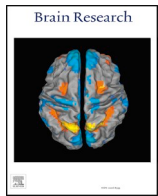




ELSEVIER

Contents lists available at ScienceDirect

Brain Research

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

Research report

Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing

Falk Huettig^{a,*}, Ernesto Guerra^b^a Max Planck Institute for Psycholinguistics, The Netherlands^b Center for Advanced Research in Education-CIAE, Universidad de Chile, Chile

HIGHLIGHTS

- Three eye-tracking experiments explored limits to prediction in language processing.
- Speech rates, visual context preview time, and participant instructions were manipulated.
- A normal speech rate only afforded prediction if participants had an extensive preview.
- Even explicit instructions to predict led only to a small anticipation effect with a normal speech rate and short preview.
- These findings are problematic for theoretical proposals that assume that prediction pervades cognition.

ARTICLE INFO

Keywords:

Prediction

Language processing

Eye-tracking

ABSTRACT

There is a consensus among language researchers that people can predict upcoming language. But do people always predict when comprehending language? Notions that “brains ... are essentially prediction machines” certainly suggest so. In three eye-tracking experiments we tested this view. Participants listened to simple Dutch sentences (“Look at the displayed bicycle”) while viewing four objects (a target, e.g. a bicycle, and three unrelated distractors). We used the identical visual stimuli and the same spoken sentences but varied speech rates, preview time, and participant instructions. Target nouns were preceded by definite gender-marked determiners, which allowed participants to predict the target object because only the targets but not the distractors agreed in gender with the determiner. In Experiment 1, participants had four seconds preview and sentences were presented either in a slow or a normal speech rate. Participants predicted the targets as soon as they heard the determiner in both conditions. Experiment 2 was identical except that participants were given only a one second preview. Participants predicted the targets only in the slow speech condition. Experiment 3 was identical to Experiment 2 except that participants were explicitly told to predict. This led only to a small prediction effect in the normal speech condition. Thus, a normal speech rate only afforded prediction if participants had an extensive preview. Even the explicit instruction to predict the target resulted in only a small anticipation effect with a normal speech rate and a short preview. These findings are problematic for theoretical proposals that assume that prediction pervades cognition.

1. Introduction

The notion that “brains ... are essentially prediction machines” (Clark, 2013; cf. Friston, 2010) is becoming more and more influential in the cognitive and neurosciences. It is argued that prediction is a fundamental principle of human information processing and offers a “deeply unified account of perception, cognition, and action” (Clark, 2013). In the present study we test this hypothesis by focusing on the prevalence of prediction in language processing. Indeed, if prediction is the grand unifying principle of the human mind, then it must also be the

unifying principle of language processing.

A problematic issue in the literature is that the term prediction is often ill-defined or defined differently by researchers within a particular field of inquiry (e.g. psycholinguistics) or across fields (visual perception researchers for instance tend to talk about quite different things than psycholinguists when they talk about prediction). Perception researchers working within the predictive coding framework, for example, often use a statistical definition for prediction. Prediction according to this view is not necessarily about the future but refers to the absence of sufficient data about current but not yet fully observed events (De Lange et al., 2018). When we

* Corresponding author at: Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands.

E-mail address: falk.huettig@mpi.nl (F. Huettig).

<https://doi.org/10.1016/j.brainres.2018.11.013>

Received 26 June 2017; Received in revised form 8 November 2018; Accepted 9 November 2018

Available online 12 November 2018

0006-8993/ © 2018 Elsevier B.V. All rights reserved.

use the term prediction here, we refer to prediction in a psycholinguistic sense, namely prediction being strictly and straightforwardly about the future. We define prediction in language processing as any pre-activation of up-coming input (Huettig, 2015, but see Ferreira and Chantavarin (2018); Kuperberg and Jaeger, 2016, for a different view). This definition avoids making a rather arbitrary decision about what constitutes prediction and what does not.

People can predict up-coming language. To the best of our knowledge there is no contemporary language researcher who believes otherwise. The experimental evidence suggesting an important role for prediction in language processing is overwhelming coming from a variety of experimental methods such as EEG (e.g. DeLong et al., 2005; Van Berkum et al., 2005; Wicha et al., 2004), MEG (e.g. Dikker and Pykkänen, 2013; Dikker et al., 2010; Fruchter et al., 2015; Gagnepain et al., 2012; Sohoglu et al., 2012), rTMS (Lesage et al., 2012), ultrasound recordings (Drake and Corley, 2015), fMRI (e.g. Boylan et al., 2014; Willems et al., 2015), eye-tracking during reading (e.g. Boland et al., 1995; Staub and Clifton, 2006), and eye-tracking during listening (i.e. visual world eye-tracking, e.g. Altmann and Kamide, 1999; Arai and Keller, 2013; Kaiser and Trueswell, 2004; Kamide et al., 2003; Kukona et al., 2011; Lowder and Ferreira, 2016). Note that people do not necessarily predict individual words but pre-activate information at various levels of representations (e.g. semantic, syntactic, phonological, etc.) and that there are considerable individual differences in people's tendencies to pre-activate upcoming language (see Huettig, 2015; Pickering and Gambi, 2018, for discussion).

This wealth of experimental evidence and a Zeitgeist of prediction becoming increasingly important as a theoretical construct in the sciences studying the mind have resulted in a large number of theories of predictive language processing (e.g. Altmann and Mirković, 2009; Dell and Chang, 2014; Federmeier, 2007; Ferreira and Chantavarin, 2018; Hale, 2001; Hickok, 2012; Huettig 2015; Kuperberg and Jaeger, 2016; Levy, 2008; Norris et al., 2016; Pickering and Garrod, 2013; Pickering and Gambi, 2018; Van Petten and Luka, 2012). A comprehensive discussion of the various theories is beyond the remit of the current paper. An important constraint for any theory of predictive language processing and one that we partly attempt to address in the current study is how routinely people predict language.

Given the strong claims about prediction it is surprising that few studies have actually investigated this directly. Whilst there are some recent studies which suggest clear limits to prediction in language processing (Frisson et al., 2017; Luke and Christianson, 2016), the number of such studies is very small. One interpretation of the experimental evidence thus is that prediction during comprehension is ubiquitous. However, considering the known problems with getting 'null results' published (e.g. van Assen et al., 2014) and the extent of false positives and lack of replication in psychology and related sciences (e.g. Kochari and Ostarek, 2018; Simmons et al., 2011; Zwaan et al., 2018), such a conclusion may be unsafe.

What then is the evidence that people *routinely* predict when comprehending language? One reason why this question is surprisingly difficult to answer is that most previous studies on predictive language processing have used very high cloze probability sentences (i.e. sentences in which the target word is very predictable). Research with low cloze probability items is required to answer the question of the importance of prediction for language understanding (Staub, 2015). If prediction is pervasive in language processing then we should find evidence for it even with less predictable sentences. In our lab for example we have observed that practically every single participant predicts that 'cake' will be the visual referent referred to next when participants heard 'The boy will eat a ...' (cf. Altmann and Kamide, 1999). In another study in which participants could use spoken gender-marked determiners in Dutch as cue for predicting upcoming nouns (i.e. a much less predictive cue than the association or semantic overlap between 'eat' and 'cake'), however, we observed that individual differences in working memory and processing speed had a large influence on whether or to what extent participants predicted the visual target object (Huettig and Jansse, 2016).

A second problematic issue with the published research is that although several methods have been used to investigate prediction (see above), the majority of studies on predictive language comprehension have relied on only two experimental methods, namely, electrophysiological studies presenting printed words and visual world eye-tracking studies presenting visual objects. A look at the literature on prediction in language processing should at least raise some questions about the ecological validity of the core methods employed (see Huettig, 2015; Ito et al., 2017; Mani & Huettig, 2016; Nieuwland et al., 2018; for extensive discussion). Most electrophysiological studies on prediction in language processing, for instance, have presented written sentences word by word in a very slow manner far removed from a normal reading scenario.

A third problematic (and related) issue is that most experiments on prediction in language processing have interpreted effects as reflecting prediction, even though this was not the only plausible interpretation of the results. Arguably, a paradigm shift (Kuhn, 1970) has taken place in the field: results of experiments are routinely interpreted as reflecting prediction even if the experiments do not provide direct evidence for prediction and similar results have previously been interpreted within a different theoretical framework (cf. Ferreira and Chantavarin, 2018). The vast majority of electrophysiological studies for example have measured the electrophysiological marker of anticipation, for example a reduced N400 ERP component, during but not before the target word. Therefore it cannot be ruled out that many studies were in fact measuring word integration difficulties rather than prediction. In other words, readers may not have predicted proactively but instead integrated the (bottom-up) activated word meaning with its context post-lexically (cf. Ito et al., 2017). Only a very small number of EEG studies have avoided this interpretation problem by using a clever experimental manipulation measuring electrophysiological correlates of prediction before the target word (DeLong et al., 2005; Van Berkum et al., 2005; Wicha et al., 2004). Recent failed replications however also cast doubt on these studies as presenting robust evidence for predictive language processing (e.g. DeLong et al., 2005, see Nieuwland et al. (2018); and Otten and Van Berkum (2009), see Kochari and Flecken (2018)). In Nieuwland et al. (2018), for example, nine EEG labs (N = 334) could not replicate the crucial electrophysiological correlate of prediction before the target word. It is important to emphasize here that the results of Nieuwland et al. should *not* be interpreted as suggesting that the phonological form of upcoming spoken words (for instance) is never pre-activated. The high-powered studies of Nieuwland et al. and Kochari and Flecken do however indicate that the EEG evidence for *routine* pre-activation is elusive at best.

Visual world eye-tracking studies, on the other hand, present spoken language (typically a sentence) while participants concurrently view a set of visual objects that are related to the speech input. Participants may hear for instance a sentence such as 'The boy will eat a cake' while viewing a cake and some other unrelated (and uneatable) objects. In this context, adults (Altmann and Kamide, 1999) and even 2-year-olds (Mani and Huettig, 2012) show anticipatory eye movements to the cake on hearing 'eat' before hearing 'cake'. Visual world eye-tracking studies arguably therefore have found the strongest evidence supporting the notion that prediction pervades language processing. In the present study we test the pervasiveness of predictive language processing using visual world eye-tracking.

Although visual world prediction studies have provided compelling evidence for prediction, their interpretation is also not without problems. When using predictive sentences such as 'The boy will eat a cake' for instance it is hard to distinguish active forecasting mechanisms (Huettig, 2015; cf. Kukona et al., 2011) from other effects such as semantic priming (EAT-CAKE, cf. Huettig and Altmann, 2005, Altmann and Mirković, 2009). Moreover, experimenters tend to present well-articulated and fairly slow speech rates to participants. Such an experimental setting could potentially create an artificial sentence processing time lag, which could in turn be exploited by listeners for prediction. Indeed, some research suggests that listeners predict less when they are exposed to casual speech compared to when they listen to the well-articulated speech typically presented in laboratory experiments (Brouwer et al., 2013). Another issue with visual world eye-tracking studies is that the commonly-used picture preview might

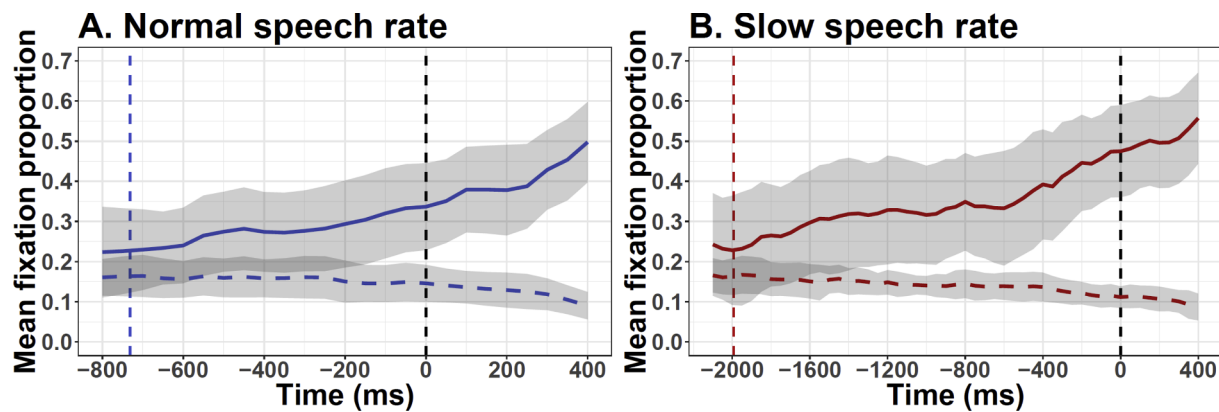


Fig. 1. Time-course plots of the proportion of fixations to target objects and average distractors in time steps of 50 in Experiment 1. Panel A. shows a 2400 ms time window from two seconds before the onset of the critical noun to 400 ms after it on slow speech rate trials. Panel B shows a 1200 ms time window, beginning from 800 ms before critical noun onset to 400 ms after it on normal speech rate trials.

provide critical scaffolding for the finding of prediction effects (Huettig, 2015; Huettig and Mani, 2016). Indeed, it has been demonstrated that printed words and visual objects can prime the retrieval of related spoken words (McQueen and Huettig, 2014). A final issue that remains unclear is to what extent participants are influenced by the instructions given by the experimenter. Previous studies strongly suggest that participants engage in prediction without being explicitly instructed to predict. It is however conceivable that not everybody engages in prediction all the time unless they are specifically told to do so (i.e. that the task is crucial, cf. Salverda et al., 2011; but see Salverda and Altmann, 2011).

In short, we identified a number of issues in the context of the investigation of prediction during sentence processing, which prevent strong conclusions from being drawn about the prevalence of prediction in language comprehension. Prediction experiments a) tend to use highly predictable sentences (high-cloze probability) rather than less predictable ones; b) produce results that are often interpreted as evidence of prediction, a conclusion that is not always warranted; and c) often use prediction-encouraging experimental set-ups (i.e., slow word presentation/speech rates, picture-preview time, instructions).

2. The present study

In the present study we conducted three visual world eye-tracking experiments to address these issues directly. We used the spoken and picture stimuli from a previous eye-tracking study on predictive language processing that found clear prediction effects (Huettig and Janse, 2016). The same spoken sentences and visual displays were used and we manipulated the speech rate presentation of the sentences within trials. Importantly, these sentences were simple instructions, without high-cloze probability scores. In addition, we varied the visual context preview time and participant instructions across experiments. Participants listened to Dutch sentences such as (translated to English) ‘Look at the displayed bicycle’ while viewing four objects (a target, e.g. a bicycle, and three unrelated distractors). The Dutch target nouns (e.g. bicycle) were preceded by definite gender-marked determiners, which allowed participants to predict the target object because only the targets but not the unrelated distractors agreed in gender with the determiner. A benefit of using determiners as cue for prediction is that with such an experimental manipulation we avoided the possibility that prediction effects may reflect simply semantic priming.

In Experiment 1, the visual context preview was set to four seconds and sentences were presented in a slow and a normal speech rate, pseudo-randomly alternated between trials. Experiment 2 was identical to Experiment 1 except that participants were given only one second as preview. Experiment 3 was identical to Experiment 2 except that participants were explicitly told to try to predict the targets. We expect these contrasting conditions within and between experiments to provide clear insights with regard to the role of situational contingencies on predictive language

processing. Finding effects of prediction in all experiments and conditions would suggest that indeed, even under adverse conditions language users are constantly engaged in anticipating incoming linguistic stimuli. However, if participants predict only under favorable (i.e. prediction-encouraging) set-ups or when they are instructed to do so, the results will raise important challenges to views of prediction as indispensable for language comprehension.

2.1. Experiment 1

2.1.1. Participants

Twenty-six native speakers of Dutch with normal or corrected-to-normal vision from the MPI for Psycholinguistics participant pool took part in the experiment. All of them gave informed consent and received monetary compensation for participation.

2.1.2. Results

Fig. 1 shows the mean fixation proportion to the target objects, the average distractors and their corresponding 95% confidence intervals in time steps of 50 ms for both speech rate conditions. Both plots are time-locked to the onset of the critical noun (vertical black dashed line at time zero). The vertical colored dashed lines mark the average onset of the determiner in the spoken sentence (and not the exact onset of it). Fig. 1A and B are presented at a different scale in the x-axis due to the large difference in trial duration between experimental conditions (see the above Materials and design section). The solid lines represent fixation proportion to the target objects and dashed lines show fixation proportion to average distractors. Grey shaded areas in the graphs represent the upper and lower bounds of the (adjusted) 95% confidence intervals. Fig. 1 reveals that the target objects were anticipated well before the acoustic onset of the target noun in both experimental conditions. In the slow speech rate and the normal speech rate trials, anticipatory eye-movements occurred rapidly after the average onset of the critical determiner (1993 ms and 732 ms before the onset of the spoken target noun, respectively).

These results reflect that participants could predict the target objects as soon as they heard the determiner in both the slow and the normal speech rate condition. In Experiment 2 we tested whether these prediction effects are contingent on the extensive preview of the visual display participants were presented with. It is possible that a substantial preview is required for participants to pre-activate possible target objects and that without such extensive preprocessing of the visual referents prediction is challenging. More so, it is also possible that both preview and speech rate play a role. In Experiment 2 we reduced the preview to one second, otherwise the experiment was identical to Experiment 1 (i.e., including again a slow and a normal speech rate condition).

2.2. Experiment 2

In Experiment 2, the materials, design and data analysis were identical to those in Experiment 1. The procedure was identical to Experiment 1 except for the preview time, which was limited to 1000 ms before the onset of the spoken sentence. All other aspects of the procedure were kept the same.

2.2.1. Participants

A new sample of 27 native speakers of Dutch from the MPI for Psycholinguistics participant pool took part in Experiment 2. All of them had normal or corrected-to-normal vision, read and signed an informed consent form and received monetary compensation for their participation.

2.2.2. Results

The results are presented in the same way as for Experiment 1. Fig. 2 reveals that target objects were anticipated well before the acoustic onset of the critical noun in the slow speech condition (Panel B.). This effect emerged at about a second before the onset of the critical noun (1000 ms after the onset of the determiner in the spoken sentence) and becomes clearly significant 400 ms before the critical is mentioned. By contrast, as can be seen in Fig. 2A, such anticipatory gaze behavior was absent in the normal speech condition. Target objects were only preferred to the average distractor after 300 ms from the acoustic onset of the target noun. In other words, participants only looked preferentially at the target object once the target noun was heard. See also Appendix A.

In Experiment 2 was identical to Experiment, except for the preview time; participants were given only a one second preview. The results are very clear. The eye-movement record reveals that with this shorter preview, participants only predicted the target objects in the slow but not in the normal speech condition. Though the difference to Experiment 1 is striking, it is conceivable that participants, in principle, could predict the target also in the normal speech condition if they were encouraged to do so. Experiment 3 tested this hypothesis. After all, in line with previous visual world prediction experiments participants were not explicitly asked to predict. Although we believe that the present visual world set-up encourages prediction in a largely automatic fashion, it is possible that there are individual differences in the extent to which participants engage spontaneously in prediction in the current task situation. In Experiment 3 we instructed participants explicitly to predict the target object. An alternative approach shown in Appendix A, lead to the same results.

2.3. Experiment 3

In Experiment 3, the materials, design and data analysis were identical to those in the first two experiments. The preview time was set to 1000 ms as in Experiment 2. The instructions, however, differed from those given in

the previous two experiments. In Experiment 3, participants were explicitly instructed to predict the object that was about to be mentioned. All other aspects of the procedure were kept the same as in Experiment 2.

2.3.1. Participants

Twenty-six new participants took part in Experiment 3. They were all native speakers of Dutch from the MPI for Psycholinguistics participant pool. All of them had normal or corrected-to-normal vision, read and signed an informed consent form and received monetary compensation for their participation.

2.3.2. Results

The results are presented in the same way as in the previous experiments. Given the 200 ms generally assumed to be necessary for programming and initiating a saccadic eye movement (e.g., Dahan et al., 2001; Martin et al., 1993; Saslow, 1967), a bias in looks to the target (compared to the averaged distractors) until approximately 200 ms after word onset most likely reflects prediction of the target (as fixation proportions until 200 ms after word onset were most likely triggered before the acoustic onset of the spoken target word). Fig. 3A reveals that participants might have anticipated the target in the normal speech trials, yet, only at the onset of the critical noun (around 200 ms before the noun was fully-fledged and accessible for the listener). By contrast, Fig. 3B shows that, as in the previews two experiments, the the slow speech condition allowed a large anticipation effect of the incoming target. Inferential analysis corroborate the results (see Appendix A).

In sum, the crucial finding from Experiment 3 is that even when participants were explicitly instructed to predict the target object, we observed at best a very small anticipation effect with a normal speech rate and the short preview.

3. Discussion

In three visual world eye-tracking experiments, we examined the effects of speech rates, preview time of the visual stimuli, and the instructions given to participants on anticipatory eye movement behavior during online processing of (low-cloze probability) spoken sentences. In these experiments, target nouns (e.g. bicycle) were preceded by definite determiners, which were gender-marked. Participants could use the gender cue to predict the target object because only the targets but not the unrelated distractors agreed in gender with the determiner. In Experiment 1, participants had a four second preview of the visual display before the spoken sentence was initiated. The results showed that participants predicted the target objects as soon as they heard the determiner in both speech rate conditions. Experiment 2 was identical except that participants were given a one second preview of the visual display before the onset of the spoken sentence. A new group of participants predicted the target objects in the slow speech but did

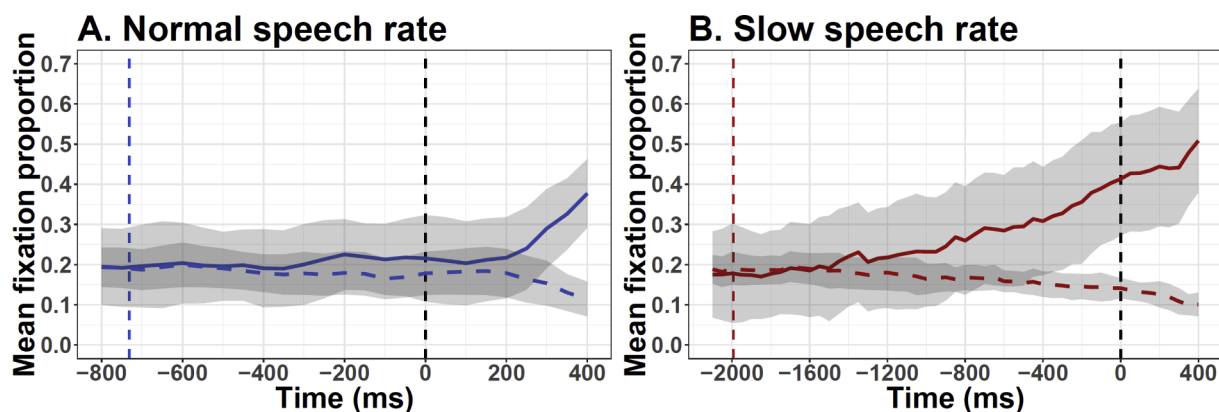


Fig. 2. Time-course plots of the proportion of fixations to target objects and average distractors in time steps of 50 in Experiment 2. Panel A. shows a 2400 ms time window from two seconds before the onset of the critical noun to 400 ms after it on slow speech rate trials. Panel B shows a 1200 ms time window, beginning from 800 ms before critical noun onset to 400 ms after it on normal speech rate trials.

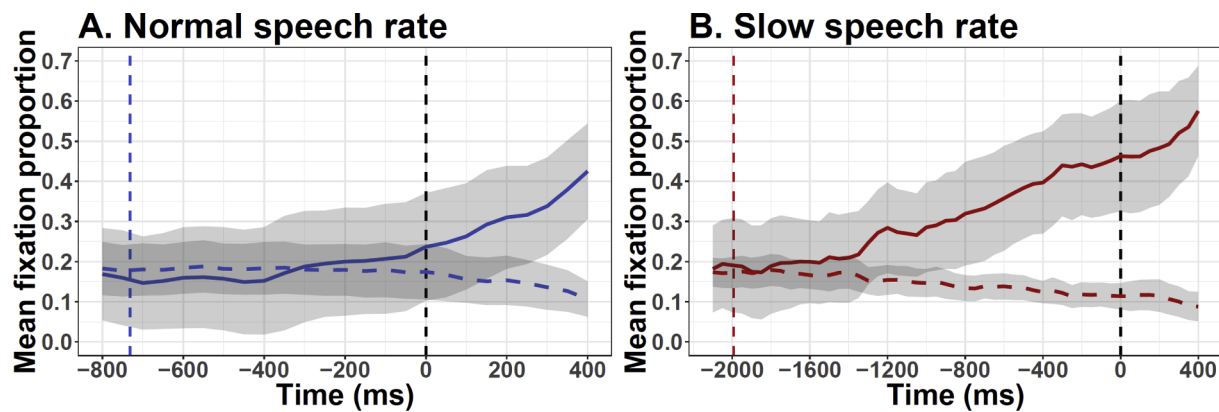


Fig. 3. Time-course plots of the proportion of fixations to critical objects and average distractors in time steps of 50 in Experiment 3. Panel A. shows a 2400 ms time window from two seconds before the onset of the critical noun to 400 ms after it on slow speech rate trials. Panel B shows a 1200 ms time window, beginning from 800 ms before critical noun onset to 400 ms after it on normal speech rate trials.

not in the normal speech condition. Experiment 3 was identical to Experiment 2 except that participants were explicitly told to predict the target objects. Participants again predicted the incoming target in the slow speech condition, but explicit instructions led only to a small prediction effect in the normal speech condition (see Appendix A for the results using mixed models).

The present results suggest that whether a language user predicts or not is contingent on the situation the comprehender finds herself in. Slow speech resulted in prediction in all three experiments. A normal speech rate however only afforded clear and substantial prediction (using gender markers) if participants had an extensive preview of the visual referents. Even the explicit instruction to predict the target object resulted in only a small anticipation effect, when a normal speech rate and a short preview were used.

We believe that these findings provide some important constraints and challenges for theoretical proposals that assume that prediction pervades language comprehension. Our results suggest that people do not anticipate appropriate referents to a similar extent when these referents are not visually co-present for a substantial time. Similarly problematic for such accounts is that the speech rate greatly influenced participants' tendency to predict. A slow speech rate (arguably a feature of many visual world experiments, but rarely explicitly acknowledged) led to prediction in all experiments but 'normal speech prediction' appeared to rely on support from the visual surroundings and /or explicit task instructions. Slow speech is of course a relatively rare occurrence in real world environments. Prediction effects during sentence processing, in other words, appear to be scaffolded by certain kinds of (non-linguistic) stimuli or certain kinds of tasks. Two important questions therefore arise therefore from the present results. To what extent is such scaffolding provided in natural conversation? How does language processing in natural conversation proceed without such scaffolding? We believe that future research must explore these questions seriously.

This leads us to possible objections to our conclusion. One may argue that we tested a very specific prediction cue, namely, the anticipation of target objects based on the gender of the determiner in the spoken instruction. We acknowledge that prediction occurs with many different prediction cues (see Huettig, 2015, for review), but as pointed out in the introduction, we believe that to make progress, research needs to move away from studying what happens when participants are presented with sentences in which the target is very predictable. If we want to test claims about the pervasiveness of prediction in language processing, it is not useful to study sentence constructions in which participant performance is at ceiling.

A second objection to our study may be that the present task situation is an atypical example of predictive language processing. However, to reiterate, a large body of the experimental evidence for the importance of prediction in language processing comes from similar set-ups. Moreover, we

do believe that exploring the nature of prediction with visual world eye-tracking experiments is particularly ecologically valid. Note that the 'prediction benefit' of the explicit task instruction was relatively modest (approximately 200 ms). This suggests that the standard set up of visual world prediction experiments taps into default processes involved in the integration of language with the visual world. Indeed, during every day interactions, prediction in language processing is often akin to choosing among several pre-activated referents and natural conversation is frequently about things in the here and now. What our study clearly shows, however, is that even with prediction-encouraging experimental set-ups such as the current one, prediction often occurs only when scaffolded by certain visual environments or certain tasks the language user may be engaged in. Note that we even inserted the word "afgebeelde", *displayed*, as padding between determiner and noun in the spoken instructions to ensure participants had in principle ample time to anticipate the target object.

A final objection to our interpretation of our (and similar) results may be to suggest that prediction always occurs but that sometimes (or even often) we fail to measure it (i.e. the experimental method was not sensitive enough to capture prediction). We acknowledge that this is a theoretical possibility but at the same time believe that such a view is not useful as it is unfalsifiable (Popper, 1959).

4. Conclusion

People often predict upcoming language. Here we have shown that speech rate, preview time, and participant task instructions all have a substantial influence on participants' tendency to engage in prediction. Our findings imply that there are many real-life situations in which people may not predict language, or what concurrent language is referring to. This contrasts with some influential accounts of human information processing that assume that prediction pervades cognition (Clark, 2013; Friston, 2010). We conjecture that it is premature to conclude that prediction is the fundamental characteristic of human cognition (cf. Frisson et al., 2017; Huettig, 2015; Huettig and Mani, 2016; Ito et al., 2017; Luke and Christianson, 2016). Further work is required to explore the how, why, what, and when of predictive language processing.

5. Method

5.1. Materials and design

Materials consisted of 38 unique visual displays and 38 accompanying spoken sentences. Each of the visual displays contained an array of four spatially distinct objects: a target object and three unrelated distractor objects. The object depictions were selected from a published set of line drawings (Severens et al., 2005). Pictures on each subset of four objects (i.e., each visual display) were matched for word frequency, picture naming time,

number of pictures, and h-statistics (see Huettig and Janse, 2016, for further detail). The corresponding sentences were simple instructions in Dutch such as “*Kijk naar de afgebeelde fiets*” (‘look at the displayed bicycle’). The instruction always referred to the target object (e.g., bicycle). Half of the instructions contained common gender words (“de”) and the other half neuter gender words (“het”), depending on the gender of the target word (i.e., a ‘de-word’ or a ‘het-word’). For example, on one trial (Fig. 4A) the target object bicycle had the common gender “de” but the three unrelated distractors (cookie, egg, ear) had neuter gender (“het”) words. Conversely, on other trials (Fig. 4B) the target was a neuter gender (“het”) word (e.g., piggybank) but the three unrelated distractors (ant, axe, backpack) had the common gender “de”. As the experiment included a speech rate manipulation, the word “*afgebeelde*” (‘displayed’), was inserted between determiner and noun in the spoken instructions to ensure participants had (in principle) ample time to anticipate the target object. Dutch native speakers judged the sentence constructions (e.g., “*Kijk naar de afgebeelde fiets*”, ‘look at the displayed bicycle’) acceptable (i.e. inserting the word ‘displayed’ did not make the sentences sound ‘strange’).

Each experimental spoken sentence had two versions; one read at a normal speech rate (mean duration_{NORMAL} = 1815.58 ms, *SD*_{NORMAL} = 145.08 ms) and another read at a slow speech rate (mean duration_{SLOW} = 4170.05 ms, *SD*_{SLOW} = 311.69 ms; mean duration_{DIFFERENCE} = 2354.47 ms, *SD*_{DIFFERENCE} = 251.88 ms). These sentences were recorded by two female native speakers of Dutch in a sound-proof room and then stored on a computer for their presentation (sample rate 44.1 kHz, 16 bit sampling resolution). In the normal speech condition, the average onset of the critical determiner occurred 732 ms before the onset of the critical noun, while in the slow speech condition the average onset of the critical determiner occurred 1993 ms before the onset of the critical noun. We crossed the two speech rate levels by creating a 2 × 1 Latin square, resulting in two experimental lists. Accordingly, each of these lists contained all critical items presented once and an equal number of trials in the slow and normal speech condition (19 items per condition). Lists were pseudo-randomized for each participant with a maximum of two consecutive trials in the same experimental condition.

5.2. Procedure

Participants’ eye movements were recorded using an SR Research EyeLink 1000 Tower mount system sampling at 1000 Hz. Before the experiment started a 9-points calibration procedure was carried out. The system was recalibrated whenever necessary. Each trial started with a fixation dot in the middle of the screen with which the experimenter could evaluate the drift on the calibration. Once participants fixated this central cue the trial was initiated by the experimenter. The visual display was presented for 4000 ms before sentence onset and throughout the sentence

until one second after the sentence offset. A central fixation dot marked the beginning of the next trial. Spoken materials were presented via computer speakers at a comfortable volume. Participants were instructed to listen to the sentence carefully. They were told that they could move their eyes freely around the visual display but were asked to not take their eyes off the monitor. The experiment lasted about 20 min.

5.3. Data analysis

Four areas of interest corresponding to location and size of each of the four pictures on the display were defined using the Experiment Builder software (SR research). A summary of participants’ fixations with their duration and coordinates on the display was produced using the Data Viewer software (SR research). Fixation proportions were calculated based on the individualization of all fixations to the target picture and to the distractor pictures. Fixation proportions were time-locked to the onset of the critical noun in the spoken sentence. Every millisecond per participant and trial was considered and a value of 1 was given to the area of interest that participants were fixating at a given moment (and a value of 0 to all other interest areas). If no fixation was detected by the eye tracker, all regions were given a 0 value. The proportion of fixation (i.e., percentage of fixation to a given object across trials) was calculated per participant and experimental condition for the four areas of interest (the target and the three distractors pictures on the display), and subsequently aggregated by object, participant, and experimental condition, as well as into 50 ms time windows. The first aggregation allows us to calculate confidence intervals (by participants), which were corrected for within-subject designs (see Cousineau and O’Brien, 2014) and for multiple comparison (see Hsu, 1996). The second aggregation help to lessen auto-correlation between fixation proportions over time.

In order to assess the effects of the gender-marked determiners and speech rates on anticipatory eye movements during sentence interpretation, we used a statistical approach that focuses on quantifiable differences between proportion of fixation, confidence intervals, the time course and size of prediction effects rather than a null hypothesis testing approach (see Cumming, 2014; Cumming and Finch, 2005; Huettig and Janse, 2016). The confidence intervals approach provides a detailed graphical description across time of the exact gaze pattern for each experimental condition showing a continuous evaluation (every 50 ms) of the difference in the effect of interest (fixation proportions) between conditions. Thus, results can be interpreted as a function of the timing of fixations towards the target objects, and the magnitude and reliability of the effects observed. In Appendix A, on reviewer request, we additionally provide an inferential analysis using mixed models in which we compared two critical time 200 ms windows around the onset of the critical word. All our data and scripts are available online at the Open Science Framework <https://osf.io/grk94/>.

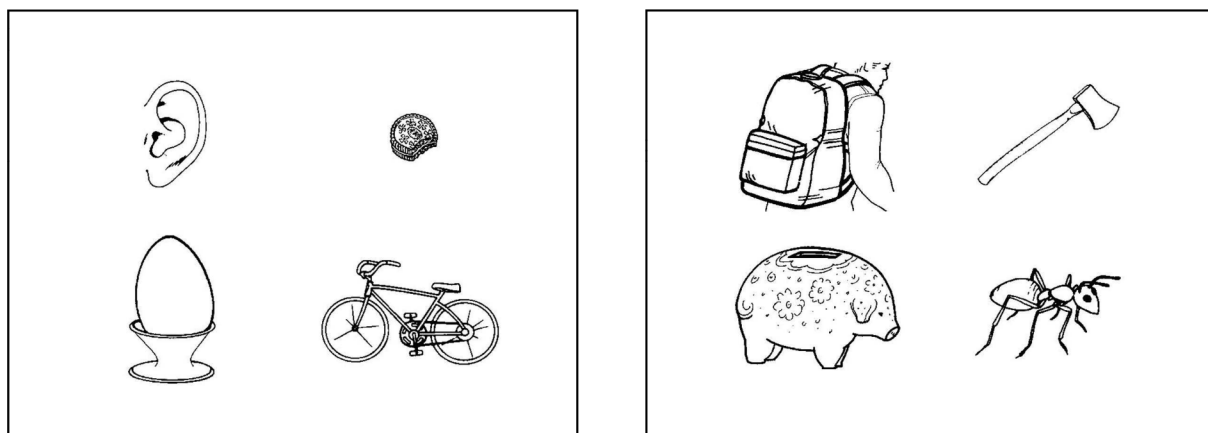


Fig. 4. Visual displays examples for a “de” only object and “het” only object trials. On the left, the target “*de fiets*” (the bicycle) is accompanied by three “het” distractor objects. On the right, the target “*het spaarvarken*” (the piggybank) is presented with three “de” distractor objects.

Appendix A

Experiment 1

In Experiment 1 participants inspected the visual context for four seconds before the onset of the critical sentence. They were instructed to pay attention to what they hear and see.

Linear mixed effects regression (LMER)

Contrasts in the LMER model are motivated by our research question: *Can we observe anticipatory eye movements in the normal speech condition?* First, the normal speech rate condition is set as reference group. Thus, the intercept is informative with regards to the overall preference for the target or the average distractors; if the intercept is significantly above or below zero, it means that there is reliable preference for the target object or the average distractor, respectively. If the intercept is not different from zero, it means that there is no preferences between the target object and the average competitor. In addition, the successive difference contrast of time blocks evaluates the changes over time (specifically comparing the 200 ms before and after the onset of the critical word) of the the empirical logit difference between target and average distractor for the normal speech condition alone.

```
# sliding contrast
contrasts(diff_data$time) <- contr.sdif(2)

# maximal model
m <- lmer(diff ~ condition + time + (1 + condition + time | subj) + (1 + condition +
time | trial), diff_data)
summary(m)

## Linear mixed model fit by REML ['lmerMod']
## Formula: diff ~ condition + time + (1 + condition + time | subj) + (1 +
##   condition + time | trial)
##   Data: diff_data
##
## REML criterion at convergence: 11958.7
##
## Scaled residuals:
##   Min       1Q   Median       3Q      Max
## -2.6680 -0.7222 -0.1449  0.7809  2.9099
##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
##   subj        (Intercept)         4.365692 2.08942
##              conditionslow      2.319271 1.52292  -0.39
##              time2-1             0.001226 0.03501  -0.93  0.02
##   trial       (Intercept)         1.309452 1.14431
##              conditionslow      0.664557 0.81520  0.79
##              time2-1             0.005450 0.07382  0.42 -0.22
## Residual                    23.181488 4.81472
## Number of obs: 1976, groups:  subj, 26; trial, 19
##
## Fixed effects:
##
##              Estimate Std. Error t value
## (Intercept)    2.1936    0.5102    4.300
## conditionslow  1.8779    0.4136    4.540
## time2-1        0.5344    0.2174    2.458
##
## Correlation of Fixed Effects:
##              (Intr) cndtns
## conditinslw -0.154
## time2-1     -0.007 -0.007
```

The results of the LMER model shows a significant intercept ($Beta = 2.19$, $se = 0.51$, $t = 4.30$), a significant difference between the normal and slow speech conditions ($Beta = 1.87$, $se = 0.41$, $t = 4.54$) and a significant difference between the two time windows ($Beta = 0.53$, $se = 0.21$, $t = 2.45$). Since we use the normal condition as reference group, the intercept reflects that overall (i.e., between -200 ms before, and 200 ms after the onset of the critical word) the empirical logit difference between target and average distractor is above zero, reflecting a preference for the target. The difference between the intercept and the slow condition reflects that this preference is larger in the slow speech rate condition relative to the normal speech rate condition. Finally, the reliable difference between time windows, suggest a relatively rapid increase in the normal speech condition- These patterns can be observed in the bar graph below Fig. A1.

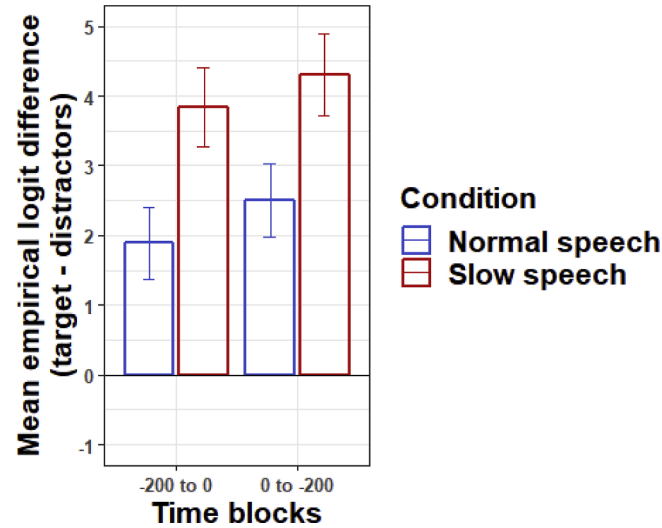


Fig. A1. Mean empirical logit differences between target and average distractor aggregated into two time blocks (200 ms before the onset of the critical noun to onset, and critical word onset to 200 ms after) in Experiment 1.

Experiment 2

In Experiment 2 participants inspected the visual context for one second before the onset of the critical sentence. They were instructed to pay attention to what they hear and see. The same processing procedures described above apply to Experiment 2, however, we set the preview time to a 1000 ms.

Linear mixed effects regression (LMER)

Same contrasts in the LMER model as in Experiment 1. The normal condition is set as reference group, and a successive difference contrast of time

blocks evaluates the changes over time of the the empirical logit difference between target and average distractor for the normal speech condition alone.

```
# sliding contrast
contrasts(diff_data$time) <- contr.sdif(2)

# maximal model
m <- lmer(diff ~ condition + time + (1 + condition + time |subj) + (1 + condition +
time |trial), diff_data)
summary(m)

## Linear mixed model fit by REML ['lmerMod']
## Formula: diff ~ condition + time + (1 + condition + time | subj) + (1 +
##   condition + time | trial)
## Data: diff_data
##
## REML criterion at convergence: 12222
##
## Scaled residuals:
##   Min       1Q   Median       3Q      Max
## -2.8407 -0.6656 -0.2895  0.7601  3.0327
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## subj    (Intercept)          4.00116  2.0003
##          conditionslow      2.67319  1.6350   0.11
##          time2-1             0.02495  0.1580   0.97 -0.15
## trial   (Intercept)          0.09595  0.3098
##          conditionslow      0.62856  0.7928   1.00
##          time2-1             0.03946  0.1987   1.00  1.00
## Residual                    21.18257  4.6025
## Number of obs: 2052, groups:  subj, 27; trial, 19
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    0.5219    0.4170   1.252
## conditionslow  2.5839    0.4164   6.205
## time2-1        0.2013    0.2105   0.957
##
## Correlation of Fixed Effects:
##              (Intr) cndtns
## conditinslw  0.035
## time2-1     0.166  0.079
```

The results of the LMER model show an intercept not significantly different from zero ($Beta = 0.52$, $se = 0.42$, $t = 1.25$) and a significant difference between the normal and slow speech conditions ($Beta = 1.68$, $se = 0.41$, $t = 4.16$). The difference between conditions reflect that, while no preference is observed in the normal condition, an overall reliable preference for the target is observed in the in the slow speech condition. Finally, no reliable effect of time window is observed ($Beta = 0.20$, $se = 0.21$, $t = 0.96$), suggesting no increase in the normal speech condition within the 200 ms after critical word onset (Fig. A2).

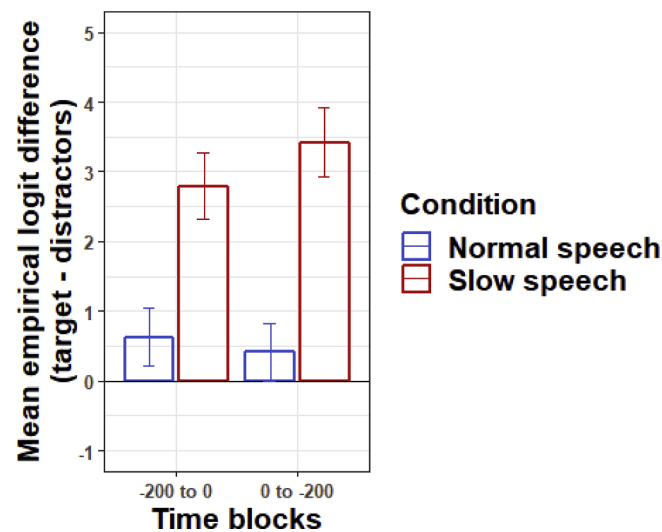


Fig. A2. Mean empirical logit differences between target and average distractor aggregated into two time blocks (200 ms before the onset of the critical noun to onset, and critical word onset to 200 ms after) in Experiment 2.

Experiment 3

In Experiment 3 participants inspected the visual context for one second before the onset of the critical sentence. By contrast to the previous two experiment, they were instructed to attempt to anticipate the object that would be mentioned in the sentence. The same processing procedures described for Experiment 2 apply to Experiment 3, with the preview time set to a 1000 ms.

Linear mixed effects regression (LMER)

Same contrasts in the LMER model as in Experiment 1 and 2. The normal condition is set as reference group, and a successive difference contrast of time blocks evaluates the changes over time of the the empirical logit difference between target and average distractor for the normal speech

condition alone.

```
# sliding contrast
contrasts(diff_data$time) <- contr.sdif(2)

# maximal model
m <- lmer(diff ~ condition + time + (1 + condition + time |subj) + (1 +
condition + time |trial), diff_data)
summary(m)

## Linear mixed model fit by REML ['lmerMod']
## Formula: diff ~ condition + time + (1 + condition + time |subj) + (1 +
## condition + time |trial)
## Data: diff_data
##
## REML criterion at convergence: 11913.5
##
## Scaled residuals:
##   Min       1Q   Median       3Q      Max
## -2.7515 -0.7272 -0.1835  0.7677  2.9541
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## subj    (Intercept)          8.11015  2.8478
##         conditionslow      5.80275  2.4089  -0.58
##         time2-1            0.06979  0.2642   0.44  0.48
## trial   (Intercept)          0.11372  0.3372
##         conditionslow      0.31454  0.5608   1.00
##         time2-1            0.02261  0.1504   1.00  1.00
## Residual                    22.64358  4.7585
## Number of obs: 1976, groups:  subj, 26; trial, 19
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    0.8397    0.5838   1.438
## conditionslow  3.1294    0.5344   5.856
## time2-1        0.5418    0.2230   2.430
##
## Correlation of Fixed Effects:
##              (Intr) cndtns
## conditinslw -0.532
## time2-1     0.118  0.135
```

The results of the LMER model show an intercept not significantly different from zero ($Beta = 0.84$, $se = 0.58$, $t = 1.44$) and a significant difference between the normal and slow speech conditions ($Beta = 3.13$, $se = 0.53$, $t = 5.86$). The non-significant intercept reflects that overall there is no preference for the target or the average distractor. However, we found a reliable main effect of time window ($Beta = 0.54$, $se = 0.22$, $t = 2.43$), suggesting that in the 200 ms following the onset of the critical word, there is an increase in the preference for the target normal speech condition. This pattern can be observed in the bar graph below [Fig. A3](#).

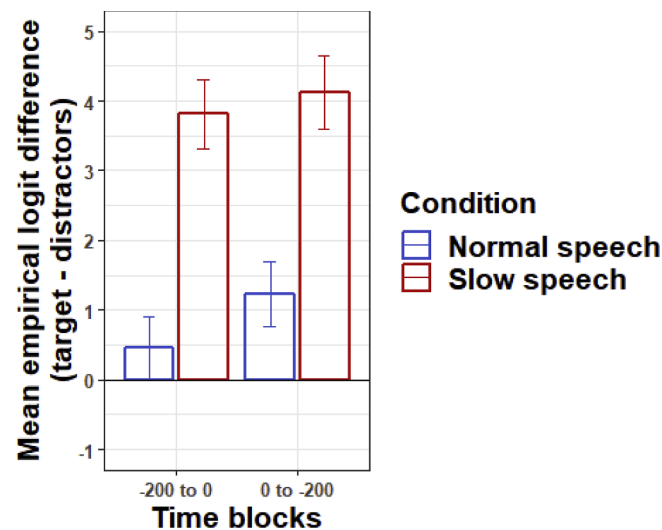


Fig. A3. Mean empirical logit differences between target and average distractor aggregated into two time blocks (200 ms before the onset of the critical noun to onset, and critical word onset to 200 ms after) in Experiment 3.

References

- Altmann, G.T.M., Kamide, Y., 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264.
- Altmann, G.T., Mirković, J., 2009. Incrementality and prediction in human sentence processing. *Cognit. Sci.* 33 (4), 583–609.
- Arai, M., Keller, F., 2013. The use of verb-specific information for prediction in sentence processing. *Lang. Cognit. Process.* 28 (4), 525–560.
- Boland, J.E., Tanenhaus, M.K., Garnsey, S.M., Carlson, G.N., 1995. Verb argument structure in parsing and interpretation: Evidence from wh-questions. *J. Memory Lang.* 34 (6), 774–806.
- Boylan, C., Trueswell, J.C., Thompson-Schill, S.L., 2014. Multi-voxel pattern analysis of noun and verb differences in ventral temporal cortex. *Brain Lang.* 137, 40–49.
- Brouwer, S., Mitterer, H., Huettig, F., 2013. Discourse context and the recognition of reduced and canonical spoken words. *Appl. Psycholinguist.* 34, 519–539.
- Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204.
- Cousineau, D., O'Brien, F., 2014. Error bars in within-subject designs: a comment on Baguley (2012). *Behav. Res. Methods* 46, 1149–1151.
- Cumming, G., 2014. The new statistics: why and how. *Psychol. Sci.* 25, 7–29.
- Cumming, G., Finch, S., 2005. Inference by eye: confidence intervals and how to read pictures of data. *Am. Psychol.* 60 (2), 170–180.
- Dahan, D., Magnusson, J.S., Tanenhaus, M.K., Hogan, E.M., 2001. Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Lang. Cognit. Process.* 16, 507–534.
- De Lange, F.P., Heilbron, M., Kok, P., 2018. How do expectations shape perception? *Trends Cognit. Sci.* 22 (9), 764–779.
- Dell, G.S., Chang, F., 2014. The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Phil. Trans. R. Soc. B* 369 (1634), 20120394.
- DeLong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8 (8), 1117–1121.
- Dikker, S., Pylkkänen, L., 2013. Predicting language: MEG evidence for lexical pre-activation. *Brain Lang.* 127 (1), 55–64.
- Dikker, S., Rabagliati, H., Farmer, T.A., Pylkkänen, L., 2010. Early occipital sensitivity to syntactic category is based on form typicality. *Psychol. Sci.* 21 (5), 629–634.
- Drake, E., Corley, M., 2015. Articulatory imaging implicates prediction during spoken language comprehension. *Memory Cognition* 43 (8), 1136–1147.
- Federmeier, K.D., 2007. Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44 (4), 491–505.
- Ferreira, F., Chantavarin, S., 2018. Integration and prediction in language processing: a synthesis of old and new. *Curr. Direct. Psychol. Sci.*
- Frisson, S., Harvey, D.R., Staub, A., 2017. No prediction error cost in reading: evidence from eye movements. *J. Memory Lang.* 95, 200–214.
- Friston, K.J., 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.
- Fruchter, J., Linzen, T., Westerlund, M., Marantz, A., 2015. Lexical preactivation in basic linguistic phrases. *J. Cognit. Neurosci.* 27 (10), 1912–1935.
- Gagnepain, P., Henson, R.N., Davis, M.H., 2012. Temporal predictive codes for spoken words in auditory cortex. *Curr. Biol.* 22 (7), 615–621.
- Hale, J. (2001, June). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1–8). Association for Computational Linguistics.
- Hickok, G., 2012. Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* 13 (2), 135–145.
- Huettig, F., 2015. Four central questions about prediction in language processing. *Brain Res.* 1626, 118–135.
- Huettig, F., Janse, E., 2016. Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Lang., Cognit. Neurosci.* 31, 80–93.
- 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, B23–B32.
- Huettig, F., Mani, N., 2016. Is prediction necessary to understand language? Probably not. *Lang., Cognit. Neurosci.* 31, 19–31.
- Hsu, J., 1996. *Multiple Comparisons: Theory and Methods*. Chapman and Hall, London.
- Ito, A., Martin, A.E., Nieuwland, M.S., 2017. How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Lang., Cognit. Neurosci.* 8, 954–965.
- Kaiser, E., Trueswell, J.C., 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94 (2), 113–147.
- Kamide, Y., Scheepers, C., Altmann, G.T., 2003. Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *J. Psycholinguistic Res.* 32 (1), 37–55.
- Kochari, A., Flecken, M., 2018. Lexical prediction in language comprehension: a replication study of grammatical gender effects in Dutch. *Lang., Cognit. Neurosci.*
- Kochari, A.R., Ostarek, M., 2018. Introducing a replication-first rule for PhD projects (commentary on Zwaan et al., 'Making replication mainstream'). *Behav. Brain Sci.* 41, e138.
- Kuhn, T.S., 1970. *The Structure of Scientific Revolutions*. University of 'Chicago Press', Chicago.
- Kukona, A., Fang, S.Y., Aicher, K.A., Chen, H., Magnusson, J.S., 2011. The time course of anticipatory constraint integration. *Cognition* 119, 23–42.
- Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction in language comprehension? *Lang., Cognit. Neurosci.* 31, 32–59.
- Lesage, E., Morgan, B.E., Olson, A.C., Meyer, A.S., Miall, R.C., 2012. Cerebellar rTMS disrupts predictive language processing. *Curr. Biol.* 22 (18), R794–R795.
- Levy, R., 2008. Expectation-based syntactic comprehension. *Cognition* 106 (3), 1126–1177.
- Lowder, M.W., Ferreira, F., 2016. Prediction in the processing of repair disfluencies: evidence from the visual-world paradigm. *J. Exp. Psychol.: Learn., Memory, Cognit.* 42 (9), 1400–1416.
- Luke, S.G., Christianson, K., 2016. Limits on lexical prediction during reading. *Cognit. Psychol.* 88, 22–60.
- Mani, N., Huettig, F., 2012. Prediction during language processing is a piece of cake – but only for skilled producers. *J. Exp. Psychol.: Human Percept. Perform.* 38, 843–847.
- Martin, E., Shao, K., Boff, K., 1993. Saccadic overhead: information processing time with and without saccades. *Percept. Psychophys.* 53, 372–380.
- McQueen, J.M., Huettig, F., 2014. Interference of spoken word recognition through phonological priming from visual objects and printed words. *Attention, Percept., Psychophys.* 76, 190–200.
- Nieuwland, M.S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer, Z., Wolfsturn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D.J., Rousselet, G., Ferguson, H.J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E.M., Donaldson, D.I., Kohút, Z., Rueschemeyer, S.-A., Huettig, F., 2018. Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife* 7, e33468.
- Norris, D., McQueen, J.M., Cutler, A., 2016. Prediction, Bayesian inference and feedback in speech recognition. *Lang., Cognit. Neurosci.* 31 (1), 4–18.

- Otten, M., Van Berkum, J.J., 2009. Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Res.* 1291, 92–101.
- Pickering, M.J., Garrod, S., 2013. An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36 (4), 329–347.
- Pickering, M.J., Gambi, C., 2018. Predicting while comprehending language: a theory and review. *Psychol. Bull.* 10, 1002–1044.
- Popper, K.R., 1959. *The Logic of Scientific Discovery*. Hutchinson, London.
- Salverda, A.P., Altmann, G., 2011. Attentional capture of objects referred to by spoken language. *J. Exp. Psychol.: Human Percept. Perform.* 37 (4), 1122–1133.
- Salverda, A.P., Brown, M., Tanenhaus, M.K., 2011. A goal-based perspective on eye movements in visual world studies. *Acta Psychologica* 137 (2), 172–180.
- Saslow, M.G., 1967. Latency for saccadic eye movement. *J. Optical Soc. Am.* 57, 1030–1033.
- Severens, E., Van Lommel, S., Ratinckx, E., Hartsuiker, R.J., 2005. Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica* 119, 159–187.
- Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22 (11), 1359–1366.
- Sohoglu, E., Peelle, J.E., Carlyon, R.P., Davis, M.H., 2012. Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* 32 (25), 8443–8453.
- Staub, A., 2015. The effect of lexical predictability on eye movements in reading: critical review and theoretical interpretation. *Lang. Linguist. Compass* 9, 311–327.
- Staub, A., Clifton Jr., C., 2006. Syntactic prediction in language comprehension: evidence from either ... or. *J. Exp. Psychol.: Learn., Memory, Cognit.* 32 (2), 425–436.
- Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol.: Learn., Memory, Cognit.* 31, 443–467.
- Van Assen, M.A., van Aert, R.C., Nuijten, M.B., Wicherts, J.M., 2014. Why publishing everything is more effective than selective publishing of statistically significant results. *PLoS One* 9 (1), e84896.
- Van Petten, C., Luka, B.J., 2012. Prediction during language comprehension: benefits, costs, and ERP components. *Int. J. Psychophysiol.* 83 (2), 176–190.
- Wicha, N.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J. Cognit. Neurosci.* 16 (7), 1272–1288.
- Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., Van den Bosch, A., 2015. Prediction during natural language comprehension. *Cerebral Cortex* 26 (6), 2506–2516.
- Zwaan, R.A., Etz, A., Lucas, R.E., Donnellan, M.B., 2018. Making replication mainstream. *Behav. Brain Sci.* 41, e120.