



Is prediction necessary to understand language? Probably not

Falk Huettig & Nivedita Mani

To cite this article: Falk Huettig & Nivedita Mani (2016) Is prediction necessary to understand language? Probably not, *Language, Cognition and Neuroscience*, 31:1, 19-31, DOI: [10.1080/23273798.2015.1072223](https://doi.org/10.1080/23273798.2015.1072223)

To link to this article: <http://dx.doi.org/10.1080/23273798.2015.1072223>

 Published online: 07 Sep 2015.

 Submit your article to this journal [↗](#)

 Article views: 223

 View related articles [↗](#)

 View Crossmark data [↗](#)

 Citing articles: 2 View citing articles [↗](#)

Is prediction necessary to understand language? Probably not

Falk Huettig^{a,b} and Nivedita Mani^c

^aMax Planck Institute for Psycholinguistics, P.O. Box 310, AH, Nijmegen 6500, The Netherlands; ^bDonders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands; ^cPsychology of Language Group, University of Göttingen, Göttingen, Germany

ABSTRACT

Some recent theoretical accounts in the cognitive sciences suggest that prediction is *necessary* to understand language. Here we evaluate this proposal. We consider arguments that prediction provides a unified theoretical principle of the human mind and that it pervades cortical function. We discuss whether evidence of human abilities to detect statistical regularities is necessarily evidence for predictive processing and evaluate suggestions that prediction is necessary for language learning. We point out that not all language users appear to predict language and that suboptimal input makes prediction often very challenging. Prediction, moreover, is strongly context-dependent and impeded by resource limitations. We also argue that it may be problematic that most experimental evidence for predictive language processing comes from “prediction-encouraging” experimental set-ups. We conclude that languages can be learned and understood in the absence of prediction. Claims that all language processing is predictive in nature are premature.

ARTICLE HISTORY

Received 15 December 2014
Accepted 7 July 2015

KEYWORDS

Cognition; language processing; learning; prediction; predictive coding

Recently, there has been a wealth of research on the importance of prediction for language comprehension (e.g. Altmann & Mirković, 2009; Dell & Chang, 2014; Federsmeier, 2007; Huettig, 2015; Kutas, DeLong, & Smith, 2011; Pickering & Garrod, 2007, 2013). Many researchers explicitly or implicitly appear to support the notion that prediction is necessary to understand language (in line with recent proposals that prediction is a or *the* fundamental principle of human information processing, e.g. Clark, 2013; Friston, 2010). Here, we examine whether the role of prediction in language processing has been overstated. Indeed, many linguists (especially within the generative linguistics framework) have traditionally argued that prediction plays no or a minor role in language understanding because language users can select words from a vast number of possibilities (e.g. Jackendoff, 2007). We would like to make clear at the outset of this article that we are in favour of an intermediate view. We suggest that prediction contributes to understanding in many situations because it provides a “helping hand” for dealing with specific situations. Language understanding, we conjecture, however, does *not* always involve prediction and as such is not necessary for language processing. Languages can be learnt and understood in absence of prediction. We will restrict our discussion to prediction in language processing but will draw on evidence from non-linguistic research on prediction when relevant. Our conclusions are (naturally)

conclusions about prediction in language understanding and not necessarily relevant for prediction, for instance, in object recognition or perception and action research. Our arguments, however, are relevant more generally to cognitive research whenever the claim is made that prediction is necessary for cognition. If prediction is the grand unifying principle of the human mind, then, of course it must also be the unifying principle of language processing. We first critically discuss potential arguments why prediction may be necessary for language understanding. We then provide arguments that prediction provides a “helping hand” but is not necessary for language processing. Finally, we discuss potential avenues for future research on this issue.

1. Potential arguments that prediction is necessary for language processing

1.1. Prediction provides a unified theoretical framework for the cognitive sciences

Theorists such as Andy Clark (2013) have proposed that “brains ... are essentially prediction machines”. He argues that prediction “offers a distinctive account of neural representation, neural computation, and the representation relation itself” and a “deeply unified account of perception, cognition, and action”. However, it is worth questioning whether we really need a deeply

unified principle underlying all functioning of the human mind? While Occam's razor may support such unification, some scholars disagree and – so do we. In a commentary to Clark's article, Anderson and Chemero (2013) argue that there can be no grand principle of brain functioning because a complex organ such as the brain almost certainly uses a diverse set of principles. Sloman (2013) points out that people, young children in particular, often focus on extending competences and engage in learning by exploration rather than prediction. There are also classic effects in the attention literature (e.g. Carrasco, Ling, & Read, 2004) for which a predictive framework makes either false predictions or offers no explanation (see Block and Siegel, 2013, for discussion; and Bowman, Filetti, Wyble, & Olivers, 2013, for a similar point). Finally, Rasmussen and Eliasmith (2013) point out that Clark's unified framework lacks too many implementational details and architectural commitments to be evaluated seriously. The latter point, we suggest, is particularly critical. Clark's general framework about prediction remains to be tested thoroughly (theoretically as well as empirically) and is currently too underspecified for it to be a convincing argument that prediction may also be necessary for language processing.

1.2. Prediction pervades cortical function

Does prediction have a neural base which pervades cortical function? Many neuroscientists and theorists from related disciplines would answer this question with a resounding "yes". Karl Friston (e.g. 2010), for instance, argues that the brain is fundamentally engaged in predictive coding and computes prediction errors, which are assumed to bias our minds towards making correct inferences. According to Friston, predictive coding involves the minimizing of prediction error through recurrent or reciprocal interactions among levels of cortical hierarchy. Higher hierarchical levels are thought to create forward models of lower level (cortical or subcortical) activity. Importantly, lower level activity is assumed to only contain the prediction error (often called the "surprisal", i.e. the extent to which the predictions are disconfirmed) between predicted activity and actual activity at lower levels. The prediction error is supposed to be used to update the forward models of lower level cortical activity.

The idea of predictive coding has become increasingly popular over recent years also among language researchers (e.g. Farmer, Brown, & Tanenhaus, 2013; Gagnepain, Henson, & Davis, 2012; Lewis & Bastiaansen, *in press*; Willems, Frank, Nijhof, Hagoort, & van den Bosch, *in press*; cf. Chang, Dell, & Bock, 2006; Kleinschmidt & Jaeger, *in press*). It is important to note here,

however, that experimental evidence that our brains engage in predictive coding during language processing is very sparse. This may well be because the neuroscientific methods available today have important limitations and are (currently still) ill-suited to address this question. One interesting proposal is that oscillatory activity during language processing provides a measure of such predictive coding. Alpha and beta oscillations are thought to index top-down processing whereas gamma oscillations are presumed to index bottom-up processing (Bastos et al., 2012; Wang, 2010). More concretely, Friston, Bastos, Pinotsis, and Litvak (2015) appear to suggest that alpha and beta oscillatory activity reflects the forward models of lower level (cortical or subcortical) activity (i.e. the predictions), whereas gammaoscillatory activity indicates processing of prediction errors to update the predictions (see also Bressler & Richter, 2015; Engel & Fries, 2010; Lewis & Bastiaansen, *in press*, for similar proposals).

What evidence is there that these assumptions are correct? Prediction could potentially be involved in syntactic unification operations (cf. Hagoort, 2005, 2013). There are indeed some studies that have found higher power in the beta frequency range in syntactically correct sentences than in sentences containing syntactic violations (e.g. Bastiaansen, Magyari, & Hagoort, 2010; Davidson & Indefrey, 2007; Kiehl, Meltzer, Moreno, Alain, & Bialystok, 2014). This is consistent with the notion that beta oscillations indicate syntactic unification providing a potential link between beta oscillations and syntactic prediction. It has also been observed that semantic violations result in lower power in the beta frequency range relative to semantically correct sentences (e.g. Kiehl et al., 2014; Luo, Zhang, Feng, & Zhou, 2010; Wang et al., 2012) consistent with the explanation that beta oscillations are linked to predictions. However, the direction of observed oscillatory activity appears to be sometimes inconsistent. Some studies, for instance, have found higher power in the gamma frequency range for highly predictable words than for semantically anomalous words (e.g. Hald, Bastiaansen, & Hagoort, 2006; Penolazzi, Angrilli, & Job, 2009; Rommers, Dijkstra, & Bastiaansen, 2013) whereas others have found higher gamma power for world knowledge *violations* and no increase in gamma oscillations for semantically correct sentences (Hagoort, Hald, Bastiaansen, & Petersson, 2004).

We acknowledge that there have been attempts to explain these divergent sets of findings (i.e. the differences in the direction and nature of alpha, beta, and gamma oscillations) within a predictive coding framework (e.g. Lewis & Bastiaansen, *in press*). Moreover, we cannot rule out that future research will provide

evidence that oscillatory activity is related to predictive coding (the line of work by Poeppel and colleagues, e.g. Giraud & Poeppel, 2012, for instance looks promising to us) but we believe that it is fair to say that the currently available experimental evidence does not provide particularly strong support that prediction pervades cortical function at least as far as language processing is concerned.

1.3. Humans are adept in detecting sequential statistical regularities in language input

Connectionist approaches to structure extraction have provided compelling accounts that language learners are skilful in detecting statistical relationships in language input. In Elman (1990), for instance, information about the distributional constraints on the context in which particular chunks co-occur causes the network to learn representations that correspond to syntactic and semantic categories. This could be interpreted as the network learning from errors in its own predictions to approximate the conditional probabilities of successive chunks within the input. Importantly, it has been demonstrated that recurrent networks are able to encode long-distance dependencies, which occur, for example, in *wh*-questions and relative clauses.

Indeed, even very young language learners are skilful in detecting statistical relationships in the input. Core evidence comes from studies examining infant learning of statistical dependencies in the input (see Romberg & Saffran, 2010, for a review). For instance, Saffran, Aslin, and Newport (1996), presented eight-month-olds with a continuous spoken sequence of trisyllabic words from a nonsense language (e.g. *pabikutibudogolatudarpitibudodaropi ...*). Note that the only cues that could be used to segment the words and detect the boundaries between words in the sequence were differences in the transitional probabilities of the syllables between and within words, i.e. pairs of syllables within words co-occur more often together relative to syllable-pairs spanning word boundaries. Saffran and colleagues found that eight-month-olds were able to calculate transitional statistics with regard to the frequency of syllable co-occurrences and use these statistics to segment continuous speech streams without explicit acoustic cues to the boundaries between words in the input. These results could be interpreted as indexing infants' prediction of one syllable upon hearing another syllable based on the high frequency of these syllables co-occurring together in their previous experience. Alternatively, these results could also be interpreted as indexing the ease of infants' recognition of frequently co-occurring syllables, independent of any prediction-based processing.

Using eye-tracking in reading McDonald and Shillcock (2003a, 2003b) also suggested that readers make use of statistical knowledge in the form of transitional probabilities, i.e. the likelihood of two words occurring together. They presented some evidence that transitional probabilities between words influence fixation durations. Frisson, Rayner, and Pickering (2005) replicated the findings of McDonald and Shillcock (2003a, 2003b) in a first experiment but, in their second experiment, when items were matched for Cloze values, no effect of transitional probabilities was found. Frisson et al. concluded that low level transitional probabilities do not explain prediction above "regular" predictability effects typically determined by the use of a Cloze task. Moreover, according to many statistical learning accounts there should be interactive effects of frequency and predictability, i.e. predictability effects should be larger for low frequency than for high frequency words (Levy, 2008; McDonald & Shillcock, 2003a, 2003b; Norris, 2006). In other words, reading a low frequency word in a context in which it is highly expected should be easier whereas reading a high frequency word in a predictive context should result in less of a benefit (since it is quite likely to occur anyhow). A great number of studies have failed to find such a significant frequency and predictability interaction nor do they report any consistent trends (Altarriba, Kroll, Sholl, & Rayner, 1996; Ashby, Rayner, & Clifton, 2005; Gollan et al., 2011; Hand, Miellet, O'Donnell, & Sereon, 2010; Kennedy, Pynte, Murray, & Paul, 2013; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Ashby, Pollatsek, & Reichle 2004; Kretzschmar, Schlesewsky, & Staub, *in press*; Staub, 2011; Staub & Benatar, 2013; Whitford & Titone, 2014).

However, some evidence of a link between the extraction of statistical regularities and language prediction comes from studies showing that performance in a statistical learning task correlates positively with sensitivity to word predictability when perceiving degraded spoken sentences (Conway, Bauernschmidt, Huang, & Pisoni, 2010; see also Misyak, Christiansen, & Tomblin, 2010). Despite this kind of correlational evidence that individuals who are good at detecting statistical relationships in implicit learning tasks are also good at predicting language input, there is as far as we know currently no *direct* experimental evidence available that unequivocally links the detection of sequential statistical regularities to mechanisms of predictive language processing.

Finally, and more generally, there is evidence that random input can lead to the formation of better representations of items than regular input. Tremblay, Baroni, and Hasson (2013) presented participants with long series of four distinct bird chirps, which were concatenated either randomly or following strong transitional

constraints. Participants' task was to report the number of unique chirps they could hear in the input. Participants performed much better when hearing the random series (a mean of approximately 4) than when hearing the regular series (a mean of approximately 3.5). Especially given recent findings that sharper representations support prediction (cf. Mani & Huettig, 2014), the findings of Tremblay et al. raise questions as to the nature of the linkage between representation detail, the extraction of statistical regularities and prediction. In other words, if prediction is enhanced by the robustness of the representations involved, and if the representations formed in learning from statistically regular input are less robust, then there may not be as strong a link between the extraction of sequential regularities and prediction in language processing after all.

1.4. Without prediction there would be no learning

Even if one were to accept that prediction may underlie infants' learning of forward statistical regularities, the fact that prediction may play an important role in language learning does not necessitate that language learning *always* involves prediction. Though Elman (2009) has argued that predictive dependencies play an important role for language learning, he has also stated that "prediction is not the major goal of the language learner" (Elman, 1990, p. 193). Others, however, appear to go so far as to claim that without prediction no language learning would be possible.

Chang, Dell, and Bock (2006, see also Kidd, 2012; Rowland, Chang, Ambridge, Pine, & Lieven, 2012), for instance, argue that "abstraction occurs because prediction occurs" (cf. Bates & Carnevale, 1993; Elman, 1991; Hahn & Oaksford, 2008; Johnson, 2004; Lewis & Elman, 2001; MacWhinney, 2004; Rohde & Plaut, 1999; Seidenberg & MacDonald 1999). Chang, Kidd, and Rowland (2013) also claim that prediction in language processing is a by-product of language learning. These authors propose that language acquisition mechanisms rely on a form of error-based learning mechanisms and that this error-based learning is prediction. The dual path model of Chang, Dell, and Bock, (2006) includes a learning algorithm (the sequencing pathway, cf. Elman, 1990) which compares predicted (next) words with words that are actually uttered (i.e. production-based prediction, see also Dell & Chang, 2014; cf. Pickering & Garrod, 2013). Any mismatch (i.e. the "prediction error") is used to adjust the model's representations. In other words, learning occurs when the model predicts the next word at each point in the sentence. Chang and colleagues (Chang, Dell, & Bock, 2006; Chang, Kidd, & Rowland,

2013; see also Dell & Chang, 2014) argue that error-based learning can explain structural priming (Bock, 1986) and, importantly that "this ability requires that prediction-for-learning is constantly taking place during language comprehension" (Chang, Kidd, & Rowland, 2013). Syntactic structure is learned because the learner's syntactic representations are gradually adjusted in order to be able to predict sentences. Chang and colleagues argue that structural priming in adults occurs because these error-based learning mechanisms stay on in proficient adult language users. Prediction in adult language processing, according to this view, is a consequence of language learning.

However, it is relevant for the notion that language learning necessitates prediction that there is evidence that infants (Pelucchi, Hay, & Saffran, 2009) and adults (Perruchet & Desauty, 2008) track backward statistics in fluent speech and that backward transitional probabilities are often more informative than forward statistics (see also St. Clair, Monaghan, & Ramscar, 2009). In languages with grammatical gender such as German, backward transitional probabilities are much more informative to learn which of the articles (i.e. *der*, *die*, or *das*) precedes a noun because the noun is often paired with the article whereas the article itself is a very poor predictor of a particular noun. Similarly, Willits, Seidenberg, and Saffran (2009) have shown in corpus analyses that backward transitional probabilities in English are much more informative for learning the grammatical category "noun" than forward transitional probabilities. The tracking of backward transitional probabilities during language learning and processing is therefore a clear example of how language learning can take place in the *absence* of prediction since backward transitional probabilities cannot be used for prediction. In short, the notion that all language learning involves prediction is unlikely to be correct. Finally, even if prediction were absolutely necessary for language learning, it does not follow that prediction is necessary for language comprehension.

Indeed, Mani and Huettig (2012) found that the (linguistic) prediction skills of two-year-olds were significantly correlated with their productive vocabulary size. Children with large production vocabularies predicted upcoming linguistic input but low producers did not. Further analysis showed that children's prediction abilities were tied specifically to their production skills rather than their comprehension skills. These findings are consistent with production-based prediction but they are also consistent with the notion that language learning can occur in the *absence* of prediction since the low producers in Mani and Huettig's (2012) study showed comprehension of all the sentences in the

study. It is important to point out in this regard that no study conducted so far has *directly* tested whether children can learn new words/grammars without prediction. Future research could usefully be directed at this topic.

1.5. *There is a wealth of experimental evidence that people predict in language processing*

Last but not least it could be argued that there is a great deal of experimental evidence for prediction and that the sheer wealth of evidence for prediction in language tasks supports the notion that prediction is necessary for language understanding. We acknowledge that there is much evidence that language users predict in many situations (e.g. Altmann & Kamide, 1999; Borovsky, Elman, & Fernald, 2012; DeLong, Urbach, & Kutas, 2005; De Ruiter, Mitterer, & Enfield, 2006; Kamide, Altmann, & Haywood, 2003; Kutas & Hillyard, 1984; Mani & Huettig, 2012; Nation, Marshall, & Altmann, 2003). Most of this evidence for prediction, however, is not relevant for answering the question about the *precise importance* of prediction for language understanding. This is because the vast majority of studies on predictive language processing have used sentences in which the target word was extremely predictable, i.e. very high Cloze probability sentences (a notable exception is a recent study by Wlotko & Federmeier, 2013). Further research with low Cloze probability items is required to answer the question of whether prediction is necessary to understand language.

1.6. *Interim summary*

We have critically evaluated five potential arguments, which could (and often are) used to claim that prediction is necessary for language processing. First, we have argued that theoretical frameworks that propose that prediction provides a deeply unified principle of the functioning of the human mind (e.g. Clark, 2013) are at present too underspecified to be able to offer sufficient theoretical support for our question of interest. Second, we conjecture that the currently available experimental evidence does not provide strong support from the domain of language processing for the claim that prediction pervades cortical function. Third, findings that individuals are able to extract forward sequential regularities from speech tell us little about the extent to which such results are driven by prediction. Fourth, there is little support for the claim that prediction is absolutely necessary for language learning. Indeed, evidence for the informativity and use of backward transitional probabilities suggest that language learning (at

least partly) takes place without predictive learning. Fifth, most of the experimental studies on prediction in language processing are uninformative with regard to the question of whether prediction is *necessary* to understand language.

In contrast (as spelled out earlier), we suggest that prediction contributes to understanding in many situations because it provides a “helping hand” for dealing with specific situations. However, we conjecture that language understanding does *not* always involve prediction and as such is not necessary for language processing. Languages can be learnt and understood in the absence of prediction. We will now turn to arguments in support of this notion.

2. Arguments in line with the notion that prediction provides “a helping hand” but is not necessary for language processing

2.1. *Not everybody predicts*

One source of support for the view that prediction plays an important but not a necessary role in language processing comes from studies finding considerable variability – from no effects of prediction to weak prediction – in developing language users (both children and second language users). For instance, a number of recent studies suggest that children’s anticipation of upcoming linguistic input is strongly influenced by children’s vocabulary knowledge with differences between the studies as to whether the driving factor here is children’s comprehension (Borovsky et al., 2012; but see Nation et al., 2003) or production vocabulary size (Mani & Huettig, 2012). Variation in the amount of prediction of course does not necessarily mean absence of prediction. However, Borovsky et al. (2012) find that children with lower scores in a sentence completion task and children with lower vocabulary scores both do not fixate a related target image even in a strongly predictive context, e.g. fixate the image *ship* upon being presented with the context “The pirate chases the ...”. Relatedly, Mani and Huettig (2012) find that children with low productive vocabulary scores do not fixate a related target image *cake* in a strongly predictive context, e.g. “The boy eats the ...”.

Similarly, results from older children (Mani & Huettig, 2014) and even adult bilinguals (Martin et al., 2013) and adult illiterates (Mishra, Singh, Pandey, & Huettig, 2012) suggest that not all listeners anticipate upcoming language input, and that anticipation of upcoming language input – but crucially not language processing – is strongly modulated by other factors, such as listeners’ literacy skills (see also Huettig & Brouwer, 2015). For

instance, Martin et al. (2013) show that L2 learners do not show a prediction effect in L2 processing. Here, participants were presented with sentences containing either a predictable or an unpredictable noun at the end of the sentence. ERPs were time locked to articles (preceding the sentence-final nouns), which were either consistent or inconsistent with the sentence-final nouns. For instance, participants read the sentence “Since it is raining, it is better to go out with a/an...” where *umbrella*, the expected continuation of the sentence would be consistent with the article *an* and inconsistent with the article *a*. L2 speakers did not show an increase in the N400 to inconsistent articles, which suggests that L2 speakers may find it more difficult to use contextual cues to anticipate upcoming language input relative to native speakers.¹ Mishra et al. (2012) compared language-mediated anticipatory eye gaze to visual objects in low and high literates. On hearing the semantically and syntactically biasing adjective and well before the acoustic onset of the spoken target word, high literates started to look more at the target object than unrelated distractors. High literates shifted their eyes towards the targets approximately 1000 ms before the low literates. Low literates’ eye gaze on the targets only started to differ from looks to the unrelated distractors once the spoken target word acoustically unfolded (cf. Huettig, Singh, & Mishra, 2011). In other words, low literates used information from unfolding spoken words to direct their eye gaze (ruling out that the anticipation effect in low literates was absent due to “noise”, or that they understood the sentences in exactly the same way as the highly literate participants but somehow were less willing or able to shift their eyes to the targets), they just did not use such information for prediction.

In all these cases showing reduced or no prediction of upcoming linguistic input in certain populations (see Federmeier, Kutas, & Schul, 2010 for a similar argument based on reduced prediction in aged populations), one would not – and could not – argue that these groups of participants cannot comprehend language, i.e. extract meaning and structure from linguistic stimuli at the fast pace that language is typically presented to the listener. Indeed, Mani and Huettig (2014) explicitly examined this by testing children’s’ prediction of upcoming linguistic input and the speed and accuracy of their processing of non-predictive sentences against the background of their reading skills and found that while participants’ reading ability correlated with their prediction skills, there was no correlation between their language abilities (measured by a standard naming task in the Intermodal preferential looking paradigm, as well as a syllable detection task, a non-word reading skills task) and participants’ reading skills [see Hahne & Friederici, 2001, for similar

findings that proficient bilinguals appear to be uniquely impaired in their prediction of upcoming language input (cf. Martin et al., 2013) but not in their processing of language, per se]. Taken together, it appears that there are a wide range of participants who show either reduced or no anticipation of upcoming language input (at least according to standard prediction measures), but who are, nevertheless, competent language users, at least in comparison to their predicting peers. This would suggest that while prediction may be important to language comprehension, language comprehension does not always involve prediction. Relatedly, however, these findings could also be interpreted to suggest that participants who showed nil or reduced prediction in the studies reviewed above were *not* predicting per se but rather that they were slower to predict relative to the groups who showed more predictive language processing. Thus, were we to give such participants more time to respond, they would show similarly predictive effects in language processing relative to the other groups. However, we note that even were these participants to be delayed predictors, such a conclusion would argue against a necessary role for prediction in language processing since their language processing appears to keep up with the pace of the stimuli presented but their prediction appears to lag behind. Second, we note that Borovsky et al. (2012) find that low predictors also performed poorly in a sentence completion task where participants were asked to provide a semantically and syntactically appropriate ending of a sentence at their own pace. The poor performance of the non-predictors in this study suggests that these participants have difficulties with regard to narrowing down the choice of potential candidates that could occur in certain sentence contexts. Admittedly, such participants also have lower vocabulary sizes, which might be indicative of impaired language abilities in general – which, in turn, would suggest that one reason for their impaired language abilities is the absence of a fundamentally important predictive support system. However, a reduced vocabulary size does not automatically imply that such participants have difficulties recognizing the words they know. Thus, word recognition – at the very least – can and does proceed independently of prediction-based mechanisms.

2.2. Suboptimal input makes prediction less (rather than more) likely

Much is made of the benefit of a predictive approach, especially with regards to the processing of noisy or ambiguous input. Thus, for instance, Pickering and

Garrod (2007) suggest that prediction is a powerful tool that listeners can use especially when required to compensate for noisy input, due to strong top-down influences on interpretation in such cases. In particular, they suggest that the influence of production-based prediction mechanisms increases inversely to quality of the input. Evidence in favour of this suggestion comes from studies showing increased top-down semantic influences in the interpretation of implausible sentences in noise (e.g. Gibson, Bergen, & Piantadosi, 2013).

However, recent research suggests that, if anything, noisy or reduced speech input makes no difference or prediction even less likely. Mitterer and Russell (2013) investigated how Dutch listeners recognize past participles in which the prefix has undergone Schwa reduction. They found that full forms benefited as much from predictability as reduced forms. This result does not fit with the proposal that prediction compensates for a noisy or reduced bottom-up signal. More direct evidence was obtained in a recent study by Brouwer, Mitterer, and Huettig (2013). They observed that strongly supportive discourse context led to prediction of the target word only in sentences with well-articulated canonical word pronunciations but not in the sentences containing phonological reductions. This suggests that when listeners are exposed to casual speech containing many phonological reductions they may often be unable to predict because they are more uncertain what they have just heard. In other words, prediction can be very challenging if the input on which to base predictions is poor.

2.3. Prediction is strongly context-dependent

While there are numerous studies reporting evidence of prediction in language processing, we note that there are increasingly more studies that find considerable context dependence in language prediction. Huettig and Guerra (2015) tested this issue directly and observed that prediction effects can disappear altogether when participants are not given adequate time to view potential thematically appropriate targets beforehand. In this study, Dutch participants listened to simple sentences such as (translated to English) “Look at the displayed piano” while viewing four objects (a target, e.g. a piano, and three unrelated distractors, e.g. a plate, a pig, some paper). Target nouns (e.g. piano) 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. These sentences were presented

either in a slow or a normal speech rate. 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 only a one-second preview of the visual display before the spoken sentence. A new group of participants predicted the target objects in the slow speech but not in the normal speech condition. These 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 both experiments. A normal speech rate, however, only afforded prediction (using gender markers) if participants had an extensive preview of the visual referents. These findings are problematic for theoretical proposals that assume that prediction pervades language comprehension. We suggest that prediction is definitely an important aspect but not a necessary characteristic of language processing.

We note, however, that a potential objection to our argument is worth discussing. Namely, the absence of experimental evidence for prediction in certain populations (see Section 2.1) or in certain situations may simply reflect the fact that, due to less experience, some populations have less confidence in their predictions. The argument could be that prediction is always occurring and that the output of the “prediction system” leads to stronger (or more accurate or reliable) or weaker (or less accurate or reliable) predictions given prior context or experience. However, predictive processing needs to reach a certain threshold level before a behaviourally observable action is initiated. In contexts in which the language user can draw on no or less experience, predictions may be unreliable and thus the “prediction system” may not initiate any action. Such an explanation arguably is compatible with Bayesian implementations, which output the confidence of a prediction given the context, i.e. $P(A|B)$. According to Bayesian accounts, no action will be initiated when predictive probabilities are too weak. How could these accounts explain the data we have presented? The findings that certain populations (e.g. older adults) do not predict in certain situations could be explained by a model whose connections were slowly damaged or lesioned (Section 2.1). Consequently, predictions would become less accurate and lead to less confidence in the computed predictions. Phonological reductions in the speech input (Section 2.2, cf. Brouwer, Mitterer, & Huettig, 2013) may also reduce confidence in the predictions. Our take on these arguments is a pragmatic one. How fundamental is prediction to language processing if it is so difficult to observe in many contexts and populations? There also seems to be a problem with falsification (cf. Popper, 2014) here in that it may be impossible

to falsify accounts which postulate that prediction occurred in absence of a behavioural manifestation.

2.4. Prediction is (frequently) impeded by resource limitations

Christiansen and Chater ([in press](#)) have recently argued that processing speech input is severely limited, resulting in a “Now or Never” bottleneck. Specifically working memory capacity is assumed to shape the structure of language perception and the solutions for dealing with the problems imposed by the bottleneck. Christiansen and Chater argue that “only an incremental, predictive language system ... can deal with the onslaught of linguistic input, in the face of severe memory constraints of the now-or-never bottleneck”.

We believe that they overlook that such a bottleneck also imposes important constraints and limits on prediction in language processing. Memory constraints and sheer speed of incoming input mean that often there are simply not enough time or enough resources available for prediction to occur. Indeed, evidence suggests that predictive processing may actually be inefficient for select groups of language users. Rayner and colleagues (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006; Rayner & Clifton, 2009) have suggested that older readers adopt a riskier reading strategy than younger adult readers, with older readers skipping words more often, possibly on the basis of their guess of what the next word will be. This finding could be interpreted as indicating that older adults predict *more* than younger adults to compensate for age-related cognitive decline. On the other hand, Federmeier and colleagues repeatedly found that older adults showed *smaller* and *delayed* effects of contextual constraint compared to young adults, which was attributed to decreased reliance on predictive processing in older age (Federmeier et al., 2010; Huang, Meyer, & Federmeier, 2012; Wlotko & Federmeier, 2012). Indeed, it should be noted that the older adults in Rayner et al. (2006), in keeping with the findings reported by Federmeier and colleagues, also regressed more to earlier words. Wlotko and Federmeier (2012) speculate, in line with the argument put forward by Peelle, Troiani, Wingfield, and Grossman (2010) that older adults’ decreased predictive processing may be due to less efficient functional connectivity, or that predictive processing has become too costly or inefficient for older adults due to decreased availability of neural resources.

Similarly, Huettig and Janse ([in press](#)) show that prediction effects in language processing are modulated by individual differences in working memory and processing speed, such that participants with poorer

working memory abilities and processing speed showed decreased prediction effects relative to others. Huettig and Janse ([in press](#), see also Huettig, Olivers, & Hartsuiker, 2011) suggest that language-mediated anticipatory movements require considerable visual and spatial working memory capacities in order for participants to correctly encode and retrieve the range of possible target alternatives that then guide eye-movements in visual world prediction tasks. Further research is needed to assess the extent to which prediction in language processing is impeded by resource limitations.

2.5. Much experimental evidence comes from “prediction-encouraging experimental set-ups”

We have already highlighted the number of studies that find evidence of predictive language processing (see Section 1.6) and are convinced, especially against the background of this literature that language users often predict upcoming spoken and written language input. We question, however, the extent to which evidence of such predictive language processing is indicative of the *necessity* of prediction-based mechanisms for language acquisition and processing. This is especially so, given the kinds of tasks that are typically employed in a majority of prediction-based experiments. In particular, we refer here to the fact that the visual stimuli presented in visual world eye-tracking experiments on prediction may provide critical scaffolding for the finding of such effects. Typically such experiments present participants with images of thematically appropriate and inappropriate referents prior to the critical auditory stimuli (Huettig, Rommers, & Meyer, 2011). Recent work on the visual world paradigm suggests that children and adults alike retrieve implicitly the label of visually fixated images (in line with cascaded activation accounts, Huettig & McQueen, 2007; cf. Mani, Durrant & Floccia, 2012; Mani & Plunkett, 2010, 2011; McQueen & Huettig, 2014; Meyer, Belke, Telling, & Humphreys, 2007). Is it, therefore, worth asking whether participants anticipate thematically appropriate targets to a similar extent when these targets are not displayed in front of them?

Some visual world eye-tracking work suggests that there is prediction of upcoming linguistic input even when appropriate targets are not present in the visual display (Rommers, Meyer, Praamstra, & Huettig, 2013). Thus, Rommers et al. (2013) find that participants fixate images overlapping in shape with the intended targets, e.g. a round object, upon hearing contextually constraining sentences such as “In 1969 Neil Armstrong was the first man to set foot on the moon”, even before they hear the word “moon” and in the absence of any visual referent of the target “moon”. Nevertheless, we suggest

that while there might be a component to prediction that is uninfluenced by the visual context provided in typical prediction tasks, it is worth examining the extent to which the strength of the prediction effects reported in the literature can be attributed to the visual presentation of the appropriate target.

On the other hand, it is certainly true that during every day interactions, prediction in language processing is often akin to choosing among several pre-activated referents. Natural conversation is frequently about things in the here and now. However, in order to argue that prediction is a *necessary* characteristic of language processing, it is important that we distinguish between purely language-based prediction effects and language-mediated anticipatory eye-movements (and changes in brain activity) which may be led by the presentation of isolated thematically appropriate images.

Another issue regarding the nature of commonly used prediction tasks, especially against the background of working memory constraints on prediction discussed in Section 2.4, concerns the kind of stimuli that are typically presented to participants in such paradigms. Typically, the auditory stimuli used are perfectly articulated sentences presented in a slow speaking rate in order to allow adequate time for participants to initiate predictive eye-movements. Indeed, this is especially the case in studies with young children. Given that working memory capacities and cognitive efficiency impact prediction performance even in such ideal situations (Huettig & Janse, *in press*), it is worth questioning the extent to which prediction performance in natural conversation is impacted by working memory and processing speed abilities. This is especially so given the differences in the quality of auditory and visual input provided in natural conversations compared to the ideal prediction tasks.

Finally, note that methodological worries about the generalisability of the available evidence for predictive language processing are not restricted to visual world eye-tracking. Most electrophysiological studies (another method of choice to investigate predictive language processing) present written sentences word by word in a (often slow) manner far removed from normal reading situations. Given that (in addition) the vast majorities of ERP studies measure the electrophysiological sign of anticipation (e.g. a reduced N400 ERP component) during the target word only (and not before), it cannot be ruled out that many studies have measured word integration difficulties rather than prediction (but see DeLong et al., 2005; Van Berkum, Brown, Zwitterlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004; for important exceptions to this).

2.6. Interim summary

We have presented five arguments that question the claim that prediction is a necessary part of language processing. In particular, we present evidence that not all language users predict, drawing mainly from studies with developing (i.e. children and adult second language learners) and illiterate language users. We also suggest that prediction effects may be highly dependent on the context in which they are obtained, showing – on the one hand – that prediction effects may disappear in contexts where participants are not provided required scaffolding in the form of slower speech or sufficient time to view possible alternatives and – on the other hand – that studies reporting robust prediction effects tend to provide participants with prediction-encouraging paradigms that question the extent to which prediction underlies natural language processing. We also discuss that – contrary to claims that predictive language use may aid processing of noisy input – prediction may actually be reduced given noisy input, or increased working memory demands. We interpret these arguments in the following manner: While we believe that language users do often predict upcoming input, we do not believe there is sufficient evidence for the claim that prediction is a necessary characteristic of language use. The population, context and resource-dependence of prediction effects in the literature strongly suggest that successful language processing can and does take place in the absence of prediction.

3. The way forward

We suggest that further resolution of this debate requires more focus on understanding why prediction effects are not found in some studies, in contrast to the large number of studies that find reliable prediction effects. In particular, we suggest it is important that future research more rigidly examines the factors contributing to differences across the two groups of studies. For instance, if prediction effects are not found in certain populations, to what extent do these populations also suffer from impoverished language skills or general cognitive skills that might explain the absence of robust prediction effects? Or if prediction effects are scaffolded by certain tasks, or certain kinds of stimuli or working memory demands, then to what extent is such scaffolding provided in natural conversation and how does language processing in natural conversation proceed without such scaffolding (and consequently without predictive processing). We believe therefore that it is critical that research on prediction in language processing

focuses more on “real world” situations. Important features of natural settings are casual speech and language which often has a low Cloze probability. Finally, on a different note, if research continues to suggest that prediction is necessary for language processing, e.g. with regard to language acquisition, or the learning of statistical regularities, it is critical that this work more accurately outlines the precise contribution of prediction to these processes and the extent to which they may be dependent or independent from prediction.

4. Conclusion

In sum, we believe that there are significant constraints for claims that prediction is necessary for language understanding. We conclude that claims that all language processing is predictive in nature are premature. Sometimes, processing words when they occur may be more efficient and economical than predicting them.

Disclosure statement

No potential conflict of interest was reported by the authors.

Note

- Note that L2 speakers did show differences in ERPs time-locked to the final noun, where the Cloze probability of expected nouns was higher than that of unexpected nouns (e.g. a raincoat in the example above). Here, L2 speakers, similar to L1 speakers, showed an N400 effect time-locked to the onset of the noun. This finding is not, however, evidence for prediction, and could merely index the ease of integration of a high Cloze probability word in a sentence context following presentation of the word. Indeed, the authors conclude, on the basis of the results reported above, that L2 readers do not predict upcoming words in a sentence context.

References

- Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed language sentences: Evidence from eye fixations and naming times. *Memory & Cognition*, 24, 477–492.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4), 583–609.
- Anderson, M. L., & Chemero, T. (2013). The problem with brain GUTs: Conflation of different senses of “prediction” threatens metaphysical disaster. *Behavioral and Brain Sciences*, 36(03), 204–205.
- Ashby, J., Rayner, K., & Clifton, C., Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology*, 58A, 1065–1086.
- Bastiaansen, M. C. M., Magyar, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, 22(7), 1333–1347.
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695–711.
- Bates, E., & Carnevale, G. (1993). New directions in research on language development. *Developmental Review*, 13, 436–470.
- Block, N., & Siegel, S. (2013). Attention and perceptual adaptation. *Behavioral and Brain Sciences*, 36(03), 205–206.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387.
- 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.
- Bowman, H., Filetti, M., Wyble, B., & Olivers, C. (2013). Attention is more than prediction precision. *Behavioral and Brain Sciences*, 36(03), 206–208.
- Bressler, S. L., & Richter, C. G. (2015). Interareal oscillatory synchronization in top-down neocortical processing. *Current Opinion in Neurobiology*, 31C, 62–66.
- Brouwer, S., Mitterer, H., & Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics*, 34, 519–539.
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7, 308–313.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234.
- Chang, F., Kidd, E., & Rowland, C. F. (2013). Prediction in processing is a by-product of language learning. *Behavioral and Brain Sciences*, 36(04), 350–351.
- Christiansen, M. H., & Chater, N. (in press). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*. doi:10.1017/S0140525X1500031X
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181–204.
- Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356–371.
- Davidson, D. J., & Indefrey, P. (2007). An inverse relation between event-related and time frequency violation responses in sentence processing. *Brain Research*, 1158, 81–92.
- Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120394.
- DeLong, K., Urbach, T., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121.
- De Ruyter, J. P., Mitterer, H., & Enfield, N. J. (2006). Predicting the end of a speaker’s turn; a cognitive cornerstone of conversation. *Language*, 82(3), 515–535.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.

- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195–225.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4), 547–582.
- Engel, A. K., & Fries, P. (2010). Beta-band oscillations – signalling the status quo? *Current Opinion in Neurobiology*, 20(2), 156–165.
- Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences*, 36(03), 211–212.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505.
- Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149–161.
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 862–877.
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- Friston, K. J., Bastos, A. M., Pinotsis, D., & Litvak, V. (2015). LFP and oscillations – what do they tell us? *Current Opinion in Neurobiology*, 31C, 1–6.
- Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, 22(7), 615–621.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20), 8051–8056.
- Giraud, A., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517.
- Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140, 186–209.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9), 416–423.
- Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, 4, 416.
- Hagoort, P., Hald, L. A., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669), 438–441.
- Hahn, U., & Oaksford, M. (2008). Inference from absence in language and thought. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 121–142). Oxford: Oxford University Press.
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4, 123–141.
- Hald, L. A., Bastiaansen, M. C. M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, 96(1), 90–105.
- Hand, C. J., Miell, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: It depends where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1294–1313.
- Huang, H-W., Meyer, A. M., & Federmeier, K. D. (2012). A “concrete view” of aging: Event related potentials reveal age-related changes in basic integrative processes in language. *Neuropsychologia*, 50, 26–35.
- Huetting, F. (2015). Four central questions about prediction in language processing. *Brain Research*. doi:10.1016/j.brainres.2015.02.014
- Huetting, F., & Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia – Evidence from eye-tracking. *Dyslexia*, 21(2), 97–122.
- Huetting, F., & Guerra, E. (2015). *Testing the limits of prediction in language processing: Prediction occurs but far from always*. Proceedings of the 21st Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP 2015), Malta.
- Huetting, F., & Janse, E. (in press). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2015.1047459
- Huetting, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language*, 54, 460–482.
- Huetting, F., Olivers, C. N., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, 137(2), 138–150.
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137, 151–171.
- Huetting, F., Singh, N., & Mishra, R. K. (2011). Language-mediated visual orienting behavior in low and high literates. *Frontiers in Psychology*, 2, 285.
- Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, 1146, 2–22.
- Johnson, K. (2004). Gold's theorem and cognitive science. *Philosophy of Science*, 71, 571–592.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Kennedy, A., Pynte, J., Murray, W. S., & Paul, S.-A. (2013). Frequency and predictability effects in the Dundee Corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology*, 66, 601–618.
- Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, 48(1), 171–184.
- Kielar, A., Meltzer, J., Moreno, S., Alain, C., & Bialystok, E. (2014). Oscillatory Responses to Semantic and Syntactic Violations. *Journal of Cognitive Neuroscience*, 26, 2840–2862.
- Kleinschmidt, D. F., & Jaeger, T. F. (in press). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*. doi:10.1037/a0038695
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262–284.

- Kretzschmar, F., Schlesewsky, M., & Staub, A. (in press). Dissociating Word Frequency and Predictability Effects in Reading: Evidence from Coregistration of Eye Movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <http://dx.doi.org/10.1037/xlm0000128>
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). New York, NY: Oxford University Press.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Lewis, A. G., & Bastiaansen, M. C. M. (in press). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*. doi:10.1016/j.cortex.2015.02.014.
- Lewis, J., & Elman, J. L. (2001). A connectionist investigation of linguistic arguments from poverty of the stimulus: Learning the unlearnable. In Proceedings of the 23rd annual conference of the Cognitive Science Society, 552–57.
- Luo, Y., Zhang, Y., Feng, X., & Zhou, X. (2010). Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience*, 169(2), 654–664.
- MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31, 883–914.
- Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language*, 66, 612–622.
- 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.
- Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, 126, 264–279.
- Mani, N., & Plunkett, K. (2010). In the infant's mind's ear evidence for implicit naming in 18-month-olds. *Psychological science*, 21(7), 908–913.
- Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, 121(2), 196–206.
- 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.
- McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6), 648–652.
- McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43(16), 1735–1751.
- McQueen, J. M., & Huettig, F. (2014). Interference of spoken word recognition through phonological priming from visual objects and printed words. *Attention, Perception, & Psychophysics*, 76(1), 190–200.
- Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, 14(4), 710–716.
- Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research*, 5(1), 3, 1–10.
- Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, 2(1), 138–153.
- Mitterer, H., & Russell, K. (2013). How phonological reductions sometimes help the listener. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 977–984.
- Nation, K., Marshall, C. M., & Altmann, G. (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.
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological review*, 113(2), 327–357.
- Peelle, J. E., Troiani, V., Wingfield, A., & Grossman, M. (2010). Neural processing during older adults' comprehension of spoken sentences: Age differences in resource allocation and connectivity. *Cerebral Cortex*, 20, 773–782.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113(2), 244–247.
- Penolazzi, B., Angrilli, A., & Job, R. (2009). Gamma EEG activity induced by semantic violation during sentence reading. *Neuroscience Letters*, 465(1), 74–78.
- Perruchet, P., & Desauty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, 36(7), 1299–1305.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105–110.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329–347.
- Popper, K. (2014). *The logic of scientific discovery*. London: Routledge.
- Rasmussen, D., & Eliasmith, C. (2013). God, the devil, and the details: Fleshing out the predictive processing framework. *Behavioral and Brain Sciences*, 36(03), 223–224.
- Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 720–732.
- Rayner, K., & Clifton Jr, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology*, 80(1), 4–9.
- Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, 21(3), 448–465.

- Rohde, D., & Plaut, D. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, 72, 67–109.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906–914.
- Rommers, J., Dijkstra, T., & Bastiaansen, M. C. M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25, 762–776.
- Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, 51(3), 437–447.
- Rowland, C., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition*, 125(1), 49–63.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Seidenberg, M., & Mac Donald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569–588.
- Sloman, A. (2013). What else can brains do? *Behavioral and Brain Sciences*, 36(03), 230–231.
- Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review*, 18(2), 371–376.
- Staub, A., & Benatar, A. (2013). Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin & Review*, 20, 1304–1311.
- St. Clair, M., Monaghan, P., & Ramscar, M. (2009). Relationships between Language Structure and Language Learning: The Suffixing Preference and Grammatical Categorization. *Cognitive Science*, 33, 1317–1329.
- Tremblay, P., Baroni, M., & Hasson, U. (2013). Processing of speech and non-speech sounds in the supratemporal plane: Auditory input preference does not predict sensitivity to statistical structure. *Neuroimage*, 66, 318–332.
- 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. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 443–467.
- Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J., Magyari, L., ... Bastiaansen, M. C. M. (2012). Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping*, 33(12), 2898–2912.
- Wang, X. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, 90, 1195–1268.
- Whitford, V., & Titone, D. (2014). The effects of reading comprehension and launch site on frequency-predictability interactions during paragraph reading. *Quarterly Journal of Experimental Psychology*, 67, 1151–1165.
- Wicha, N. Y. 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. *Journal of Cognitive Neuroscience*, 16, 1272–1288.
- Willems, R. M., Frank, S. L., Nijhoff, A. D., Hagoort, P., & Van den Bosch, A. (in press). Prediction during natural language comprehension. *Cerebral Cortex*. doi:10.1093/cercor/bhv075
- Willits, J., Seidenberg, M., & Saffran, J. (2009). *Verbs are looking good in language acquisition*. In Proceedings of the 31st Annual Conference of the Cognitive Science Society (pp. 2570–2575).
- Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, 49, 770–785.
- Wlotko, E. W., & Federmeier, K. D. (2013). Two sides of meaning: The scalp-recorded N400 reflects distinct contributions from the cerebral hemispheres. *Frontiers in Psychology*, 4, 181.