fixation probabilities on distractor objects in the predictable and unpredictable conditions in the no-load condition (top) and in the load condition (bottom) in Experiment 1. Time 0 ms shows target word onset. The left-most dashed line on the y-axis direction (y = -1959 ms) indicates mean verb onset; the next dashed line (y = -1090 ms) indicates mean verb offset. Standard error bars are represented using transparent thick lines. The significance of the model ($|t|>2$) is shown on the top of the graphs, with a solid circle (●) showing a significant effect of Predictability.

We additionally explored whether there was a relationship between participants' recall performances and their degree of prediction. To test this, we computed the log-transformed target fixation proportion difference between the predictable and the unpredictable conditions from 250 ms before the target noun onset until the target noun onset

(the time window where L1 participants in the load condition showed a significant effect of predictability). We computed a correlation between this fixation proportion difference and the mean recall score for each participant, but the correlation was not significant, $r(22) = .002$, $p > .99$).[3]

Effects of cognitive load in the unpredictable condition

Figure 2 suggests that identification of unpredictable target objects is delayed by the cognitive load. In order to compare the fixation probability difference between the target and distractors without violating the assumption of independence between observations. We subtracted log odds of fixations on target objects from averaged log odds of fixations on distractor objects, and used this as a dependent variable. We analysed this dependent variable using a linear mixed-effects model, which included a fixed effect of Load (no-load vs. load), and random intercepts by participants and by items. The model run in each 50 ms window showed a significant effect of Load from 50 ms until 500 ms after the target word onset ($|t|$s > 2). Therefore, cognitive load influenced eye movements to target objects in the unpredictable condition as well.

Semantic competitor effect

We conducted analyses on the semantic competitor that were parallel to the analyses on the target. The linear mixed-effects model showed no effects of Predictability or Load, nor an interaction of Predictability by Load, on the log-transformed proportion of looks to the semantic competitor objects in any of the time windows ($|t|$s < 2). We also conducted the analyses parallel to the analyses in the unpredictable condition in order to test whether there is a semantic competitor effect in neutral sentence contexts. The linear mixed-effects model did not show a significant effect of semantic competitor (intercept term) or a significant effect

of Load in any time bin between -1500 ms to 500 ms relative to the target word onset ($|t|$s <

2). Therefore, there was no indication of a semantic competitor effect in Experiment 1.

2.3   Discussion

Experiment 1 investigated whether making successful predictive eye movements

during language comprehension is affected by cognitive load. We found that predictive eye

movements in L1 speakers occurred whether or not those speakers were faced with the

additional cognitive load. However, the cognitive load led to those predictive eye movements

being delayed. It seems that the additional load caused participants to have fewer cognitive

resources that could be allocated to making predictive eye movements.

3   Experiment 2

Experiment 2 addressed similar questions to Experiment 1, but used L2 speakers of

English. It asked whether predictive eye movements in L2 speakers occurred under

conditions of load and no load, and whether load caused any predictive eye movements to be

delayed. Given the results of Experiment 1, we hypothesised that predictive eye movements

in L2 speakers would also be delayed under load. Alternatively, L2 speakers may not make

predictive eye movements at all under load, due to fewer resources available during L2

comprehension.

3.1   Method

Participants

Forty-eight L2 English speakers studying at the University of Edinburgh participated

in Experiment 2. Native languages of the L2 participants were Chinese (20), Polish (3),

Spanish (3), Romanian (2), Norwegian (2), German (2), Lithuanian (2), Malay, French,

Czech, Dhivehi, Greek, Bulgarian, Swedish, Russian, Urdu, Catalan, Slovak, Dutch, Hindi,

and Armenian. They filled in a language background questionnaire before the experiment.

The collected L2 proficiency measures for participants in each condition are shown in Table

1 (along with their native languages). L2 participants in the no-load and load conditions did

not differ on any of these measures, ps > .15.

	No-load	Load
Length of stay in the UK (months)	13.5 (9.6)	23.8 (33.5)
Length of exposure to English (years)	11.8 (4.7)	13.1 (4.8)
Self-rated English proficiency (1 = not good at all, 10 = very good)	8.3 (1.2)	8.0 (1.0)
Native language	Chinese (14), Lithuanian (2), Romanian, Russian, Dutch, Urdu, Catalan, Hindi, Slovak, German	Chinese (6), Polish (3), Spanish (3), Norwegian (2), Romanian, Dhivehi, French, Czech, Armenian, German, Bulgarian, Malay, Swedish, Greek

Table 1. The means of the English proficiency measures in L2 participants (with standard deviations in parentheses) and their native languages.

Stimuli and procedure

The stimuli and the procedure in Experiment 2 were identical to those in Experiment

1.

3.2   Results

Behavioural task accuracy

Because of a software error, the mouse-clicking responses were not recorded for

participants who were assigned to the no-load condition. The accuracy for the target clicking

task for participants in the load condition was 98%. Incorrectly answered trials were excluded

from the eye-tracking analyses. In the working memory task, the mean number of correctly

recalled words was 3.3 (SD = .78; range =2.1-4.6).

Eye-tracking data analyses

The eye-tracking data were analysed as in Experiment 1.

Effects of prediction and load

Figure 3 shows the fixation probabilities on target objects and the averaged fixation probabilities on distractor objects in the predictable and unpredictable conditions for participants in the load condition and in the no-load condition separately. The time was synchronised at the target noun onset, and verb onset and offset are the means of all the critical items. The graphs show the time window from 2000 ms before to 500 ms after the target noun onset. The model testing the fixed effects and interaction of Predictability (predictable vs. unpredictable) and Load (no-load vs. load) showed a significant interaction of Predictability by Load ($|t|s > 2$) in every 50 ms window from 850 ms before the target noun onset all until 500 ms after onset. The significant interactions in the time window before target word onset indicate that participants showed more predictive eye movements when they were not under cognitive load than when they were under cognitive load.

To explore the interaction, we ran a model evaluating the fixation probability on target objects as predicted by Predictability (predictable vs. unpredictable), including random intercepts by participants and by items. We ran this model for the no-load and load conditions separately. Participants in the no-load condition were more likely to look at target objects in the predictable condition than in the unpredictable condition from 1150 ms to 1050 ms before the target word onset, and from 950 ms before the target word onset onwards (shown as ● in Figure 3). In contrast, participants in the load condition did not show a significant effect of Predictability in consecutive bins until 100 ms after the target word onset. As in Experiment 1, we can conclude that predictability effects on eye movements were significantly delayed as a result of the additional cognitive load.

Although we did not find evidence for predictive eye movements before the target word onset in L2 participants, we followed Experiment 1 and examined whether L2 participants with a better recall performance show greater prediction effects. We computed a correlation between the mean recall score of each participant and the log-transformed fixation proportion difference in the same time window we used in Experiment 1 (from 250 ms before the target noun onset until the target noun onset). The correlation was not significant, $r(22) =$ -.08, $p > .7$).[4]

We further examined the relationship between the extent of L2 participants' prediction and their English proficiency. We computed the log-transformed target fixation proportion difference between the predictable and unpredictable conditions from 200 ms after the mean verb offset (= 890 ms before the target noun onset) until the target noun onset, and used this as a proxy for the extent of prediction. We computed the correlation between this measure and L2 proficiency measures for participants in the load condition and for those in the no-load condition separately. In both groups of participants, the extent of prediction did not correlate with their self-rated proficiency scores (no-load condition, $r(22) = $ -.015, load condition, $r(22) = .19$, ps $> .3$), with their lengths of stay in the UK (no-load condition, $r(21)$[5] $= .016$, load condition, $r(22) = .23$, ps $> .2$), or with their length of exposure to English (no-load condition, $r(22) = $ -.22, load condition, $r(21)$[6] $= .082$, ps $> .3$).[7]
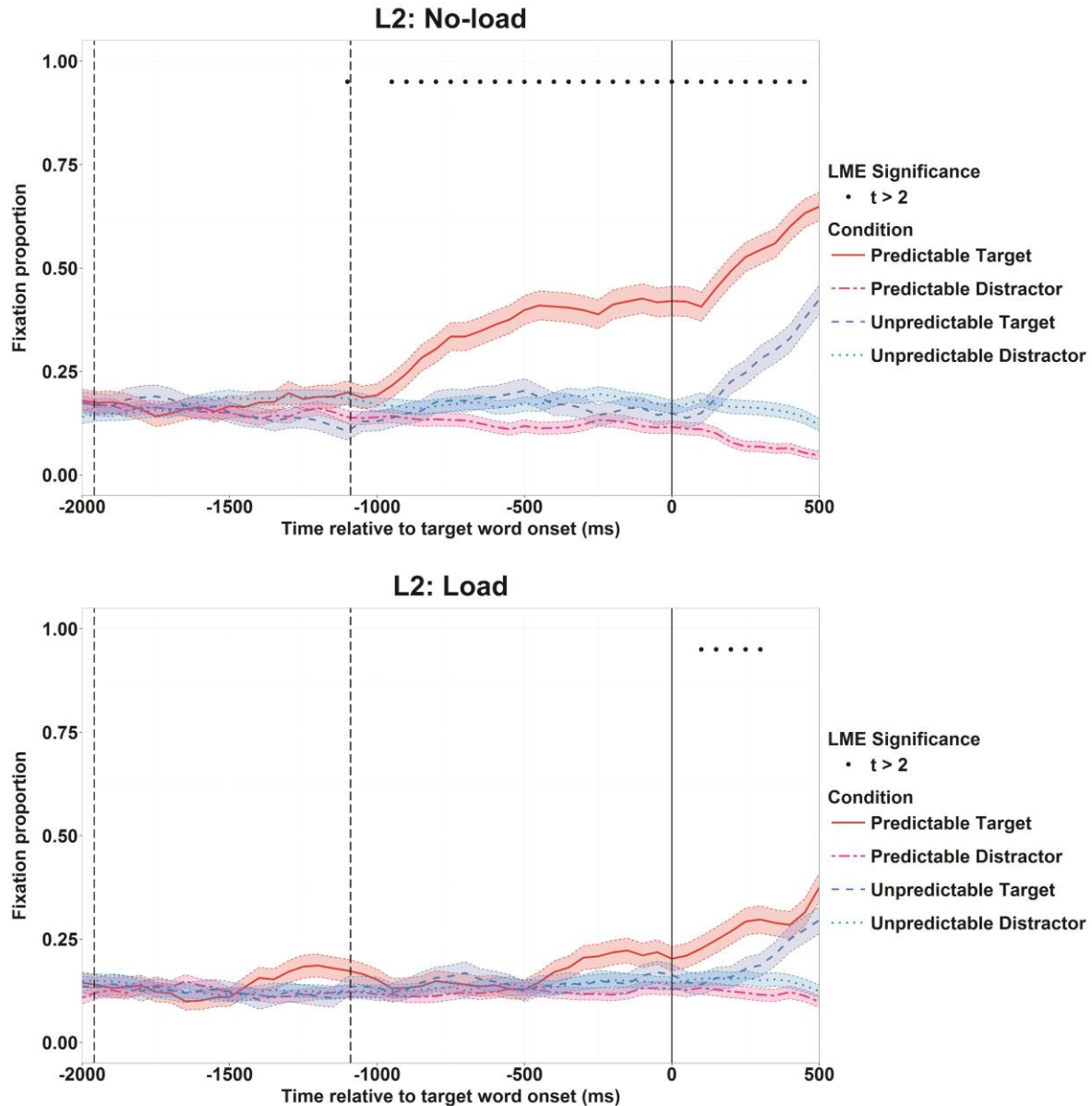
Figure 3. Fixation probabilities on target objects and mean fixation probabilities on distractor objects in the predictable and unpredictable conditions in the no-load condition (top) and in the load condition (bottom) in Experiment 2. Time 0 ms shows target word onset. The left-most dashed line on the y-axis direction (y = -1959 ms) indicates mean verb onset; the next dashed line (y = -1090 ms) indicates mean verb offset. Standard error bars are represented using transparent thick lines. The significance of the model ($|t|>2$) is shown on the top of the graphs, with a solid circle (●) showing a significant effect of Predictability.

Effects of cognitive load in the unpredictable condition

Experiment 1 found that cognitive load may interfere with general identification of

target objects. We examined if the cognitive load affected eye movements in the

unpredictable condition in Experiment 2 as well. The same linear mixed-effects model as in

Experiment 1 examined the log odds of fixations on target objects subtracted from averaged log odds of fixations on distractor objects predicted by fixed effects of Load (no-load vs. load). The model did not show a significant effect of Load in any time window from 1500 ms before the target word onset to 500 ms after the target word onset. Therefore, unlike in L1 participants, cognitive load did not influence the identification of target objects in the unpredictable condition in L2 participants.

Semantic competitor effect

The linear mixed-effects model run for the semantic competitor objects did not show any significant effect of Predictability or Load in any of the time windows ($|t|$s < 2). Following Experiment 1, we also ran a linear mixed-effects model parallel to the model in the unpredictable condition on semantic competitor objects, which showed a significant effect of semantic competitor (intercept term), but only from 950 ms before the target word onset until 850 ms before the target word onset, and a significant effect of Load from 1500 ms before the target word onset until 1400 ms before the target word onset ($|t|$s > 2). In these time windows in the non-predictable condition, target objects did not yet start to attract more fixations than distractors, so the explanation in terms of pre-activation of the semantic information of target objects does not fit.

Interaction of load by language group on prediction effects

A comparison of Experiments 1 and 2 indicates that the pattern of results was similar for L1 and L2 participants in the respect that the effect of predictability was significantly delayed in the load condition relative to the no-load condition. To explore a potential effect of language group, we ran another linear mixed-effects model on the data including both L1 and L2 participants, testing main effects and interactions of Predictability (predictable vs. unpredictable), Load (no-load vs. load) and Language group (L1 group vs. L2 group). The

model included random intercepts by participants and by items, but random slopes were not included as their inclusion did not allow the model to converge. The model showed a significant interaction of Predictability by Load throughout the time window from 1000 ms before the target noun onset until 500 ms after the target noun onset. The three-way interaction was not significant in any 50 ms time bin from 1500 ms before the target noun onset until 500 ms after the target noun onset. Therefore, we can conclude that the cognitive load manipulation affected predictive eye movements similarly in L1 and L2 speakers.

4   General Discussion

We investigated effects of cognitive load on predictive eye movements in L1 and L2 speakers. In Experiment 1, L1 English speakers listened to predictive and non-predictive sentences and clicked on an object that was mentioned in the sentence. Half the participants performed an additional working memory task of remembering word lists. In Experiment 2, fairly advanced L2 speakers of English were tested under the same conditions (i.e., either under a load or under no load). The results showed that both L1 and L2 participants directed their eyes to a predictable target object before it was mentioned (and did not show such predictive looks to the same object when the sentence was non-predictive), which suggests that they made predictions about upcoming referents. Participants who were under a cognitive load showed increased looks to predictable objects much later compared to those who did not perform the concurrent working memory task. This pattern of results was similar for L1 and L2 participants. Taken together, the results suggest that predictive eye movements draw on some of the cognitive resources that are used for remembering words.

4.1   Evidence for prediction in L2

The current findings suggest that L2 speakers can make use of the information extracted from each word to predict a likely referent in a similar manner to L1 speakers when

there is no additional cognitive load. This conclusion is consistent with Chambers and Cooke (2009) and Dijkgraaf et al. (2016), but inconsistent with Mitsugi and MacWhinney (2016). However, Mitsugi and MacWhinney's study (2016) was different in several respects from the studies of Chambers and Cooke (2009) and Dijkgraaf et al. (2016), and from our study. In Mitsugi and MacWhinney's study (2016), L2 speakers had to use syntactic information provided by case markers in addition to the meaning of encountered words to make predictions. But this combinatorial utilisation of the cues might have been particularly difficult in L2, because the manipulated syntactic rules were specific to the L2. In Chambers and Cooke (2009), Dijkgraaf et al. (2016), and our studies, the experimental sentences were syntactically simple (no double-object structure), and L2 speakers did not have to rely on L2-specific cues for predictions, so predictions were probably easier.

Another explanation for the inconsistency with Mitsugi and MacWhinney (2016) relates to proficiency. Mitsugi and MacWhinney's L2 participants were intermediate learners, having studied the L2 for 4.3 years on average. Our participants had been exposed to English for more than 12.5 years on average, and participants in Chambers and Cooke (2009) for 11.9 years. Participants in Dijkgraaf et al. were also highly proficient, judging from the vocabulary test score and self-rated proficiency. A higher proficiency may help explain successful prediction for our participants who were not under a cognitive load and in Chambers and Cooke (although the relationship between L2 prediction and L2 proficiency is not very clear, as we discuss below). Finally, the sentences in our study were spoken slowly with pauses, so our participants had longer time to process contextual information and to generate predictions compared to participants in Mitsugi and MacWhinney. Consistent with this explanation, a recent study has shown that a slower reading rate enhanced predictive processing in L1 speakers (Ito, Corley, Pickering, Martin, & Nieuwland, 2016). However, Dijkgraaf et al. (2016) found predictive eye movements while participants listened to sentences at a natural

pace. In sum, the inconsistent results between Mitsugi and MacWhinney and our studies could be explained by different types of cues, speech rate, or proficiency.

We did not find any relationship between L2 participants' proficiency scores and their predictive eye movements. This is surprising, given that Chambers and Cooke (2009) used a similar set of proficiency measures and found a robust correlation between L2 proficiency and the extent of predictive eye movements. One possibility is that our participants who did not perform the working memory task were predicting as much as is possible (i.e., at a ceiling level). Although the experimental design is similar in Chambers and Cooke's study and our own, the time between the critical verb onset and the predictable noun onset was longer in our study (1959 ms) than in Chambers and Cooke (1220 ms). This may have made predictions easier in our study. It could be that other proficiency measures such as vocabulary knowledge are related to predictive processing. However, Dijkgraaf et al. (2016) did not find any effect of L2 vocabulary knowledge on predictive eye movements (see also Nation, Marshall, & Altmann, 2003, for no effect of L1 vocabulary), and evidence is mixed as to which vocabulary knowledge (production or comprehension) is related to predictive eye movements (Borovsky et al., 2012; Mani & Huettig, 2012).

Finally, our L2 participants differed in their L1, unlike the previous studies on L2 prediction (Chambers & Cooke, 2009; Dijkgraaf et al., 2016; Mitsugi & MacWhinney, 2016). It might be that L2 participants whose L1 was more similar to English showed more predictive eye movements than L2 participants whose L1 was less similar. However, as we have noted, our L2 participants who showed predictive eye movements (i.e., those who were in the no-load condition) could have been performing at a ceiling level. Potential effects of linguistic background might be observable when prediction is more complex, where L2 speakers do not show a similar level of prediction to L1 speakers (e.g., prediction involving the utilisation of case markers).

4.2 Cognitive load affects predictive eye movements

We found that cognitive load affects predictive eye movements similarly in L1 and L2 participants. This suggests that cognitive resources are required for making predictive eye movements across different groups of participants. This finding is compatible with Kaan's (2014) claim that the mechanisms of prediction in L1 and L2 are fundamentally equivalent. The effect of load suggests that participants who were under cognitive load had to allocate the cognitive resources to the working memory task, additionally to prediction, whereas those who were not under cognitive load could focus more on prediction. Our results are therefore compatible with Huettig and Janse (2016), who showed that people with better working memory capacities made more predictive eye movements. People with larger working memory capacities have more resources available compared to those with a smaller working memory capacity. Hence, both studies found that predictive eye movements are stronger when there are more resources available.

When comprehenders made predictive eye movements in our task, they had to identify objects in the visual scene, comprehend the utterance, predict object type based on the utterance (e.g., foldable objects), judge which object in the scene is compatible with this prediction (the scarf), and finally move the eyes to this object. Load cannot simply have affected language comprehension (in general), because it did not lead to delayed fixations on target objects in the unpredictable condition in L2 participants (though it did affect such fixations in L1 participants). However, it is not clear which of the components to prediction was affected by the cognitive load.

Cognitive load may primarily have interfered with visual processes: identifying the objects in the scene or moving the eyes to the predictable object. This is consistent with the

evidence that short-term memory is responsible for storing temporary spatial information (Baddeley, 2012).[8]

Cognitive load may also have interfered with purely linguistic processes (i.e., predicting object type based on the utterance) or with the process of integrating such predictions with the visual scene (in this case, determining which object is compatible with the prediction of being foldable). The former possibility is compatible with accounts in which predictions rely on the production system (Dell & Chang, 2014; Pickering & Garrod, 2007, 2013), as participants in the load condition may well have been rehearsing words to be recalled and such rehearsal involves production processes (Baddeley, 2000). But it is also possible that semantic predictions (e.g., something a lady would fold) are largely resource-free, and it is the identification of the scarf as foldable (and the piano as unfoldable) which involves costly inference.

One remaining question is whether similar patterns of results would be found if people listened to sentences presented at a more natural pace. The no-load condition in our study provided an optimal condition for predictive eye movements, as it used slowly-spoken sentences with high-frequency words and a simple structure (i.e., Subject-Verb-Object). It is possible that predictive eye movements are less likely when people listen to sentences at a natural pace, and hence the effect of cognitive load on predictive eye movements might be reduced.

4.3 *No semantic competitor effects*

Our study found no evidence that semantic competitor objects were more likely to be fixated than distractors. It may be that the target object attracted fixations so strongly that it prevented fixations on semantic competitors (cf. Huettig et al., 2011). The lack of evidence could also be due to the constraining sentence contexts, which made semantic competitors

implausible referents. In support of this explanation, we note that Dahan and Tanenhaus (2004; Experiment 1) did not find more looks to contextually implausible phonological competitor objects than to non-competitor objects.

However, we did not find a semantic competitor effect in the unpredictable sentences either. This finding contrasts with Huettig and Altmann (2005), who used similar sentence contexts and similar semantic relationships to us. The difference could be a consequence of our clicking task (not used by Huettig and Altmann), which may have caused our participants to focus more attention on target objects and hence less on the competitor objects. Another possibility is that our two distractors were semantically related to each other as well as to target and semantic competitor objects, whereas the distractors were semantically unrelated in other studies (Huettig & Altmann, 2005; Yee & Sedivy, 2006). Participants in these studies could have determined that two objects were semantically related and predicted that the sentence would refer to one of them on the basis of the visual scene alone. Alternatively, fixations on one related object might have preferentially led to fixations on its related partner – something that could not happen for the unrelated objects.

5   *Conclusion*

We reported two experiments that investigated whether L1 and L2 speakers' predictions are subject to processing limitations. We found similar predictive eye movements in L1 and L2 speakers, but these predictive eye movements were delayed for participants who performed a working memory task of remembering words concurrently. Thus, making predictive eye movements appears to require cognitive resources that are used for remembering words. Similar effects of cognitive load on L1 and L2 speakers are compatible with the account that the mechanisms of prediction are not fundamentally different in L1 and L2 comprehension.

Appendix

Critical sentences and names of the depicted objects

Sentence	Target	Semantic competitor	Distractors
The boy will close/touch the cabinet.	cabinet	table	teddy bear, yo-yo
The lady will fold/find the scarf.	scarf	high heels	piano, violin
The boy will catch/describe the dragonfly.	dragonfly	eagle	shark, whale
The man will fire/bring the gun.	gun	bomb	watering can, water hose
The man will fly/ check the airplane.	airplane	motorcycle	computer, television
The boy will beat/choose the drum.	drum	guitar	video game, puzzles
The magician will bend/move the spoon.	spoon	cup	coin, paper
The housewife will mop/wash the floor.	floor	carpet	pants, skirt
The teacher will answer/open the door.	door	window	book, letter
The woman will heat/fetch the pan.	pan	knife	purse, ring
The child will dress/borrow the doll.	doll	board game	newspaper, comic book
The woman will climb/use the stairs.	stairs	elevator	bus, train
The woman will light/clean the lamp.	lamp	bed	plate, bowl
The woman will iron/wear the shirt.	shirt	shoes	earrings, necklace
The man will shoot/need the bow.	bow	sword	spatula, fork
The girl will sharpen/buy the pencil.	pencil	ruler	slippers, running shoes

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Baayen, H. R., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baayen, H. R., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). Philadelphia, PA.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29.

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

Bates, D. M., Maechler, M., & Dai, B. (2008). Lme4: Linear mixed-effects models using S4 classes.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–11.

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.

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

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.

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*, 20120394.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2016). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*.

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: An eye-tracking study. Are 18-month-old late talkers really different from their typical peers? *Journal of Communication Disorders*, *58*, 43–157.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*(4), 516–547.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic

processing. *Psychological Science : A Journal of the American Psychological Society / APS*, *13*(5), 425–430.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135.

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

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.

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.

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.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2016). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning Memory and Cognition*.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic*

Approaches to Bilingualism, *4*(2), 257–282.

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

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive procesing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127.

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.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language Cognition & Neuroscience*, *31*(1), 32–59.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, *44*(4), 978–990.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

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.

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.

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.

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314–329.

Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*(6), 464–496.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Siences*, *36*, 329–392.

R Development Core Team. (2015). A language and environment for statistical computing. R Foundation for Statistical Computing.

Saslow, M. G. (1967). Latency of saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1030–1033.

Segalowitz, N., & Hulstijn, J. H. (2009). Automaticity in bilingualism and second language learning. In *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 371–388).

Slevc, L. R., & Novick, J. M. (2013). Memory and cognitive control in an integrated theory of language processing. *The Behavioral and Brain Sciences*, *36*(4), 373–4.

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

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated
    and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and
    Language*, *50*(3), 289–335.

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

Footnotes

[1] In another experiment that is not reported here, we used the same visual stimuli with colour terms in order to increase the time available to make predictions. Therefore, the pictures are always coloured. No colour terms were used in the current experiments, but they were used in the predictability pre-test.

[2] We selected this time window because the shortest duration of the target nouns was 574 ms, and a longer time window would include a time in which participants were seeing a blank or a recall screen, depending on the load condition.

[3] We also computed the same correlation using a time window from 250 ms before the target noun onset until 200 ms after the target noun onset, considering a time lag for people to make saccadic eye movements in response to a stimulus (Saslow, 1967). The correlation using this time window was not significant either, $r(22) = .012$, $p > .9$.

[4] For the same reason as Footnote 2, we computed the same correlation using a time window from 250 ms before the target noun onset until 200 ms after the target noun onset. The correlation using this time window was not significant either, $r(22) = -.097$, $p > .6$.

[5] One participant did not provide this information.

[6] One participant did not provide this information.

[7] We computed this correlation using another time window that included 200 ms after the target noun onset as well, following Footnotes 2 and 3. In this analysis either, the extent of prediction in L2 participants did not correlate with their self-rated proficiency scores (no-load condition, $r(22) = -.016$, load condition, $r(22) = .19$, ps $> .3$), with their lengths of stay in the UK (no-load condition, $r(21)^7 = .16$, load condition, $r(22) = .23$, ps $> .2$), or with their length of exposure to English (no-load condition, $r(22) = -.22$, load condition, $r(21)^7 = .082$, ps $> .3$).

[8] People may not encode spatial location robustly in naturalistic tasks (e.g., Tanenhaus, Magnuson, Dahan, & Chambers, 2000). However, participants saw only four objects in our experiment (as in most visual world studies). In fact, the occurrence of eye movements to the appropriate object indicated that they had typically remembered its location (see Huettig, Olivers, & Hartsuiker, 2011).