View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Edinburgh Research Explorer

THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Predicting while comprehending language

Citation for published version:
Pickering, M & Gambi, C 2018, 'Predicting while comprehending language: A theory and review' Psychological Bulletin, vol. 144, no. 10, pp. 1002-1044. DOI: 10.1037/bul0000158

Digital Object Identifier (DOI):
10.1037/bul0000158

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Psychological Bulletin

Publisher Rights Statement:
©American Psychological Association, 2018. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 05. Apr. 2019

Predicting while comprehending language: A theory and review

Psychological Bulletin – in press

Martin J. Pickering
University of Edinburgh
and
Chiara Gambi
University of Edinburgh
and
Cardiff University

Short Title: Predicting language

First Author's Address:
Martin Pickering
Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
United Kingdom
Phone: +44 (0) 131 650 3447
Fax: +44 (0) 131 650 3461
Email: martin.pickering@ed.ac.uk

Abstract

Researchers agree that comprehenders regularly predict upcoming language, but they do not always agree on what prediction is (and how to differentiate it from integration) or what constitutes evidence for it. After defining prediction, we show that it occurs at all linguistic levels from semantics to form, and then propose a theory of which mechanisms comprehenders use to predict. We argue that they most effectively predict using their production system (i.e., prediction-by-production): They covertly imitate the linguistic form of the speaker's utterance and construct a representation of the underlying communicative intention. Comprehenders can then run this intention through their own production system to prepare the predicted utterance. But doing so takes time and resources, and comprehenders vary in the extent of preparation, with many groups of comprehenders (non-native speakers, illiterates, children, and older adults) using it less than typical native young adults. We thus argue that prediction-by-production is an optional mechanism, which is augmented by mechanisms based on association. Support for our proposal comes from many areas of research (electrophysiological, eye-tracking, and behavioral studies of reading, spoken language processing in the context of visual environments, speech processing, and dialogue).

Keywords: dialogue; language comprehension; language production; prediction.

Public significance statement.
This theoretical review shows that people regularly predict upcoming language. Importantly, it also shows that in most cases people rely on their own ability to produce language to make predictions that are compatible with both the speaker's language and their intended message. This form of prediction aids, but it is not necessary for, language understanding.

What Does it Mean to Predict During Language Comprehension? (1)

Traditionally, most cognitive and perceptual psychology assumes that people deal with the world as they encounter it. More recently, however, researchers have proposed that the brain's fundamental computations are prediction and assessment of those predictions (A. Clark, 2013). Hence people may continuously use context to predict how the world might be and then compare these predictions with what they subsequently encounter. People are therefore as prepared as they can be for the stimuli that are likely to occur, and the benefit from getting predictions right most of the time may outweigh any difficulty from occasionally getting them wrong.

There is now extensive evidence that prediction is important for language comprehension, just as it is for perception and cognition more generally. In this paper, we propose an integrated theory of the mechanism of prediction during language comprehension (Section 2). We then use this theory as a guide to conduct a systematic review of the experimental evidence, across domains and methodologies, including electrophysiology, eye movements, speech, and dialogue (Section 3). Finally, we discuss broader implications of our proposal (Section 4). The central claim of the theory is that comprehenders predict with mechanisms that are used for producing language – and because the predictions that comprehenders make using these mechanisms are similar to those that they would make if they were producing themselves, the predictions tend to be accurate and successful. These mechanisms are not always used, but comprehenders can always fall back on general-purpose associative mechanisms.

In Section 1 we first discuss what it means to predict language. We start by asking to what extent language is predictable (1.1), and then go on to distinguish prediction from integration, both in the context of experimental work (1.2) and in the context of computational accounts of language processing (1.3). Finally, we set out methodological criteria that studies must meet in order to demonstrate prediction (1.4).

How predictable is language? (1.1).

For at least 40 years, it has been clear that people interpret language extremely rapidly. Comprehenders do not delay a word, phrase, or sentence before performing lexical access, parsing, and semantic analysis. In fact, they analyze each word as they encounter it and integrate it with prior context in a highly incremental fashion. In one of the first demonstrations of incrementality, Marslen-Wilson (1973) had participants shadow speech and found that their errors were determined by the prior context, even when they lagged little more than 250ms behind the speech that they heard. This finding suggests that they were not just repeating what they heard, but were immediately trying to combine it with previous context. In reading, Just and Carpenter (1980) found evidence for lexical, syntactic, and semantic processing as soon as a word was fixated. Similarly, Swinney (1979) showed that listeners used context to select the appropriate meaning of lexically ambiguous words (e.g., *bat*) within a few hundred milliseconds. When syntax is ambiguous, people also rapidly select or favor a syntactic analysis during both reading (e.g., Frazier & Rayner, 1982; Trueswell, Tanenhaus, & Garnsey, 1994) and listening (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Such incrementality means that comprehenders may sometimes adopt an analysis that turns out to be incorrect, but importantly it enables comprehenders to process linguistic input much more rapidly than would otherwise be possible.

But in fact people may comprehend even faster than is suggested by the evidence for incrementality. They may not only analyze each word as they encounter it, but also predict what they are going to encounter. Traditionally many researchers argued against prediction and instead assumed "bottom-up priority" (e.g., Forster, 1979; Marslen-Wilson, 1987). One major criticism of the prediction view was simply that most words are not predictable. Predictability was (and is) typically assessed by asking people to complete a sentence context, using the so-called *Cloze procedure* (Taylor, 1953)[1]; if most people produce the same completion, the context is deemed *predictive* or *constraining* and this completion highly predictable or *high-Cloze* (with alternative completions being *low-Cloze*). But most naturally occurring contexts are not highly predictive or constraining, with people providing many different completions (i.e., they are medium- to low-Cloze). Therefore prediction appeared to have a very limited value, and researchers assumed either that it never occurs or that it occurs only in unusual contexts (e.g., Stanovich & West, 1979). Hence, it would not be characteristic of language comprehension in general.

It is true that very few words are highly predictable (e.g., Luke & Christianson, 2016), but many words are moderately predictable. Moreover, language involves grammar, sounds, and meaning, and one or more of these may be predictable even if the word itself is not. For example, consider (1):

1. The boy went out to the park to fly a kite.

People may not be able to predict *fly*, but they can be fairly confident that the upcoming word will be a verb. They may not be able to predict that the boy went to the park, but could predict that he must have gone into a sufficiently large place, such as a shop or a beach. So even if people do not regularly predict words, they might predict some aspects of language. Moreover, the traditional argument against prediction depends on isolated utterances. But dialogue involves extensive repetition (Garrod & Anderson, 1987) and predictable sequences (e.g., question-answer pairs; Sacks, Schegloff, & Jefferson, 1974), and so may support prediction to a greater extent than isolated language. In conclusion, traditional psycholinguistics underestimated the predictability of language and may also have underestimated how often prediction actually occurs.

Prediction versus integration (1.2).

So far, we have discussed the predictability of language, but in order to define what it means for a comprehender to predict, we must contrast *prediction* with *integration*. Theoretically, prediction occurs if a comprehender activates linguistic information before processing input that carries that information. We use *pre-activation* to refer to the information that is activated predictively. In (1), *kite* is highly predictable after *a*. If people predict *kite*, they must pre-activate some component of its linguistic representation, such as the sound /k/ or conceptual feature +FLYABLE, before they could have done so on the basis of encountering *kite* (e.g., while reading *fly a*). Pre-activation of course goes beyond simply building up an appropriate contextual representation; for example building up a representation for the article *a* does not in itself imply pre-activation of a consonant sound.[2]

The benefits of successful prediction are made clear by the notion of pre-activation. When comprehenders predict successfully, they pre-activate representations that they use

[1] Throughout the paper, we report mean Cloze values when available, though note that studies vary on their precise instructions (e.g., use the most natural, the most plausible, or the first completion that comes to mind; Staub, Grant, Astheimer, & Cohen, 2015) and assessment criteria (e.g., e.g., whether singular and plural responses are collapsed).

[2] Note that we do not use the terms *expectation* or *anticipation*, which appear to be used in slightly different ways in the literature, and are not necessary for our account.

when they actually encounter the predicted input. Such pre-activation therefore allows them to perform some of the processing ahead of time, and therefore explains how prediction facilitates comprehension.

In contrast, integration occurs when a comprehender combines linguistic information that is activated as a result of processing the input, with a representation of the preceding input (i.e., the context). In (1), a comprehender would integrate *kite* by processing it and thus deriving linguistic representations (e.g., +FLYABLE), and then combining these representations with a representation of the prior context up to *fly a*. As integration does not involve pre-activation, it does not facilitate comprehension in the same way and all of the processing needs to occur bottom-up.

It can be very difficult to distinguish prediction from integration, and in particular to find evidence that is compatible with prediction but not integration. Much research demonstrates that people are faster at processing a more predictable than an unpredictable word. For example, Schwanenflugel and Shoben (1985) found that people made faster lexical decisions to highly predictable words (*the tired mother gave her dirty child a* <u>*bath*</u>) than less predictable words (*the tired mother gave her dirty child a* <u>*shower*</u>), even though the sentences were equally plausible and the words did not differ in frequency. In an ERP study, Kutas and Hillyard (1984) presented participants with target words that were more or less predictable given the context and showed that the amplitude of the N400 (a negative-going wave peaking around 400ms after word onset) was inversely related to predictability. Finally, using eye-tracking Ehrlich and Rayner (1981) found that readers fixated predictable words for less time than unpredictable words and were more likely not to fixate on a predictable word at all. All of these predictability effects have been extensively replicated, and are of course compatible with the evidence for incremental interpretation discussed in Section 1.1 (e.g., Marslen-Wilson, 1973).

Importantly, while we acknowledge that most of these findings could be due to prediction, they could also be due to integration. If people predict *kite* in (1), they pre-activate aspects of its linguistic representation, and therefore will find it easier to process the word than an unpredictable alternative such as *airplane* when they encounter it. If they do not predict *kite*, they may still find it easier to process *kite* than *airplane*, but in this case the facilitation would be because it is easier to integrate *kite* than *airplane* with the context, and not because of pre-activation. Under this integration interpretation, before encountering *kite*, comprehenders would of course activate properties of the context that are relevant to *kite* (e.g., that it describes a flying event involving a boy as agent) but crucially they would not pre-activate aspects of the linguistic representation of *kite*. Instead, they only activate those aspects when they encounter *kite* and it is only then that they are facilitated, because it is easier to combine the meaning of *kite* with the context (as opposed to the meaning of a different word, such as *airplane*).

As a rather different example, contextual effects on the perception of speech sounds can also be due to prediction or integration. English listeners tend to categorize a sound "halfway" between /s/ and /ʃ/ as /s/ if it follows *tremendou*- and as /ʃ/ if it follows *repleni*- (Samuel, 2001). It is possible that they accessed the lexical entry for *tremendous* or *replenish* by this point, and therefore pre-activated /s/ or /ʃ/ before they encountered the ambiguous sound. The pre-activated phoneme then affected their perception of the ambiguous sound. But it is also possible that listeners did not pre-activate the final phoneme of the lexical entry. When they encountered the ambiguous sound, they categorized it in a way that made it easier to integrate with the preceding context, but without predicting the missing sound (e.g., Norris, McQueen, & Cutler, 2000). Indeed, some related effects must be due to integration because they are caused by disambiguating information following, rather than preceding, the ambiguous sound (Ganong, 1980).

Finally, consider associative priming (D. E. Meyer & Schvaneveldt, 1971), for example the reduced N400 effects on *queen* after *king* (Bentin, McCarthy, & Wood, 1985; Rugg, 1985), and phonological priming, as in facilitation for *thing* after *king* (Praamstra & Stegeman, 1993; see also Slowiaczek, Nusbaum, & Pisoni, 1987). The traditional account of such effects involves activation spreading from the representation of the prime word to representations of associatively (or semantically) or phonologically related words. Recent researchers have (appropriately, in our view) characterized this explanation as predictive (e.g., Lau, Holcomb, & Kuperberg, 2013; McRae, Hare, Elman, & Ferretti, 2005). But an integration-based account of such findings, with people responding to *queen* quickly after *king* because *king-queen* is an appropriate (because associatively related) combination, is also possible. This account is reminiscent of compound-cue theories of priming, in which facilitation for *queen* after *king* occurs because the *king-queen* compound is retrieved easily from memory (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988).

Suprisal and entropy. (1.3)

As well as in the experimental evidence we have just considered in Section 1.2, we see another manifestation of the tension between prediction and integration in the notions of surprisal and entropy, which are incorporated in many computational models of language processing. As discussed by Hale (2001) and Levy (2008), the *surprisal* of a word is the negative logarithm of its conditional probability, that is the probability that the word will occur given the preceding context. To give an example, the word *kite* has a lower surprisal than *airplane* following *fly* in (1). Typically, this probability is derived from a large corpus, and how exactly the context is defined can vary. As such, surprisal represents a way of measuring predictability that is an alternative to the Cloze task – low surprisal corresponds to high Cloze, and high surprisal corresponds to low Cloze.

Accordingly, words with higher surprisal are harder to process than words with lower surprisal, for example leading to longer eye fixations and self-paced reading time (N. J. Smith & Levy, 2013) and increased N400 effects in ERP studies (Frank, Otten, Galli, & Vigliocco, 2015). Surprisal can also be defined over syntactic categories or structures (e.g., the probability of a noun or a sentential complement), and again higher surprisal leads to more processing difficulty (e.g., Linzen & Jaeger, 2016). But such findings do not demonstrate prediction, any more than do Ehrlich and Rayner (1981), Kutas and Hillyard (1984), or Schwanenflugel and Shoben (1985): They are compatible with prediction, but they are also fully compatible with an integration account in which low-surprisal words are easily integrated with context. Finally, although the relationship between surprisal and processing difficulty suggests that surprisal may constitute a good formalization of predictability (and may therefore be seen as part of a "computational-level" description in Marr's [1982] terms), there is also some evidence that measures of comprehension relate more closely to measures of predictability based on cloze values (N. J. Smith & Levy, 2011).

The *entropy* of a context is a measure of the degree of uncertainty about how it will continue (and is defined as $-\sum_i p_i \log p_i$, where p_i is the probability of continuation i). The entropy is higher for contexts that are compatible with more equiprobable continuations than fewer equiprobable continuations, and is higher for contexts with equiprobable continuations than contexts with continuations that differ in probability from one another (but have the same number of continuations). Roughly speaking, a high-Cloze context has low entropy, whereas a low-Cloze context has high entropy. Like surprisal, entropy can be defined with respect to words or, for example, syntactic categories or structures.

Unlike effects of surprisal, effects of entropy on processing of the context would provide strong evidence for prediction because, by definition, they occur before the predictable word. If people read low-entropy contexts faster than high-entropy contexts, it would suggest that they predicted possible continuations differently in the two types of context (e.g., they had difficulty predicting a large range of unlikely continuations). In an fMRI study of narrative comprehension, Willems, Frank, Nijhoff, Hagoort, and Van den Bosch (2016) found brain areas whose activation was negatively correlated with the entropy of the distribution of possible continuations (and Roark, Bachrach, Cardenas, and Pallier [2009] provide related behavioral evidence from self-paced reading times). But the evidence for such effects is very limited. Finally, some studies have investigated whether the degree to which processing of the target word reduces the entropy of the previous context (entropy reduction) may index processing difficulty. However, the effects of entropy reduction are very small (Frank et al., 2015; Linzen & Jaeger, 2016) and in any case they may reflect integration rather than prediction.

In sum, experimental research couched in terms of the information-theoretic notions of surprisal and entropy is compatible with prediction-based accounts. In fact, the same is true of studies concerned with the effects of Bayesian probability on the processing of words or sounds (e.g., Gibson, Bergen, & Piantadosi, 2013; Norris & McQueen, 2008). But none of these studies demonstrates that comprehenders use predictive mechanisms when understanding language, because the experimental effects occur on the predictable word itself, rather than before it occurs.

How to demonstrate prediction (1.4).

So far we have discussed the definition of prediction, how it differs from integration, how both prediction and integration can explain classic findings, and how they relate to the computational notions of surprisal and entropy. We now ask what evidence could demonstrate prediction and distinguish it from integration. We argue that demonstrating prediction requires a fundamentally different experimental approach than the one that has been used to demonstrate facilitatory effects of predictive contexts (e.g., Schwanenflugel & Shoben, 1985). In the latter kind of study, the focus is on processing of the (more or less) predictable word itself. It is very difficult to demonstrate prediction using this approach and most such experiments do not determine whether prediction occurs – that is, whether the word or any aspects of its meaning, grammar, or form are pre-activated.

To reiterate, prediction occurs if there is pre-activation of aspects of the linguistic representation of a predictable word (or other linguistic unit such as a speech sound). By far the clearest demonstration of prediction occurs when a study reveals activation of a linguistic representation of a word before the comprehender encounters that word. So, for example, if a comprehender listening to (1) looks at a picture of a kite over, say, one of an airplane, before hearing the word *kite* (which they do; Altmann & Kamide, 1999), then we can conclude that they have predicted the meaning of *kite*. Similarly, to test whether the phonology /kaIt/ is predicted, we could use the context in (2).

(2) The boy went out to the park to fly an …

This context is only a slight variation of the context in (1), and here as well most people would predict *kite* immediately following *fly*. However, the form of the determiner *an* is incompatible with this prediction (as *kite* begins with a consonant). Therefore, if comprehenders experience difficulty at *an*, then we can conclude that they have predicted the initial sound of *kite* (as first suggested by DeLong, Urbach, & Kutas, 2005). Finally, it is sometimes possible to measure pre-activation directly (as first attempted by Dikker & Pylkkänen, 2013). For example, since activation in the left middle temporal gyrus is linked to lexical retrieval, and is higher for less frequent words, we can use activity in this brain area as an index of pre-activation of an upcoming lexical item (Fruchter, Linzen, Westerlund, & Marantz, 2015).

We noted that prediction occurs if comprehenders activate some component of a word's linguistic representation before they could have done so on the basis of encountering it. Of course, "bottom-up" activation of such linguistic representations takes some time – for example, lexical access in visual word recognition takes about 130-150ms (e.g., Sereno & Rayner, 2003; Carreiras, Armstrong, Perea, & Frost, 2014). If an effect is detected before the relevant representation could have been activated bottom up, then pre-activation – and hence prediction – must have occurred. But such effects can only be used to demonstrate prediction if the time-course of such bottom-up activation is well established. In section 2, we review studies that meet our criterion for prediction: They either show effects before the target word is encountered, or they show effects at the target word but too early for "bottom-up" activation to account for the effects. We use these studies as the primary basis for the theoretical account of prediction that we develop in section 2.

A much larger group of studies are often interpreted as supporting or demonstrating prediction, but do not meet this criterion, as they do not test for effects before the predictable word. This is the case for studies such as Schwanenflugel and Shoben (1985), Ehrlich and Rayner (1981), and Kutas and Hillyard (1984). As argued above, they are compatible with both prediction and integration accounts. Our interpretation of these studies therefore differs from Kuperberg and Jaeger (2016), who interpreted these and other findings as follows: "The simple point we wish to make at this stage is that it is logically impossible to explain these effects without assuming that the context influences the state of the language processing system before the bottom-up input is observed." (p. 33). We disagree with this claim, as such findings can be explained without pre-activation of the predictable word.

Many (otherwise interesting) studies can therefore be interpreted as due to integration just as well as to prediction. When this is the case, we do not consider the studies further in this review. But in fact there are many studies that do not test for effects before the critical word, but for which a prediction explanation is preferable to an integration explanation on the basis of additional considerations. These studies require more detailed argumentation and are reviewed in section 3.

A theory of prediction (2)

In Section 1, we defined what we mean by prediction and how it differs from both incremental interpretation and integration. Importantly, we also set out precise methodological criteria that must be met in order for a study to demonstrate prediction: It must either measure before the target word or demonstrate effects after the target word that could not be due to bottom-up processing. The aim of Section 2 is to present a theory of prediction in language comprehension that builds on evidence that meets these methodological criteria. We review this evidence thematically to build up to the theory which is presented in Section 2.6.

In a nutshell, our theory of prediction is *general* -- because it applies to prediction made at all linguistic levels, from semantics to syntax and form, and *production-based*, because it proposes that the central mechanism used by comprehenders to generate predictions is the same mechanism they use to produce their own utterances. To predict-by-production, comprehenders first covertly imitate what they have comprehended so far. They then derive the intention underlying the utterance, taking into account the linguistic context provided by the utterance which has been covertly imitated, as well as aspects of background knowledge and other extra-linguistic information that the comprehender assumes are shared with the speaker. In addition, the comprehender may compensate for differences between herself and the speaker. Crucially, this derived intention is then run through the comprehender's own production system where it triggers the retrieval and building up of production representations, which constitute the comprehender's prediction of the speaker's upcoming utterance. Representations can be activated at any linguistic level, but activation at lower levels (such as form) follows activation at higher levels (such as semantics), just as in language production, and is dependent on sufficient time and resources being available to the comprehender. Thus, prediction-by-production is an *optional* mechanism that can support comprehension but is not necessary for comprehension to take place. Finally, we propose that prediction-by-production is augmented by an additional prediction mechanisms based on associations (prediction-by-association), which is less resource-intensive but also less accurate.

In the remainder of Section 2, we review evidence for each of the key proposals mentioned in the previous paragraph, namely that: comprehenders predict at all linguistic levels (2.1); comprehenders predict using their production system (2.2); comprehenders using prediction-by-production covertly imitate the speaker's utterance, compute the derived intention, and run this derived intention through their production system; (2.3) prediction-by-production is optional (2.4); comprehenders can also predict-by-association (2.5). Finally, we present the full theory of prediction in Section 2.6.

Comprehenders Predict at All Linguistic Levels (2.1)

In this section, we describe some studies that demonstrate that comprehenders predict aspects of meaning (semantics), grammar (syntax), and form. We chose these studies because for each of them it is clear what level of representation is predicted. In accord with most psycholinguistics (e.g., Allport & Funnell, 1981; Forster, 1979; Levelt, Roelofs, & Meyer, 1999), we adopt the basic distinction of levels of representation into semantics, syntax, and form, and ignore further distinctions (which are often disputed among researchers) except when necessary. We refer to semantic properties (concepts or features) with capitals (e.g., KITE, +FLYABLE), syntactic properties using italics (e.g., *N* for noun, *ACC* for accusative, or *kite* for the lemma[3] – the syntactic component of the lexical entry), and form using standard linguistic transcriptions (e.g., /kaIt/ for the phonology of *kite*).

Semantics. (2.1.1) Altmann and Kamide (1999) demonstrated prediction of semantics using the "visual world" paradigm. In this paradigm, participants see a small number of entities, either presented as isolated entities or arranged in a coherent scene, and hear an utterance (Tanenhaus et al., 1995; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Allopenna, Magnuson, & Tanenhaus, 1998). They may act on those entities, for example picking up one of them (e.g., Tanenhaus et al. 1995), although in Altmann and Kamide they simply observed them. In their study, participants saw a scene containing a boy, various toys (ball, toy train, toy car), and a cake. After a silent preview of the scene, they heard *the boy will eat the…cake*. Listeners' looks to the different entities were measured using an eye-tracker. Before they heard *cake*, they tended to look at the cake more than when they heard *the boy will move the…cake*. Clearly, the verb *eat* requires an object that refers to an edible entity (whereas *move* does not), and the cake is the only edible entity in the scene. The participant interpreted the scene as involving a cake, and therefore accessed its conceptual properties, critically including the fact that it is edible (unlike the other entities). She then heard *the boy will eat*, retrieved the semantics of EAT, and looked at the only object in the scene whose associated conceptual representation included the feature +EDIBLE.[4] The study

[3] We also use italics to refer to words (as opposed to their referents) in examples from experiments.

[4] The participant must have normally identified the location of the cake before hearing *eat*. In other words, during the silent preview, the participant encoded the objects and their locations.

does not demonstrate whether the comprehender predicted the concept CAKE (or indeed the lemma *cake* or the form /keɪk/), nor that, in the absence of the object, the participant would predict CAKE, but it does demonstrate prediction of semantic features.

In addition, Grisoni, McCormick Miller, and Pulvermüller (2017) had participants listen to sentences that were highly constraining (around 80% Cloze) for a hand-related (e.g., *I take a pen and I …write*) or face-related verb (e.g., *I find a cigarette on the desk and I … smoke*). Negated versions of the high-constraining contexts (around 20% Cloze; e.g., *I do not take the pen and I …*) provided a low-constraining control. Using EEG, they showed that after a high-constraining (but not a low-constraining) context participants pre-activated body-specific parts of motor cortex involved in the action implied by the predictable verb (i.e., hand for *write*, face for *smoke*). This activation occurred during the 100ms before the onset of the predictable verb and shows prediction of verb meaning (just as body-specific activation occurs after the verb is encountered; e.g., Pulvermüller, Shtyrov, & Ilmoniemi, 2006). Thus, Grisoni et al., like Altmann and Kamide (1999), demonstrated that comprehenders predict conceptual features, but in addition showed they can do so in the absence of a supportive visual context. Again, this study does not demonstrate whether comprehenders predict a specific concept, but we later discuss some evidence that this is also the case (Thornhill & Van Petten, 2012; see section 2.1.2).

Syntax. (2.1.2) Wicha, Moreno, and Kutas (2004) presented Spanish participants with highly constraining contexts (80% Cloze), such as *Mis papa´s quisieron cargar poco en su viaje. Pero con lo que llevaba mi madre de ropa no les cupo todo en…* (“My parents wanted to carry little on their trip. But with what my mother took in clothing, it did not all fit in…”). An article (*el$_{MASC}$*) that was incompatible with the gender of the following predictable word (*maleta$_{FEM}$*, suitcase) led to an enhanced late positivity, 500-700ms post-stimulus (but no enhanced N400). Therefore, comprehenders used the context to predict the semantics, which in turn led to the specific prediction of the target lemma and its associated syntactic gender (the words were semantically neither male nor female).

Similarly, Van Berkum, Brown, Zwitzerlood, Kooijman, and Hagoort (2005) had Dutch participants hear high-constraint (86% Cloze) contexts (e.g., *De inbreker had geen enkele moeite de geheime familiekluis te vinden. Deze bevond zich natuurlijk achter een…*, “The burglar had no trouble locating the secret family safe. This was of course located behind a…”), followed by a gender-marked adjective and the predictable noun (*schilderij*, painting$_{neuter}$) or an unpredictable but plausible noun of the opposite gender (*boekenkast*, bookcase$_{common}$). When participants encountered the unpredictable adjective (i.e., which agreed in gender with the unpredictable noun), ERPs showed a positivity 50-250ms after the adjective (e.g., *grote*, big$_{common}$) in comparison to the response for the predictable adjective (*groot*, big$_{neuter}$). A self-paced reading experiment confirmed that readers were disrupted when they encountered the unpredictable adjective.

In two related studies, Otten and colleagues also found evidence for prediction of grammatical gender at the adjective in Dutch, although with different ERP signatures. In a listening experiment, Otten, Nieuwland, and Van Berkum (2007) found the effect emerged between 300 and 600ms after the onset of the adjective, and as a negativity rather than a positivity. In a reading experiment, Otten and Van Berkum (2008; Experiment 2) found a later negativity between 900 and 1100ms after the onset of the adjective. In sum, while the ERP correlates of the gender prediction effect vary from study to study, comprehenders of languages that include a syntactic gender category appear to predict this syntactic feature.

This is consistent with the evidence from the “blank screen” paradigm (Altmann, 2004), in which the pictures are removed before the onset of the sentence and participants fixate on the location where the critical object had been.

Form. (2.1.3) Delong et al. (2005) recorded ERPs while participants read sentences such as *The day was breezy so the boy went outside to fly…* (presented one word at a time, with a 500ms interval between the appearance of successive words) and showed that the amplitude of the N400 on the noun was larger when the sentence ended with the unpredictable *an airplane* than the predictable *a kite*. More importantly, a related effect occurred at the preceding article: The amplitude of the N400 on the article was negatively correlated with the article's Cloze (which ranged from 0% to 96% in the study). This finding at the article implies that comprehenders predicted an aspect of the form of the predictable noun – whether it began with a consonant or a vowel – and were therefore surprised if the article was not compatible with this prediction. (An integration explanation is very unlikely as it would require people to find it easier to integrate *fly a* vs. *fly an*, and as *a* and *an* do not differ semantically, they should not give rise to an N400).[5]

Ito, Pickering, and Corley (2018) had native English participants listen to very highly constraining (Cloze: 97.5%) sentences (e.g., *The tourists expected rain when the sun went behind the…*) while looking at visual displays that contained a depiction of the highly predictable word (*cloud*), a form-related competitor (*clown*), or an unrelated competitor (*globe*). (A fourth condition is discussed in section 1.4.3). As expected, participants looked at the depiction of the predictable word more than the unrelated competitor from about 600ms before the predictable word onset. Crucially, they also looked at the depiction of the form-related competitor more than the unrelated competitor between 500 and 350ms before the predictable word onset. Hence, they pre-activated the form of the predictable word well in advance of encountering it.

In a study using magnetoencephalography (MEG), Dikker, Rabagliati, Farmer, and Pylkkänen (2010) presented participants with contexts that predicted a syntactic category (noun or verb participle), and then a target noun which had visual characteristics that were typical (e.g., *soda*) or atypical (e.g., *infant*) of the orthography of nouns. So participants read a context that predicted a noun followed by a typical noun (e.g., *The tasteful soda*), a matched context that predicted a verb followed by the same typical noun (*The tastefully soda*), or either type of context followed by an atypical noun (e.g., *The cute/cutely infant*). They found enhanced activity in visual cortex after 100-130ms (an M100) for a typical noun in a verb-biasing context versus a typical noun in a noun-biasing context, but no difference between contexts for atypical nouns. The effect was thus present only when there was a mismatch between the predictable syntactic category and the visual form of the target word.

Crucially, although the effect was not found before the target word, it occurred too rapidly to be the result of integration. Lexical access in visual word recognition takes 130-150ms (e.g., Sereno & Rayner, 2003; Carreiras et al., 2014), and so the process of activating syntactic category information (which is part of the lexical entry) and trying to integrate it with the syntactic representation of the context should have taken more than 130ms. Further, the effect was localized to a visual brain area, which strongly suggests that it was not elicited

[5] We note that a recent study (Ito, Martin, & Nieuwland, 2016) did not replicate this effect. Nieuwland et al. (2017) conducted a nine-lab replication of De Long et al. (2005) which also did not show the effect (though see Yan, Kuperberg, & Jaeger, 2017 for some criticism of their methodology). Martin et al. (2013) did however find a larger N400 to unpredictable (1% Cloze) than predictable (69% Cloze) articles using a design in which the sentence context was first presented as a whole (and remained on screen until participants pressed a button), and then the article and noun were presented one at a time, with a 700ms interval. In sum, the extent to which form prediction can be detected using the form of the article (which of course might in fact be due to the form of an intervening adjective; e.g., *an orange kite*) is currently unclear.

by syntactic processing but rather by visual form processing differences.[6] This finding means that comprehenders predicted the syntactic category of the upcoming word and that category's typical visual form. When they predicted a noun, they also predicted noun-like features, and this resulted in a reduced M100 for more typical nouns (see also Dikker, Rabagliati, & Pylkkänen, 2009; Herrmann, Maess, Hasting, & Friederici, 2009).[7]

Comprehenders Predict Using Their Production Systems (2.2)

We propose that the central mechanism for prediction is what we term *prediction-by-production*. People can comprehend incomplete utterances and complete them using their language production system as they do in the Cloze task, constructing some of the representations involved in overt speech (or writing) but stopping short of overt production. We base our proposal on Pickering and Garrod (2007, 2013), but the idea that prediction uses production processes is shared with Dell and Chang (2014), who argued that the same mechanisms are used to make predictions during language comprehension and production (in their *P-chain* framework; see also F. Chang, Dell, & Bock, 2006). We also note that other theorists consider a role for production in prediction (Federmeier, 2007; Huettig, 2015).

Evidence that the production system is activated during comprehension (e.g., Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Pulvermüller et al., 2006) is of course compatible with our proposal. Our proposal also accords with evidence that the role of production mechanisms is enhanced in speech comprehension under adverse conditions (see the meta-analysis by Adank, 2012), under the assumption that adverse conditions cause comprehenders to rely more on prediction-by-production. For example, comprehenders are better at understanding a novel accent in noise after training to imitate the accent (which presumably helps them develop new production representations that can assist with comprehension; Adank, Hagoort, & Bekkering, 2010). In addition, motor evoked potentials (MEPs) elicited by transcranial magnetic stimulation (TMS) of the lip area in motor cortex are larger when participants listen to distorted compared to natural speech (particularly when listening to

---

[6] Note that all critical comparisons in Dikker et al. (2010) were within target nouns, and there were effects of form typicality even in trial-level analyses, making it extremely unlikely that M100 differences were spurious.

[7] In a related self-paced reading study, Farmer, Christiansen, and Monaghan (2006) had participants read a context that predicted a verb (e.g., *The very old man attempted to*) and found that they took longer to read a noun-like verb than a verb-like verb. However, Staub, Grant, Clifton, and Rayner (2009) did not find any equivalent effect (though see Farmer, Monaghan, Misyak, & Christiansen, 2011 for a response). Farmer, Yan, Bicknell, and Tanenhaus (2015) found typicality effects on first fixation durations using eye-tracking. Importantly, they also showed that typicality effects were stronger when the syntactic category was more predictable. If Farmer et al.'s (2006, 2015) findings are robust, they are compatible with syntax-to-form prediction as shown by Dikker et al.'s MEG measures. But note that typicality effects on their own (i.e., without evidence for a modulation by category predictability) would also be compatible with an alternative explanation: The comprehender may simply find it easier to determine that a verb-like verb is a verb than that a noun-like verb is a verb, and consequently could begin the process of integration more quickly in the former case. Thus, showing an interaction between typicality and category predictability (as in Dikker et al., 2010 and Farmer et al., 2015) is essential to demonstrate prediction. In addition, Dikker et al.'s MEG data are stronger than the eye-tracking data because the typicality by predictability interaction in Farmer et al. (2015) was only robust for gaze duration measures (but not for first fixation durations), so it is unclear whether the effect occurred sufficiently early to rule out an integration-based explanation.

sounds that require movement of the lips in their articulation; Nuttall, Kennedy-Higgins, Hogan, Devlin, & Adank, 2016).

More importantly, recent evidence shows not only that the production system is involved in language comprehension, but also that production involvement during comprehension underlies prediction. Some of this evidence comes from studies showing parallels between prediction and production (Hintz & Meyer, 2015) or correlations between prediction and production skills (Rommers, Meyer, & Huettig, 2015; Hintz et al. 2017), and is thus supportive but not conclusive. Hintz and Meyer found that Dutch participants who listened to simple mathematical equations (e.g., *3+8=11*) looked at the solution (i.e., a number displayed on a clock face) predictively, and did so with similar timing to participants who had to complete those equations (*3+8=*), but these common patterns of behavior do not necessitate a common mechanism. Rommers et al. (2015; cf. Rommers et al., 2013) showed that listeners with higher verbal fluency (i.e., production ability) make more predictive looks to a predictable picture (e.g., a picture of a moon after the sentence *In 1969 zette Neil Armstrong als eerste mens voet op de…*, "In 1969 Neil Armstrong was the first man to set foot on the …"), and Hintz et al. similarly found that verbal fluency accounts for a large proportion of between-listener variance in looks to a predictable picture (e.g., apple after *De man schilt op dit moment een*…, "The man peels at this moment…"), at least when listeners are given a long time to preview the pictures in the display (in contrast, receptive vocabulary explained individual variance in predictive looks regardless of the amount of preview). These correlational findings, however, do not demonstrate a causal role for the production system during prediction in language comprehension.

But crucially, there is direct evidence for such causal relationship. First, Drake and Corley (2015) had participants listen to highly constraining contexts and then name a picture that corresponded either to the predictable word (match) or to a word that differed in onset from the predictable word (mismatch; e.g., *cap* for *tap*). Using ultrasound recordings, they compared articulation in the match and mismatch conditions to articulation in a control condition in which participants named the same pictures in isolation. They found that articulation diverged more from the control in the mismatch than in the match condition, suggesting that listeners predicted the final word using production mechanisms, and that such predictions affected articulation.

Second, predictive looks were disrupted by cerebellar rTMS (Lesage, Morgan, Olson, Meyer, & Miall, 2012) in a study closely based on Altmann and Kamide (1999): Comprehenders took longer to fixate on *cake* when they heard *The boy will eat the …* after they had received repetitive stimulation (thought to be inhibitory) to the right cerebellum. No delay occurred when they heard *The boy will move the …*, indicating that the disruption was specific to predictive language processing. Moreover, no prediction-specific effects occurred when comprehenders received no stimulation or stimulation to a control site. These findings make sense because the cerebellum contributes to "fast and flexible motor control by predicting the sensory consequences of movements on a fine timescale" (p. R795), and most likely does so for language production as well as for other types of movement (Ito, 2008; see also Moberget, Gullesen, Andersson, Ivry, & Endestad, 2014; Miall et al., 2016).

Most importantly, Martin, Branzi, and Bar (2018) showed that prediction of the noun gender at the article in Spanish (as in Wicha et al., 2004) was reduced under articulatory suppression: Comprehenders were asked to produce the syllable /ta/ in time with visual presentation of each word in a sentence, up until three words prior to the presentation of a gender-marked article that either matched or mismatched their prediction. Whereas N400 amplitude at the (more or less expected noun) was unaffected by this manipulation, the N400 elicited by unexpected articles was reduced under articulatory suppression, suggesting that engaging the production system selectively impaired prediction and not comprenshion as a whole. Moreover, no N400 reduction occurred in participants who either listened to a recording of themselves producing the syllable /ta/ or tapped their tongue (without producing a speech sound) in time with visual presentation of the words. These findings suggest that language production interferes with prediction (as indexed by the N400 reduction), rather than language comprehension or comparable non-linguistic action. In sum, several lines of evidence support our proposal that prediction during comprehension is based on production mechanisms.

Prediction-by-production (2.3)

Having argued that comprehenders predict by production in the previous section, we now describe the three key stages of the prediction-by-production mechanism. Motivation for these stages comes in part from evidence about prediction during language comprehension and in part from evidence about language production. Our aim is to integrate these two sources of evidence with each other and into our theory.

In order to predict by production, the comprehender must first determine (via non-predictive incremental comprehension processes) the linguistic representations corresponding to the speaker's utterance so far (the linguistic context). But these representations are part of the comprehension system; in order to constrain production processes, equivalent representations need to be activated within the production system. The first stage of prediction-by-production, which we term *covert imitation* (Pickering & Garrod, 2013), involves the activation of production representations that correspond to the representations built by the comprehension system. In practice, this stage might be often facilitated by representational parity - that is the fact that the production and comprehension system share representations. In fact, we have argued elsewhere (Gambi & Pickering, 2017) that there is good evidence for shared lexico-semantic and syntactic representations at least (evidence is less clear at the form level).

The second stage involves deriving the intention underlying the speaker's utterance. In addition to what the speaker has said so far, the comprehender also takes into account shared background knowledge and the shared visual (or other extra-linguistic) context. Such additional information, which collectively we label *non-linguistic context*, constrains the process of inverse mapping (from linguistic representations to intention, rather than from intention to linguistic representations, as normally is the case in production), and affects its output. This output is the *derived intention* that the comprehender assumes will underlie what the speaker will say next. In the third and final stage, the comprehender runs the derived intention through her own production system to construct linguistic representations underlying the predicted utterance.

Covert imitation. (2.3.1) If comprehenders did not take into account the linguistic context, they would often predict completions that were incompatible with the utterance produced by the speaker. But this is not the case. In fact, comprehenders' predictions are usually constrained by the linguistic context. Consider, for example, Kamide, Altmann, and Haywood (2003a, Experiment 3). In a visual-world experiment, they contrasted Japanese sentences such as *waitress-NOM customer-DAT merrily hamburger-ACC bring* ("The waitress will merrily bring the hamburger to the customer") and *waitress-NOM customer-ACC merrily tease* ("The waitress will merrily tease the customer"). Up until *merrily*, comprehenders may construct very similar representations for both sentences, and generate very similar predictions: for example, that the speaker intends to talk about an event involving a waitress performing some action in relation to a customer. However, the linguistic context specifies that *customer* is a recipient in the former, but not the latter sentence. Accordingly, participants predicted that a theme would be mentioned (in effect, because a recipient needs to be the recipient of something), and looked at the hamburger (a possible theme) more when *customer* was marked as dative (and thus typically a recipient) than when it was marked as accusative.

Covert imitation can account for the evidence that predictions are constrained by the linguistic context. Via covert imitation, the comprehender turns comprehension representations into production representations. Such representations affect subsequent processing within the production system, causing it to be congruent with the linguistic context provided by the speaker's utterance. For example, listeners in Kamide et al. (2003a) covertly imitated the speaker's representation in which *customer*-DAT is the recipient and then predicted that the speaker would produce a plausible theme, which was the hamburger in that scene. Note that, since the lexical content (i.e., *waitress*, *customer, merrily*) did not differ across conditions, participants must have covertly imitated the thematic and syntactic structure of the utterance, so that *hamburger* was predicted more strongly when the structure of the utterance made its mention more likely.

Similarly, in two Dutch studies demonstrating that comprehenders predict syntactic gender (see Section 2.1.1), Otten et al. (2007) and Otten and Van Berkum (2008, Experiment 2) showed that predictions of gender also depend on covert imitation of the structure of the sentence: Comprehenders showed evidence of having predicted the target lemma (*sword*) and its syntactic gender (neuter) after predictive contexts (e.g., *The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for a big*$_{\text{common}}$ *but rather old sword*$_{\text{neuter}}$) more than after control contexts that contained the same content words in different structural roles (e.g., *The benevolent sorcerer saw that the dragon threatened the brave knight. Quickly he reached for a big*$_{\text{common}}$ *but rather old sword*$_{\text{neuter}}$).

Finally, Hintz et al. (2017) examined the extent to which the strength of the association between the verb (*peel*) and the predictable word (*apple*) could account for predictive looks in a visual-world study. General associations between the verb and the predictable word did not explain variance in predictive looks, but functional associations did. Functional associations take into account the structural relationship between the verb and the predictable word (i.e., they measure how likely *apple* is to be the object of *peel*; see McRae et al., 2005). Thus, these findings once again suggest that predictive looks are constrained by covert imitation of structure and are not merely driven by the lexical content of the utterance.

Deriving the intention. (2.3.2) In the previous section, we have shown that covert imitation of the linguistic context constrains the predictions made via the production system, using predictive looks towards a potential theme (hamburger) in Kamide et al.'s (2003a) Japanese experiment as one example. But in addition to the linguistic context, the comprehender also takes into account the non-linguistic context, such as shared background knowledge (e.g., that customers are likely to order hamburgers) and the shared visual context (e.g., that a hamburger is visually present). Together, shared linguistic context, visual context, and background knowledge provide the three components of the "common ground" that underlies much successful communication (H.H. Clark, 1996). By incorporating these three components in the process of computing the derived intention, the comprehender maximizes her chances of correctly predicting what the speaker will say.

But of course there is an obvious difference between speaking and predicting what the speaker will say: The speaker knows his own intention but the comprehender cannot be certain of the speaker's intention. In many cases, the information provided by the linguistic and non-linguistic context will be sufficient for the comprehender to recover the speaker's intention, because the assumption that this information is shared between the comprehender and the speaker is correct. This is likely to be the case when speaker and comprehender can recall what has been said, have access to the same visual information, and their background knowledge is sufficiently similar in the relevant domain (e.g., they share a restaurant script; Schank & Abelson, 1977).

However, the comprehender sometimes needs to compensate for differences between herself and the speaker, including differences in access to background knowledge, visual context, and indeed memory for the content of previous utterances. Some studies suggest that they fail to compensate (Keysar, Barr, & Horton, 1998; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003), at least during early processing, and are therefore egocentric. But other studies suggest they can be more successful. For example, Hanna, Tanenhaus, and Trueswell (2003) either had a speaker instruct a listener about where to place a red triangle on a grid (so that the shape unambiguously became part of common ground), or asked the listener to place a red triangle on a secret location on their grid, unbeknownst to the speaker (so that the shape was unambiguously excluded from common ground). The speaker then said *Now put the blue triangle on the red one*, referring to another red triangle (the target). Listeners looked at the target and the identical competitor which was in common ground equally often, but were far more likely to look at the target than at the competitor which was not in common ground, and did so from the very onset of *red*. This study suggests comprehenders can rapidly take into account what they believe the speaker is aware of and therefore that their derived intention incorporates adjustments for self-other differences (i.e., it is not merely what would be the comprehender's own intention under the circumstances); see Brown-Schmidt (2009) for related evidence. However, the studies by Keysar and colleagues also suggest that the egocentric perspective is not always overridden.

Importantly, self-other adjustments to the derived intention affect predictions of what the speaker is likely to say next. Thus, Barr (2008) found that common ground determines which objects listeners predict the speaker will refer to (even though when listeners later hear the object name they look at objects with similar names which are not in common ground). To further illustrate the point, we consider a visual-world eye-tracking study by Chambers and San Juan (2008). In one condition, participants moved an object around a grid, in response to instructions such as *Move the chair to Area 2.* When participants then heard *Now return the …*, they looked at the chair more often than when they heard *Now move the ….* Based on the meaning of *return*, they predicted another reference to the previously mentioned object. But when the speaker asked participants to move two objects before saying *Now return the …*, participants did not preferentially look at the chair, as it was not unambiguously the referent. Crucially, however, this pattern of effects occurred when the participant knew the speaker was also aware of this ambiguity, but not when the participant moved one of the objects unbeknownst to the speaker. Thus comprehenders predicted that the chair would be mentioned on the basis of a derived intention that incorporated their beliefs about how the speaker's intention differed from their own.

Running the intention through the production system. (2.3.3) Having derived the intention, the comprehender is now in a position to predict the speaker's upcoming utterance. To do so, she runs the derived intention through her production system, reproducing some of the processes involved in speaking but stopping before overt articulation. This process constitutes prediction-by-production and, of course, it shares many characteristics with language production; therefore below we describe each of the stages of language production (Levelt, 1989; see also Goldrick, Ferreira, & Miozzo, 2014), and once we have identified the key characteristics of each stage, we discuss evidence that they apply to prediction during language comprehension as well.

In the first stage of speaking (so-called conceptualization), people construct a semantic representation that includes entities, events, and their relations (e.g., indicating who did what to whom). In non-sentential contexts (e.g., when naming a pictured object), this stage takes around 150-200ms (Indefrey & Levelt, 2004; Indefrey, 2011), and leads to the activation of conceptual features, for example +FLYABLE for *kite*; it may also include activation of a unitary KITE concept (Levelt et al., 1999), though this is controversial (cf.

Dell, 1986; Bierwisch & Schreuder, 1992). Importantly, although related concepts such as AIRPLANE, BALLOON, or STRING, or features such as +LIGHT or +COLORFUL, also receive some (but less) activation (i.e., a space of related concepts are activated in parallel), activation is quickly directed only to those concepts that are relevant in the context of the utterance that is being produced. So after *The boy went out to the park to fly*, KITE, AIRPLANE, and perhaps BALLOON remain activated, but the activation of STRING quickly decays.

We propose that such parallel, but directed, activation characterizes prediction-by-production of semantic representations (concepts or features): for example, prediction of +EDIBLE in Altmann and Kamide (1999), or prediction of body-specific features in Grisoni et al. (2017). Using Altmann and Kamide (1999) as an example, it may be that the concept CAKE retains or increases its activation (compared to other edible entities) because of the presence of a depicted cake in the scene, which constitutes shared visual information and thus constrains the process of deriving the underlying intention. This illustrates how a comprehender using prediction-by-production first accesses a potentially large network of semantic representations (e.g., all concepts that share the feature +EDIBLE), but then directs the activation so that relevant concepts, and specifically those that are compatible with the results of covert imitation and the derived intention, retain (or gain additional) activation, whereas irrelevant concepts rapidly lose activation.

In production, conceptualization is followed by the processes of syntactic encoding (Bock & Levelt, 1994) and lexical selection (Levelt et al., 1999). Syntactic encoding refers to the process of mapping the event structure activated as part of the semantic representation (e.g., TRANSFER event) to an appropriate syntactic frame. So for example, a speaker intending to describe a scene where an agent transfers a theme object to a recipient will typically select either a prepositional object frame (PO, as in *The assassin will send a parcel to the dictator*) or a double object frame (DO, as in *The assassin will send the dictator a parcel*). Which frame is selected will depend on a range of factors including which frame(s) the speaker has selected most recently (i.e., on structural priming; Bock, 1986b; Pickering & Ferreira, 2008) and on the outcome of lexical selection for the verb (e.g., V. Ferreira, 1996) or the nouns (e.g., Bock, 1986a).

Comprehenders can similarly predict syntactic structures and categories on the basis of the event structure, and they also take previous experience and lexical restrictions into account while doing so. For example, Arai, Van Gompel, and Scheepers (2007) had participants read a PO or a DO sentence (with very similar meanings) and then listen to another PO or DO sentence, in a context containing pictures of the theme and the recipient. When they heard the verb, they tended to look at the entity corresponding to its theme if they had heard a PO but at the entity corresponding to its recipient if they had heard a DO (when the verb was repeated; see also Arai, Nakamura, & Mazuka, 2015). Another study showed that the effect could not be due to repetition of the order of animate versus inanimate entities (Carminati, van Gompel, Scheepers, & Arai, 2008), and there is consensus that structural priming is mainly driven by repetition of syntactic structure (Pickering & Ferreira, 2008). In addition, structural priming sometimes affects predictions even without verb repetition (Thothathiri & Snedeker, 2008). In sum, as in production, comprehenders appear to predict syntax and rapidly select a single structure. This process of selecting an appropriate structure of course depends on compatibility with the results of covert imitation, as already discussed in relation to Kamide et al. (2003a).

Alongside syntactic encoding, speakers engage in lexical selection. Lexical selection is the process of accessing the lemma associated with the currently most activated concept (*kite*), together with the lemmas associated with related concepts (*airplane, balloon*). According to Levelt et al. (1999), a speaker naming a single object compares these lemmas and generally selects the most activated one in around 70-90ms (Indefrey, 2011). This selection process likely requires some processing resources, although it can occur without full attention (Roelofs & Piai, 2011). Selection of a lemma leads to activation of lexicalized syntactic information, for example the grammatical gender of the selected item (e.g., Vigliocco, Antonini, & Garrett, 1997).

Similarly, during prediction-by-production, the comprehender can also access the grammatical properties of a predicted item. Importantly, the comprehender must have predicted semantics before predicting syntax, just as in production. So for example, recall that in Van Berkum et al. (2005; see also Wicha et al., 2004; Otten et al., 2007, Otten & Van Berkum, 2008, Exp. 2), participants comprehended sentence contexts that strongly predicted a noun (*painting*) and were disrupted when they encountered an adjective whose gender (non-neuter) was incompatible with this noun. This finding supports prediction-by-production, with the context leading to semantic activation of the concept PAINTING that then in turn leads to activation of the associated lemma (*painting*) and its grammatical gender.

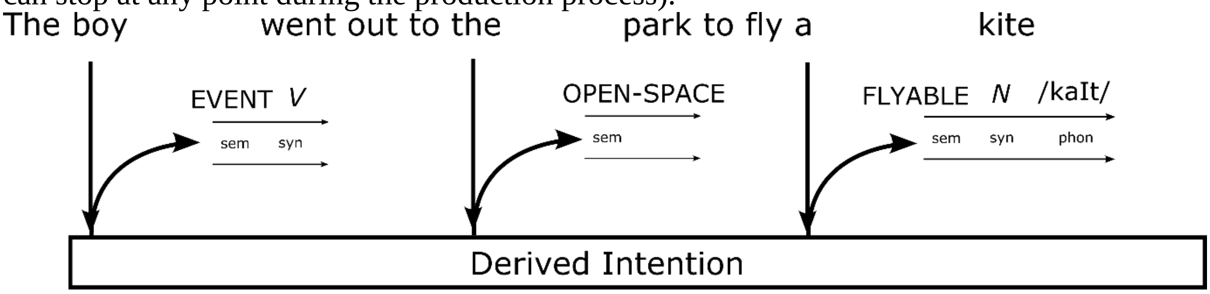
At this point, we reach a point of controversy within theories of word production. Levelt et al. (1999) assumed that speakers select a single lemma and access its word-form, which captures its phonological properties. A speaker who prepares to name a kite therefore activates /kaIt/ but not /ˈeəpleɪn/ or /bəˈluːn/. In contrast, Dell (1986) argued that speakers can activate the phonology of more than one candidate item. In fact, much evidence indicates that such parallel activation does sometimes occur, for example for near-synonyms (e.g., Peterson & Savoy, 1998). Next, the speaker constructs a phonetic representation that feeds into the process of articulation, and all alternatives except one are eventually abandoned. It takes 180-200ms to access phonological and syllabic information, and a further 110-200ms to phonetically encode and begin articulation (Indefrey, 2011), though these timings may be affected by context (Strijkers & Costa, 2016).

Similarly, comprehenders using prediction-by-production may activate the phonology of predictable words after performing lexical access. For example, participants in Ito et al. (2018) predicted the phonological (or possibly orthographic) form of *cloud* after predicting the semantic and syntactic representations associated with *cloud* on the basis of a highly constraining context. This process parallels language production, with phonological activation occurring later and being dependent on lexical activation (and we note that looks to the picture of the phonological competitor occurred after looks to the predictable word).

In summary, we have reviewed experimental findings that highlight how the key stages and components of language production, from semantics to form, are reflected in prediction during language comprehension. Comprehenders predict meaning, in which case they activate many aspects of meaning in parallel, and rapidly focus on the elements that are relevant for production. They predict syntax, in which case they also select an appropriate lemma and its syntactic properties. And they can also predict phonology (or other aspects of sound), in which case activation narrows down to a single word form and its phonological properties. As in theories of production, predictions of syntax and phonology involve fewer alternatives than predictions of semantics, take longer to construct, and are dependent on predictions of semantics.

Our examples have considered the predictions people make at a specific point in an utterance. But they can repeatedly predict during comprehension, going through cycles of predicting-by-production. As she encounters more of the utterance, the comprehender incrementally updates a representation of the speaker's intention underlying that utterance. Importantly, this representation does not only constitute the basis of the comprehender's ongoing understanding of the speaker's utterance, but it is also provides the input to cycles of prediction-by-production that repeatedly generate candidate continuations for the speaker's utterance. An example (for the utterance *The boy went out to the park to fly a kite*) is illustrated in Figure 1.

Figure 1. An example sentence with predictions that may be computed at three different time points (after *The boy*, after *The boy went out to the*, and after *The boy went out to fly a*). Downward arrows represent the process of comprehension and the derivation of the underlying intention. Upward arrows represent the activation of the production system and abbreviations stand for the three main stages of production (*sem* = semantics, *syn* = syntax, *phon* = phonology). The example illustrates the fact that comprehenders comprehend continuously (as indicated by the single long box representing the continuously updated representation of the derived intention underlying this utterance), and can predict at any time during comprehension, but do not always go through all the stages of production (and instead can stop at any point during the production process).



The Optionality of Prediction (2.4)

We have shown that people can predict by initiating the process of language production while comprehending an utterance produced by someone else, as though they were using it to complete the utterance, and that they can go through cycles of predicting-by-production, repeatedly updating their predictions as they encounter more of the utterance. Now we propose that, although comprehenders can predict-by-production, they are unlikely to do so at every word in the utterance, or to go through all the stages of production at every cycle. In other words, prediction-by-production is optional.

To illustrate this with an example, we return to predictions that might occur while comprehending the utterance *The boy went out to the park to fly a kite* in Figure 1. Horizontal arrows represent the production system, and abbreviations within them indicate the stages of production the comprehender goes through as part of different prediction-by-production cycles. For instance, after *The boy* she may predict that the speaker will describe an event (but be unsure which one) and that the speaker will use a verb. This leads to the activation of the meaning EVENT and the syntactic category V. After *The boy went out to the*, the comprehender's production system leads to the activation of the semantic category of OPEN-SPACE. After *The boy went out to the park to fly a*, it leads to the activation of the specific word *kite*, which includes all of its lexical information (e.g., +FLYABLE, noun, /kaIt/), including the fact that the first phoneme of *kite* is a consonant. In sum, this comprehender can go through all the stages of production (from semantics to form), as is the case for the prediction of *kite*, but she does not do so at every word.

There are two good reasons why comprehenders may seldom go through all the stages of production as shown in this example. First, each stage takes time, and so comprehenders may not have the later stages ready in time for the predictions to be useful. Second, prediction-by-production requires resources (just as production does), and sufficient resources may not always be available. We discuss these two points in more detail below (sections 2.4.1 and 2.4.2).

These points lead us to propose that prediction-by-production is not a necessary component of language comprehension: At some points in a sentence, comprehenders may not predict at all, and at other points they may predict early stages (e.g., semantics) but not later stages (e.g., form). Note that the proposal that prediction is optional distinguishes our account of prediction during language comprehension from predictive coding accounts of perception (Friston, 2005; A. Clark, 2013), as they essentially equate the process of perceiving with prediction; this proposal is instead shared with Huettig and Mani (2016). We conclude this section by discussing evidence for the optionality of prediction-by-production (2.4.3).

The timing of prediction-by-production. (2.4.1) Prediction-by-production can only be as fast as the comprehender's production system. The comprehender must run the intention through her production system, but before she can do that she must also determine the speaker's intention, which involves compensating for differences between herself and the speaker (and will be harder and more time-consuming if the discrepancy is greater). Importantly, prediction-by-production may occur for the earlier stages of production when it does not occur for the later stages (Indefrey & Levelt, 2004), because the earlier stages are more likely to be completed "in time" (i.e., before the speaker begins articulating the predicted word).

For the same reason, although prediction-by-production can occur in all acts of comprehension, it is more likely to reach later stages when the speaker is slower than faster. It is particularly likely when comprehending speakers that are slow or disfluent, for example when they have difficulty with what is being uttered. It is also more likely when the presentation rate is slow, as in many psycholinguistic experiments (see sections 3.1 and 3.2 for data). As most natural comprehension involves fast speech rates (e.g., Quené, 2008), comprehenders may often have insufficient time to predict form by production, but they are more likely to have enough time to predict semantics (and perhaps syntax). In addition, slower producers (such as less proficient non-native speakers) are less likely to use prediction-by-production (see section 2.4.3).

It is important to stress that comprehenders can predict well in advance of the predictable item. The comprehender incrementally updates a representation of the speaker's intention, and uses that representation to generate predictions in a continuous manner (see Figure 1). This means that the comprehender may sometimes be able to begin prediction-by-production earlier than the previous word.

Indeed, we know that speakers prepare more than one word at a time. Meyer (1996) had participants produce conjoined phrases (*cup and table*) and found that a semantic distractor affected production whether it was related to the first or the second word, but a phonological distractor affected production only if it was related to the first word. Her finding suggests that speakers plan semantics further ahead than phonology (see also Smith & Wheeldon, 2004). In a similar way, comprehenders could predict semantics further ahead than phonology. If so, they may sometimes pre-activate the semantics of the predicted word well ahead of when the word is predicted to occur. They may also be able to predict the phonology "in time" (i.e., before the word) because advanced prediction of semantics gives them a head-start when it comes to predicting phonology as well. We know of no study that has manipulated the scope of semantic prediction directly, and so we do not know whether predictions of phonological forms would be more likely when comprehenders can predict semantics further ahead.

Use of resources in prediction-by-production. (2.4.2) Several dual-tasking studies of picture naming suggest that all stages of production up to and including phonology are resource intensive and require attention (see Roelofs & Piai, 2011). Hence, prediction-by-production should also take up resources as well as time. What is more, such resources are likely shared with aspects of comprehension (cf. Kempen, 2014), as suggested by much neuroscientific evidence (Menenti, Gerhan, Segaert, & Hagoort, 2011; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012; Silbert, Honey, Simony, Poeppel, & Hasson, 2014). Consistent with this, dual-tasking studies of dialogue indicate that a secondary task performed concurrently with comprehension is disrupted the most when the comprehender is about to start speaking (Boiteau, Malone, Peters, & Almor, 2014; Sjerps & Meyer, 2015), suggesting that preparation for upcoming production is resource-intensive. We might therefore expect that prediction-by-production may sometimes even interfere with the process of comprehension.

Indeed, higher cognitive load on comprehenders appears to make prediction less likely. For example, Huettig and Janse (2016) found that comprehenders with better working memory and faster processing speed make more predictive eye-movements in the visual world paradigm. In addition, Ito, Corley, and Pickering (2017) found that such eye-movements are delayed under memory load. However, it may be that these effects of cognitive load are in part dependent on the experimental method. Using the same ERP paradigm as Otten and Van Berkum (2008), Otten and Van Berkum (2009) found that both an early and a late negativity were elicited by Dutch adjectives incompatible with the gender of an expected noun. Somewhat surprisingly, low-working memory participants showed a more marked late negativity, potentially suggesting they predicted more, but in contrast the early negativity was not affected by the comprehenders' working memory; thus, it is unclear to what extent individual differences in working memory capacity affect prediction of syntax in this paradigm. In sum, more research is needed to explore the impact of cognitive load on prediction-by-production, but there is already some evidence that the latter is resource-intensive.

Prediction-by-production does not always occur. (2.4.3) Evidence for the optionality of prediction-by-production comes from groups of comprehenders who show limited or no prediction, while still being able to comprehend. Mitsugi and MacWhinney (2016) found that non-native (L2) Japanese speakers did not use case-marking (dative/accusative) to predict in a visual world study based on Kamide et al. (2003a; Experiment 3). Thus, L2 speakers may not predict in conditions when native speakers do predict. However, Foucart, Martin, Moreno, and Costa (2014) found that both late French-Spanish and early Catalan-Spanish bilinguals reading Spanish predicted the gender of a highly predictable (81% Cloze) noun (see Foucart, Ruiz-Tada, & Costa, 2016 for similar results in speech).

In an ERP study based on DeLong et al. (2005), Martin et al. (2013) found that late Spanish-English bilinguals (who learned English after age 8) did not predict the phonological form of a highly predictable noun (unlike English monolinguals), despite being familiar with the *a*/*an* rule for English.[8] Moreover, Ito et al. (2018) found that Japanese L2 speakers of English did not look at a phonological competitor of a highly predictable noun while listening to English sentences (unlike native English speakers; see section 2.1.3). (They did look at the competitor *after* hearing the highly predictable noun, indicating they had knowledge of the phonological relationship.) In summary, non-natives sometimes, but not always, appear to predict syntactic information, but there is no evidence that they predict phonology. These findings may reflect difficulty of predicting later stages of the production process.

[8] In their replication of Martin et al. (2013), Ito et al. (2016) also found no evidence for prediction of phonology in non-native English speakers, but note that they found no evidence for native speakers either (see Footnote 5).

In addition, poor reading skills may impair or prevent prediction during spoken language comprehension. In a visual-world study, adults with high literacy fixated predictable target objects before they heard the object's name but adults with low literacy did not (Mishra, Singh, Pandey, & Huettig, 2012). Dutch adults with dyslexia predicted a target object after hearing a gender-marked article (which was followed by a gender-unmarked adjective), but did so more slowly than adults without dyslexia (Huettig & Brouwer, 2015). Moreover, listeners from both groups who had higher word-reading skills were more likely to predict the target object. Although the cause of this relationship is unclear, these results show that even native speakers differ in their prediction skills. Moreover, children's prediction skills are also related to their vocabulary (e.g., Mani & Huettig, 2012; Borovsky, Elman, & Fernald, 2012) or reading ability (Mani & Huettig, 2014). Finally, DeLong et al. (2012) replicated DeLong et al. (2005) with older adults but found no correlation between the ERP effect on the article (*a*/*an*) and the article's Cloze probability (though the older adults showed effects on the noun), suggesting that they did not predict the noun form even though their comprehension of the sentence was otherwise intact (as indicated by the N400 effect on the noun)[9].

In sum, several studies involving non-native speakers, adults with limited reading skills, children, and older adults suggest that these comprehenders do not always predict to the same extent as typical native young adults. While it is difficult to ascertain whether comprehension proceeds as rapidly or smoothly in these comprehenders as it does in typical native young adults, these findings show language comprehension can occur without prediction. At present, we do not know what makes less typical comprehenders less likely to engage in prediction, but it is possible that limited resources (e.g., in children or older adults) might contribute to these effects. In any case, together with the evidence that prediction-by-production takes time and resources, these findings make a strong case for the optionality of prediction-by-production.

Prediction-by-Association (2.5)

So far, we have discussed how comprehenders predict-by-production and our proposal is that prediction-by-production constitutes the most effective mechanism for prediction during language comprehension. We have described how the process of predicting by production is constrained by covert simulation of the speaker's utterance and relies on computing the underlying intention (with adjustments) and eventually running that intention through the comprehender's production system. We have also noted that comprehenders use prediction-by-production optionally, depending on whether time and resources allow. However, comprehenders also have another prediction mechanism at their disposal, one that is not based on production.

It is very likely that prediction is involved in the spreading of activation between related representations (Collins & Loftus, 1975), such as in semantic/associative or phonological priming. We have already mentioned in Section 1.2 that although priming of *king* by *queen* or *thing* may be explained in terms of integration, the most common explanation of such priming involves prediction. According to such an explanation, encountering *king* leads to activation of the *king* representation, and hence activation very quickly (e.g., Perea & Gotor, 1997) spreads to representations that are linked to it (in long-term memory), such as *queen* and *thing*. This spreading of activation is a form of prediction.

[9] This finding should however be interpreted with caution given the difficulty with replicating the effect in younger adults (see footnote 5).

But the process by which activation spreads between different representations during priming is not compatible with prediction-by-production: It need not be constrained by covert imitation of the speaker's utterance so far. For example, a speaker who utters *king* is unlikely to utter *thing* in close proximity, and a similar argument may be made for many items that are semantically related (e.g., *shorts* and *tuxedo*, both items of clothing but unlikely to be mentioned together). To illustrate this point further, consider a visual-world study by Kukona, Fang, Aicher, Chen, and Magnuson (2011). They had participants hear utterances such as *Bill will arrest* …, while looking at pictures of a robber and a policeman (and unrelated characters). Participants looked at both the policeman and the robber upon hearing *arrest*. It therefore appears that the word *arrest* activates, via the concept ARREST, both the concept POLICEMAN and the concept ROBBER, thus increasing the likelihood of fixations to the corresponding entities. But note that the word *policeman* is incompatible with the linguistic context so far, as an agent for *arrest* has already been specified, and *policeman* is an unlikely continuation. This finding thus contrasts with Kamide et al. (2003a, Experiment 3), where comprehenders' predictions were instead constrained by the linguistic context so far. It also serves to illustrate an interesting distinction between predicting a state of the world and predicting what the speaker will say (see Van Berkum, 2013). Listeners in Kukona et al. accurately predicted that the arresting event would involve a policeman and a robber, but they did not accurately predict order of mention.

There is other evidence that comprehenders generate predictions that are not constrained by the linguistic context. Kamide et al. (2003a, Experiment 2) found that participants who heard *The man will ride* … (while looking at a display containing a motorbike, a carousel, a beer, and a candied apple) tended to fixate a motorbike more than participants who heard *The girl will ride* … or participants who heard *The man will taste* ….. But in addition, participants who heard *The girl will ride* … tended to fixate the motorbike more than participants who heard *The girl will taste*. The latter finding suggests that the spreading of activation from RIDE is not limited to the set of concepts activated by the subject noun, in a way that contrasts with covert imitation of the context. Borovsky et al. (2012) found similar results with both adults and 3-10 year old children. In addition, Kukona, Cho, Magnuson, and Tabor (2014) found that participants looked at a white car after hearing *The boy will eat the white*, despite the fact that CAR is not compatible with the semantic representation of EAT. Finally, Sauppe (2016) found predictive looks to the agent in the verb-initial language Tagalog even when verb morphology indicated that the agent would not immediately follow the verb.

To further illustrate the difference between prediction-by-association and prediction-by-production, it is useful to consider how each mechanism would explain the findings from Altmann and Kamide (1999). An explanation of these findings in terms of prediction-by-production would be as follows: Given sufficient time to process the scene (see Hintz et al., 2017), the comprehender incorporates representations of the objects in the scene together with the covertly imitated utterance *The boy will eat* into the derived intention (because she assumes the speaker is describing the scene, which constitutes shared visual information); as a result, the derived intention involves the cake, but not one of the other objects in the scene, and also not something edible that is not in the scene. In contrast, prediction-by-association would explain the same findings as follows: Activation spreads very quickly from the representation of *eat* to representations for edible objects, such as the concept CAKE (as well as other concepts such as APPLE or STEAK). The participant then looks to a location of an object with a matching conceptual representation, and the cake just happens to be the only matching object in the scene.

In sum, there is some evidence that predictive looks in the visual-world paradigm are in part incompatible with covert imitation and thus with prediction-by-production. Such evidence, together with evidence from semantic/associative and phonological priming, shows that comprehenders sometimes appear to activate a large network of semantically, associatively, or phonologically related items very rapidly, in what appears to be a largely unconstrained or undirected manner. In practice, this means that the representations that are predicted via spreading activation often do not correspond to the upcoming words. Moreover, as this type of activation also decays very rapidly (e.g., McNamara, 2005), even when the speaker does eventually produce the related word, it may well occur too late for the prediction to be useful.

Given that the spreading of activation between linked representations in long-term memory is (presumably) resource-free (Neely & Kaan, 2001), this type of prediction could nevertheless still benefit comprehension. For example, activation triggered by multiple lexical items could accumulate over time. One example of this may be the findings by Kamide et al. (2003a), following *The man will ride*. Activation of MAN would first result in the spreading of activation to associated concepts, such as MOTORBIKE (as well as BEER, as both concepts had corresponding pictures in the scene). When RIDE becomes activated, it also spreads activation to MOTORBIKE (as well as CAROUSEL), and this results in MOTORBIKE being activated more than either BEER or CAROUSEL, because of the summation of associations with the meanings of the subject and the verb. It is however unlikely that such summed associations could support prediction over longer utterances and across unrelated intervening words, and of course we have already reviewed examples of prediction that cannot be explained this way (see section 2.3.1 on covert imitation).

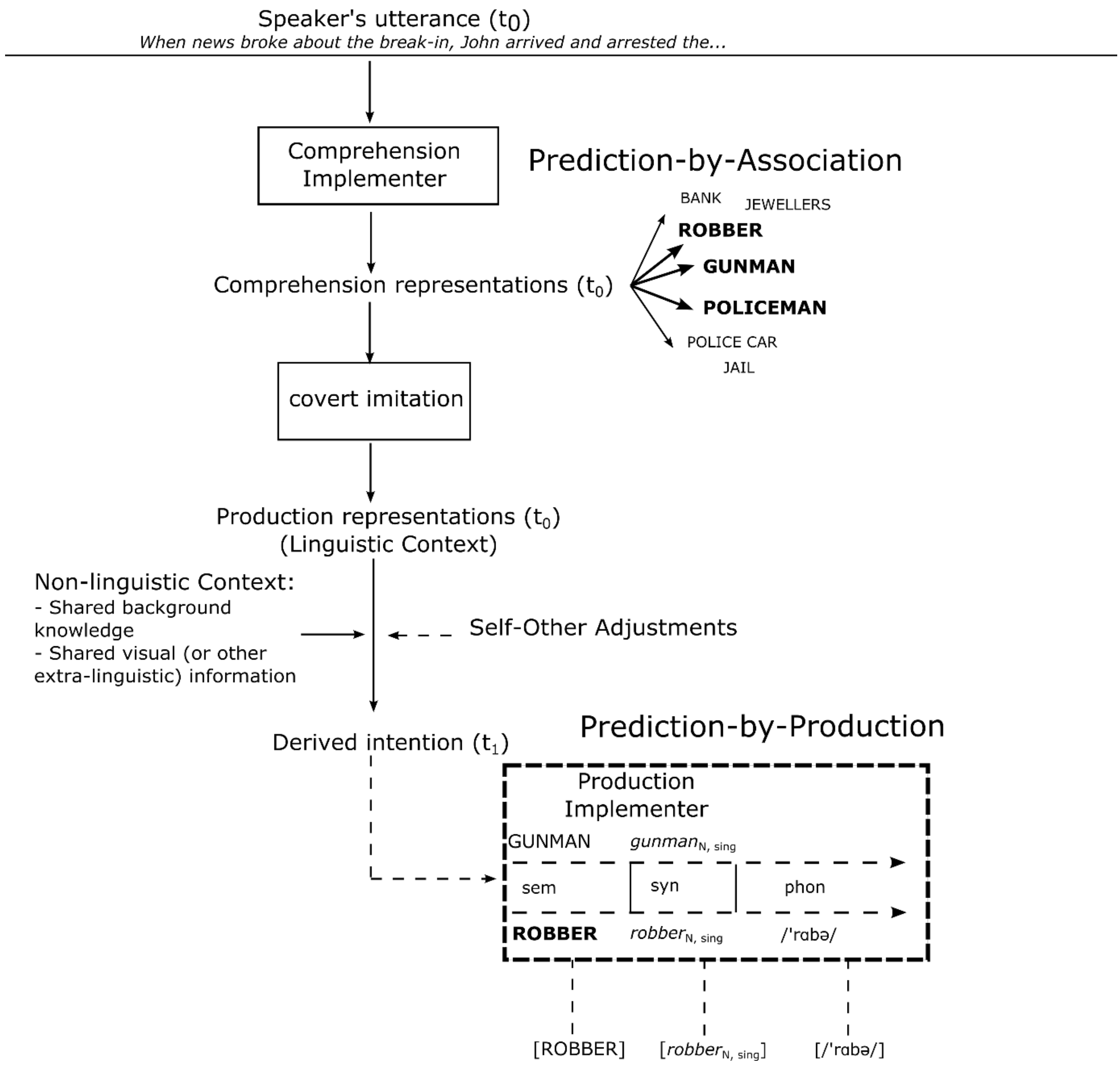
Prediction-by-association may be largely automatic and as such constitutes a non-optional prediction mechanism, one which is an inherent component of the process of language comprehension. Note that the form of spreading activation that leads to parallel activation of multiple semantically related concepts may well correspond to the initial stage of the semantic component of prediction-by-production. If so, it would be compatible with the widespread assumption that the semantic network is shared between production and comprehension (Gambi & Pickering, 2017).

Patterns of spreading activation are based on the structure of the comprehender's linguistic knowledge, which is in turn is based on experience comprehending language. Following Pickering and Garrod (2013), we therefore regard spreading activation as a form of prediction-by-association. The structure of the mental lexicon will of course constrain which parts of the network activation spreads to, but unlike in prediction-by-production the flow of activation does not need to be directional (from semantics, to syntax, and then to phonology). Crucially, the limitations of prediction via spreading activation (such as inaccuracy and decay rate) mean that it should play a comparatively small role in prediction (though there may be other forms of prediction-by-association which play a role in prediction; see Section 4.3).

A Model of Prediction (2.6)

We can now present our model of prediction (in relation to Figure 2). This model assumes that comprehenders have two mechanisms for prediction. The most important and effective, but optional, mechanism is prediction-by-production. In addition, comprehenders possess a less effective, but non-optional mechanism: prediction-by-association. Prediction-by-production depends on convert imitation and the process of constructing the derived intention. Although prediction can take place throughout an utterance (as illustrated in Figure 1), Figure 2 focuses on prediction at one point in the interest of readability.

Figure 2. An illustration of prediction-by-production and prediction-by-association. Boxes refer to processes; unboxed descriptions refer to representations. Solid lines indicate processes that are an integral part of comprehension; dashed lines are optional processes. At the top, the comprehender builds comprehension representations corresponding to the speaker's utterance at time t_0 using the comprehension implementer. Such representations are the basis for prediction-by-association, which leads to the pre-activation of several concepts (more strongly activated concepts are in bold). Comprehension representations also feed into the process of deriving the intention the comprehender would use to continue the utterance (derived intention at a later time t_1) if she were speaking. To do so, the comprehender makes use of covert imitation of the linguistic context, takes non-linguistic context into account, and may apply contextual adjustments for differences between herself and the speaker. Then, the comprehender uses her production implementer to activate production representations corresponding to the predicted word *robber*, first in semantics (ROBBER), then in syntax (*robber*$_{\text{N, sing}}$), and finally in phonology (/'rɑbə/). Square brackets around the set of predicted semantic, syntactic, and phonological representations indicate that prediction of syntax depends on prediction of semantics, and prediction of phonology in turn depends on prediction of syntax; note that predictions of later stages of production need not always occur, as indicated by the dashed arrows within the production implementer. The content of the box labelled *production implementer* depicts stages of the process of production, including activation of alternative concepts (GUNMAN) and lemmas (*gunman*$_{\text{N, sing}}$) that are ultimately abandoned.



As the speaker's utterance unfolds, the comprehender incrementally constructs comprehension representations of phonology, syntax, and semantics, using the comprehension implementer (comprehension representations at t_0 in the figure). Activation then spreads from these representations to associated representations at any of these levels *via* prediction-by-association. The comprehender then turns the comprehension representations of the utterance so far into the representations that she would have constructed if she had produced the utterance, using covert imitation. She then derives the intention that she would use to continue the speaker's utterance (derived intention at t_1, see Figure 2) via a process of inverse mapping. In doing so, she takes into account not only the linguistic context (so far), but also the non-linguistic context, which includes shared background knowledge and shared visual (or other extra-linguistic) information. In addition, she may apply self-other adjustments to compensate for differences between the comprehender and the speaker in relation to memory for the linguistic context and access to the non-linguistic context (see section 3.3 for further discussion).

Once the comprehender has derived the intention that she would use to produce the next part of the utterance (e.g., sound, word, or phrase; derived intention at t_1), she runs this derived intention through her production system to begin implementing the processes involved in speaking. She can construct semantics, syntax, and phonology in order, or may stop after any of these levels. She can also compute all stages of speaking, in which case she will complete the utterance (note this case is not depicted in Figure 2); this is of course what the comprehender does in the Cloze task when she produces a continuation and what she may do in dialogue when she completes her partner's utterance (Lerner, 2002; see Section 3.5).

To illustrate, consider predictions that may take place after comprehending *When news broke about the break-in, John arrived and arrested the...* Using prediction-by-association, the comprehender initially predicts a large network of associated concepts, including locations such as BANK, JEWELLERS, and JAIL, instruments such as POLICE CAR, and event participants such as POLICEMAN, ROBBER, and GUNMAN. Some of these predictions (more precisely, the words corresponding to these concepts) are compatible with possible continuations of the sentence. But others are not; for example, POLICEMAN is a plausible agent of *arrested*, but its agent has already been specified (i.e., by the word *John*). In addition, while locations and instruments (e.g., *jail* or *police car*) may be mentioned at some point, they are not required shortly after *arrested*, whereas a patient such as *robber* or *gunman* is required. Initially, the strength of the activation of any specific prediction depends on the strength and number of its associations with the context (i.e., on the structure of the semantic network), and particularly with the immediate context (because of the fast rate of decay of prediction-by-association). In Figure 2 we assume that at *arrested* the concepts POLICEMAN, ROBBER, and GUNMAN are more activated than any of the locations or instruments, and that POLICEMAN is activated to approximately the same extent as ROBBER or GUNMAN. Note that, although here we focus on semantics, these predictions may occur at any linguistic level. These predictions-by-association are non-optional – they are an inherent component of comprehending such a sentence.

Crucially, the comprehender then can use prediction-by-production. First, she covertly imitates what the speaker has said so far (the linguistic context), and combines this with an assessment of the non-linguistic context to construct the derived intention. In doing so, she may also take into account any differences between herself and the speaker (if she avoids egocentricity). For example, she may realize that the speaker, like herself, can see a man carrying a gun (i.e., the shared visual context) or know that the speaker is particularly interested in firearms (i.e., a difference between herself and the speaker), and therefore incorporate this information into the derived intention. She then uses this derived intention to begin the processes involved in producing a continuation.

The comprehender rapidly focuses the semantic activation on the concepts associated with potential completions, with their degree of activation depending on the likelihood of them serving as part of the completion (see *production implementer* in Figure 2). So at *arrested*, the concept ROBBER receives high activation, GUNMAN lower activation, but POLICEMAN now receives no activation, just like locations and instruments. (If the non-linguistic context supports GUNMAN, then it will of course receive higher activation than ROBBER.) Subject to time and resources, the comprehender implements more or less of the stages that would be involved in producing a completion. So she may select the lemma *robber* (while dropping *gunman*), and construct its syntax (singular count noun) in around 250ms and phonology */'rabə/* (or other aspects of its sound structure) in around 450ms. This example illustrates how prediction-by-association and prediction-by-production may be used in comprehending a single utterance. We now apply this theory of prediction to a range of studies in language comprehension.

Prediction in Language Comprehension: A Theoretical Review (3)

We now conduct a systematic review of studies of prediction and interpret them in terms of our theory. While Section 2 considered only evidence that unambiguously meets our methodological criteria for demonstrating prediction, Section 3 also considers studies that provide less clear evidence (i.e., that could be interpreted in terms of integration as well as in terms of prediction). We limit our discussion to studies for which a convincing case can be made that they should be interpreted in terms of prediction, and for those studies we carefully examine the arguments in favor of and against prediction. In addition, for the sake of readability, we did not include every study that met our methodological criteria in Section 2, and we review these additional studies here to show they also support our proposal.

We start by considering situations in which a comprehender processes language on its own, and does not produce language overtly. These are (1) electrophysiological (and related) studies of word processing in sentences (and texts), and (2) behavioral studies of reading, primarily involving eye-tracking. We then consider (3) spoken language processing in the context of non-linguistic visual environments. After this, we turn to studies of (4) speech processing and (5) dialogue. At the end of this section, we consider prediction in different populations (6), such as non-native speakers, children, and older adults. Many of these studies use stimuli that become more or less predictable at a critical point and measure predictability using the *Cloze procedure* (see section 1.1).

Electrophysiological (and Other Neuroscientific) Studies (3.1)

Much evidence for prediction comes from electrophysiological studies, in particular event-related potentials (ERPs). The ERP literature has paid particularly close attention to the prediction of words. For example, Van Petten and Luka (2012) acknowledged that comprehenders may predict semantics, but restricted their review to the prediction for "a specific word (lexical item) to occur in the future" (p. 179). In contrast, our interest is in prediction at different levels of representation (section 2.1). In fact, we have already discussed several ERP studies that demonstrate prediction at different levels (semantics: Grisoni et al., 2017; syntax: Wicha et al., 2004, Van Berkum et al., 2005, form: De Long et al., 2005, Ito et al., 2017, Kim & Lai, 2012; see section 2.1).

Each of those studies shows effects of a predictable representation before that representation could have been activated bottom-up. We now review additional studies that also show such effects but where we cannot be clear which specific representation was predicted (section 3.1.1). Then, we review findings that *need not* be due to prediction, because they measure on the target word and the timing of the effects is such that they could have occurred as a result of bottom-up processing, but where additional considerations (specific to each study) suggest that they *are in fact* due to prediction (section 3.1.2).

We exclude studies for which an interpretation based on integration is at least as likely as one based on prediction. These include many studies that looked at the N400 response to a target word and which draw on the fact that this response is greater for a less predictable than a more predictable word (Kutas & Hillyard, 1984). This relationship is extremely strong (Wlotko and Federmeier [2012b] reported an inverse correlation of .9 at the grand average level.) But as we pointed out in section 1.2, the reduced N400 for predictable words may reflect ease of integration (Neville, Kutas, Chesney, & Schmidt, 1986; Kutas, Hillyard, & Gazzaniga, 1988). For example, studies of listening show that N400 effects due to unpredictable words are time-locked to when they diverge phonetically from the predictable word (Connolly & Phillips, 1994; Van Petten, Coulson, Rubin, Plante, & Parks, 1999; Van Den Brink, Brown, & Hagoort, 2001), suggesting such N400 effects may reflect prediction of word forms (i.e., specified for sounds). But it is equally possible that participants use the sounds to activate the words ("bottom up"), and these are then integrated more or less easily with the context.

Another example of a study that may appear to demonstrate prediction but is in fact compatible with integration is Maess, Mamashli, Obleser, Helle, and Friederici (2016). Using MEG, they found that, while the magnitude of the N400 at the noun was larger for less predictable (Cloze <25%) than more predictable (Cloze >50%) nouns, the magnitude of the N400 at the verb was larger for verbs in more constraining contexts (e.g., *He conducts the orchestra*) than less constraining ones (*He leads the orchestra*). Moreover, there were strong negative correlations between the magnitude of the neural activation at the verb and at the noun (across a range of left temporal areas and the parahippocampus), which could suggest that in more constraining contexts participants pre-activated the upcoming noun and that this then facilitated later processing of the noun. However, constraining verbs may elicit stronger activation because they are semantically richer, rather than because they facilitate pre-activation of the nouns; this would make the finding compatible with integration.

Studies such as Connolly and Phillips (1994) and Maess et al. (2016) are thus excluded from our review. But many other EEG and neuroscientific studies are included and show how much of the literature can be interpreted in terms of our model. We conclude the section by arguing that neuroscientific evidence supports prediction-by-production (section 3.1.3). This conclusion is compatible with Federmeier (2007), who also proposed (largely on the basis of evidence from hemispheric differences) that ERP evidence supports the use of production mechanisms during prediction (see also DeLong, Groppe, Urbach, & Kutas, 2012; Federmeier, Kutas, & Schul, 2010; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007).

Electrophysiological studies demonstrating prediction but not the level of prediction. (3.1.1) Some studies unambiguously demonstrate prediction but do not reveal which level of representation is predicted. In an MEG study, Fruchter et al. (2015) had participants read adjective-noun pairs, in which the adjective was either highly or weakly predictive of the noun (e.g., *economic* is highly predictive of *growth* but weakly predictive of *reform*), and in which the (more or less) predictable noun was more or less frequent. They found increased activity in the left middle temporal gyrus (left MTG) just before presentation of a lower (vs. higher) frequency noun, but only when the adjective was highly predictive of the noun. When the adjective was weakly predictive of the noun, the frequency effect occurred only after presentation of the noun. Thus, participants may have used the adjective to predict a specific noun. The MTG has been associated with lexical access in comprehension (e.g., Lau, Phillips, & Poeppel, 2008) and, importantly, production (e.g., Indefrey & Levelt, 2004). However, the locus of frequency effects in word production (e.g., Jescheniak & Levelt, 1994; Caramazza, Costa, Miozzo, & Bi, 2001) or comprehension (Dahan, Magnuson, & Tanenhaus, 2001) is unclear and so we do not know whether Fruchter et al.'s study showed prediction of syntax or form.

Dikker and Pylkkänen (2011) showed participants a picture followed by a noun phrase that matched (or mismatched) the specific item in the picture (e.g., an apple) or the semantic field (e.g., a collection of food). They found an M100 effect in visual cortex when the noun phrase matched the specific item but not the semantic field. As the effect is so rapid, and it occurs in visual cortex, it must involve prediction (see Dikker et al., 2010). This conclusion was further supported by a reanalysis of Dikker and Pylkkänen (2011) conducted by Dikker and Pylkkänen (2013). They analyzed MEG responses before the target noun phrase. Because trials followed a rigid structure, participants presumably became aware when the target would appear. Just before it occurred, there was activation of left mid-temporal cortex followed by activation of visual cortex (as well as activation of ventro-medial prefrontal cortex) in the specific item versus the semantic field context. These findings suggest that people predicted the noun (its semantics) followed by its (visual) form, at the moment that was appropriate to facilitate processing of the noun. However, we can only infer this process indirectly from the localization of the activation, and so we cannot be sure what aspect of the linguistic representation of the noun was predicted.

Similarly, Boylan, Trueswell, and Thompson-Schill (2014) had participants read contexts that predicted either noun or verbs. Using fMRI, they examined activation of visual areas before the target word was presented, while participants were instructed to look for an appropriate word among a random pattern of dots. Activation in the so-called visual-form area differentiated between contexts predicting a noun and those predicting a verb, suggesting that participants predicted visual characteristics of the typical orthography of different syntactic categories. However, once again we can only infer what is being predicted using localization; it is possible that participants predicted syntax, but it is also possible that they predicted some other characteristic that differs systematically between their nouns and verbs (e.g., aspects of semantics that may lead to differences in mental imagery).

Electrophysiological studies supporting but not demonstrating prediction. (3.1.2) In Federmeier and Kutas (1999), participants read high-cloze contexts such as *They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of …* They then read the predictable word (*palms*), a semantically (i.e., categorically) related and unpredictable word (*pines*), or a semantically unrelated and unpredictable word (*tulips*). The presentation rate, as is typical for most ERP studies of reading, was fairly slow (SOA of 500ms). The N400 was reduced for the highly predictable *palms* compared to the unpredictable *tulips*. More importantly, it was also reduced for *pines* versus *tulips* (see also Federmeier, McLennan, Ochoa, & Kutas, 2002), despite the fact that both words were unpredictable (and also less plausible than the predictable *palms*). In addition, the N400 reduction for related unpredictable versus unrelated unpredictable words was greater in a very constraining context (90% Cloze) than a less constraining context (59% Cloze), suggesting that the effect depended on the predictability of *palms*.

Since these effects occurred after the predictable or a related word was encountered and processed for meaning, it is possible that the reduced N400 at *pines* was due to ease-of-integration. On this account, when participants encountered *pines,* they would have activated the corresponding concept PINES, which in turn would have activated PALMS (or some of its features), and the activation of PALMS would then have led to the N400 reduction, as PALMS integrates better with the context. However, the related unpredictable word *pines* was not a strong lexical associate of *palms*, so it is unlikely than PINES activated PALMS directly. It is instead more likely that participants predicted a feature (or features) common to both pines and palms (e.g., HAS-TRUNK, +TROPICAL), or the common category TREE.

We can explain Federmeier and Kutas's (1999) results in terms of prediction-by-production. Comprehenders first predicted HAS-TRUNK or TREE, and activation spread to related concepts. A speaker producing the sentence would also go through these stages, and then strongly activate the concept PALMS and more weakly activate other concepts, including PINES (but also other trees and tropical plants). A comprehender might similarly activate the concept PALMS most strongly but also activate PINES, leading to the N400 reduction for related words observed by Federmeier and Kutas (1999).

In a study based on Federmeier and Kutas, but using auditory rather than written stimuli, Romero-Rivas, Martin, and Costa (2016) manipulated whether the stimuli were spoken in a foreign accent or not. Listeners who listened to accented speech showed no reduction in the N400 for the related compared to the unrelated word, in contrast to participants who listened to non-accented speech. It is not clear whether the foreign accent caused participants to predict less or instead to make different predictions (e.g., just of the predictable word), but in any case these findings are compatible with the optionality of prediction-by-production (section 2.4). Moreover, using a 500ms SOA, Wlotko and Federmeier (2015) replicated Federmeier and Kutas's (1999) finding of N400 reduction to unpredictable related words compared to unpredictable and unrelated words. But when they used a 250ms SOA, they found a smaller reduction for unpredictable related words. Hence, this finding also supports the optionality of prediction-by-production: Specifically, it shows that even semantic prediction takes time.[10]

Metusalem et al. (2012) had participants read a description of an event in which kids went outside after a blizzard that ended with *They spent the whole day outside building a big* …. They found a reduced N400 both for a predictable final word *snowman* (81% Cloze) but also for an event-relevant but unpredictable (0% Cloze) and implausible word (*jacket*). Since *jacket* was not strongly associated with the words in the context or with the predictable word *snowman* (at least not more than the control word *towel*), an integration explanation is unlikely. Instead, the context likely activated the event structure representation "playing outside in the cold"; from this, activation spread to related concepts, including JACKET.

Now, let us briefly contrast these results with Federmeier and Kutas (1999). The two studies are superficially similar (they both show N400-reduction for unpredictable and implausible words). But in Methusalem et al. (2012), JACKET was not activated because of its relationship to the predicted concept SNOWMAN. Instead, comprehenders activated the snowman-building event non-predictively and this in turn led to activation of JACKET via prediction-by-association (i.e., from the event structure or schema). Importantly, a speaker would not pre-activate the word *jacket* and so a comprehender would not predict *jacket* using production mechanisms. Note that Amsel, DeLong, and Kutas (2015) replicated Metusalem et al.'s (2012) finding (while controlling for plausibility). They also showed a reduced N400 for words that had a perceptuo-motor relationship with both the predicted word and the context. Comprehenders therefore activate perceptuo-motor aspects of semantics (and not merely abstract features; see also Grisoni et al., 2017). Amsel et al.'s effects may be due to the relationship between the target word and the predicted word (i.e., involving prediction-by-production) or the target word and the context (i.e., involving prediction-by-association).

In discussing Federmeier and Kutas (1999), we noted that comprehenders might predict semantic features (e.g., HAS-TRUNK, +TROPICAL) or the conceptual category TREE. Thornhill and Van Petten (2012) replaced Federmeier and Kutas's *pines* condition with a near-synonym of the predicted word. Specifically, they had more predictive contexts (78% Cloze; e.g., *On his vacation, he got some much needed*) or less predictive contexts (30% Cloze) followed by the most predictable word (*rest*), a near-synonym (*relaxation*), or an unrelated word (*sun*). As expected, the near-synonyms led to a smaller N400 than the unrelated word (with the reduction being greater in the more predictive contexts). More importantly, in predictive contexts, there was an enhanced late (post-N400) frontal positivity for near-synonyms and unrelated words compared to predictable words. This finding suggests that comprehenders predict conceptual representations of highly predictable words, and do not merely predict semantic features (which greatly overlap between words and their near-synonyms). Therefore these results are compatible with production models in which lemmas are linked to unitary concepts (e.g., Levelt et al., 1999; *contra* Dell, 1986; Bierwisch & Schreuder, 1992). The authors interpret this late effect as indicating disconfirmation of the prediction (e.g., encountering *relaxation* rather than the predicted *rest*); see Section 4.4.

Another study suggests prediction of semantic features associated with shape. Rommers, Meyer, Praamstra, and Huettig (2013) presented participants with high-constraint

[10] Note that the N400 reduction was not diminished at the shorter SOA of 250ms if participants experienced the 500ms SOA before the 250ms SOA, perhaps because they were able to recruit additional resources to predict more quickly.

contexts (e.g., about the lunar landing; average Cloze of 72%) followed by the predictable word (*moon*), an unpredictable word referring to an object related in shape to the predictable object (*tomato*), or an unpredictable and unrelated word from the same category as the shape-related word (*rice*). The negative wave in the shape-related condition was smaller than in the unpredictable condition. The effect occurred later (500-700ms) than the standard N400, a finding which may reflect the fact that N400 effects often occur late in auditory experiments. In addition, an explanation in terms of prediction is supported by the fact that Rommers et al. also conducted a visual world experiment, which showed that people look at shape-related competitors before they hear the predictable word.

A Polish study relating to both semantics and syntax (Szewczyk & Schriefers, 2013) may be interpreted as showing that people predict animacy. This study used discourse contexts that constrained towards an animate or inanimate noun, and were either highly constraining toward a specific noun (Cloze: 89%) or less so (Cloze: 32%). The noun was preceded by an adjective that was marked for animacy. This adjective elicited a smaller N400 when it was compatible with the discourse context than when it was not. Since the effect did not depend on the predictability of a specific noun, it is possible that participants simply found it easier to integrate the adjective whose animacy was more compatible with the context. However, it is likely that comprehenders predicted animacy independent of a specific word (in our terms, that they predicted +ANIMATE), or alternatively that different comprehenders predicted different individual animate concepts rather than all animate concepts together. In addition, the study suggested that comprehenders predicted a syntactic feature (the animacy marking), because the effect depended on the syntactic match versus mismatch (cf. Wicha et al., 2004). This relationship between semantic and syntactic prediction (i.e., a link between levels) is of course compatible with prediction-by-production.

Kwon, Sturt, and Liu (2017) had Mandarin speakers read highly constraining sentence contexts (85% Cloze) followed by either the predictable noun, a related but unpredictable noun, or an unrelated noun. As in Federmeier and Kutas (1999), the related noun elicited a smaller N400 than the unrelated noun, an effect which we interpreted as showing prediction of the predictable noun. Moreover, the nouns were always preceded by a classifier. Importantly, when the classifier was not appropriate for the predictable noun, it elicited an N400 whose amplitude was smaller for classifiers congruent with related than unrelated nouns. Since Mandarin classifiers carry (some) semantic content, the effect on the classifier may reflect integration of the classifier with the preceding context, but it is more likely that it reflects pre-activation of the noun before the noun position (similarly to the animacy effect in Szewczyk & Schriefers, 2013).

Kim and Gilley (2013) had participants read ungrammatical sentences such as *The thief was caught by for police* and grammatical controls. For half of the participants, the ungrammaticality always resulted from the word *for*; for the other half, it resulted from one of seven words (*at, of, on, for, from, over, with*). For both groups, ungrammatical stimuli led to a negative deflection 170-270ms post-stimulus (an "N170"), which is compatible with integration because it did not occur early enough to rule out bottom-up processing.[11] But, in

[11] In a related study, Lau, Stroud, Plesch, and Phillips (2006) found that early syntactic anomaly effects (after around 200ms) were affected by whether the linguistic context predicted one particular syntactic category (99% Cloze for that syntactic category) for the upcoming word or was compatible with different syntactic categories. Specifically, an unpredictable preposition led to a stronger early anomaly effect if the context predicted a noun than if it was compatible with more than one syntactic category. But the effect did not occur early enough to rule out the possibility that comprehenders had time to access the syntactic category of the preposition and try to integrate it with the preceding context

addition, only the low-variability group showed a positive deflection 125-145ms post-stimulus (a "P1") that was localized to part of occipital cortex. It is just possible that the latter effect also reflects rapid bottom-up processing of a very frequent word that was repeatedly encountered in the experiment. However, it is more likely that participants in the low-variability group learned to predict the word *for*, including aspects of its form (e.g., its shape) at the critical sentence position, and determined whether this prediction matched visual input (similarly to Kim & Lai, 2012; see section 2.1). Similarly, Söderström, Horne, Frid and Roll (2016) investigated an early negativity elicited by word stem accents in Swedish and showed that its amplitude correlated with the predictability of the suffix that followed the stem.

Laszlo and Federmeier (2009) conducted a similar experiment to Federmeier and Kutas (1999), except that they used stimuli that were orthographically rather than semantically related to the most likely completion. As well as the predictable (89% Cloze) word (e.g., *bank*), they used an orthographic neighbor that was a word (*bark*), a (pronounceable) pseudoword (*pank*), or an illegal letter string (*bxnk*). For all three types of neighbors, the N400 amplitude was reduced in comparison to matched non-neighbors. As with Federmeier and Kutas, it is possible that this facilitation occurred because neighbors of the predictable word activated this word bottom-up and this word then integrated more easily with the context. But their data are more consistent with a predictive interpretation that is very similar to the one that we proposed for Federmeier and Kutas: Comprehenders predicted the semantics of the predictable word and then the predicted semantics led to activation of the predicted orthographic form. The predicted orthographic form subsequently facilitated processing of the orthographic neighbors of the predicted word, and these in turn provided support for the semantics of the predicted word (even though that word never occurred), which explains the reduced N400. Note that, if the effect were limited to words, it could be that people predicted orthographically related words without predicting orthographic representations; activation could spread from *bank* to *bark* without activation of *b*, *a*, or *k*, as long as orthographically similar words are linked to one another in the lexical network. However, the fact that the effect occurred with non-words (without lexical entries) suggests that comprehenders rather predicted individual graphemes from semantics (of the predicted word), in accord with the directionality of prediction-by-production (i.e., semantics to form).[12]

Finally, Ito, Corley, Pickering, Martin, and Nieuwland (2016) investigated both words related in form to a highly predictable word (Cloze: 94%) and words related in meaning to that word, in experiments that used a 500ms/word or a 700ms/word presentation rate. Meaning-related words led to a reduced N400 (relative to unrelated words) at both

---

(particularly as prepositions are very high-frequency words, and lexical access may proceed more rapidly than for typical open-class words).

[12] Vissers, Chwilla, and Kolk (2006) found that a misspelt word whose pronunciation is identical to a highly predictable word (e.g., *boekun* for *boeken*, 'books'; Cloze: 91%) elicited a P600 ERP effect, but this effect did not occur when the word was less predictable (Cloze: 0%). This finding is compatible with prediction of phonological (or orthographic) form, but it may of course reflect reanalysis as a consequence of integration difficulty. They also found an earlier effect (a N270) during comprehension of misspelt words, but this is not sufficiently early to rule out bottom-up activation, and it occurred only when the word was less predictable (cf. Newman & Connolly, 2004), which is difficult to reconcile with an explanation in terms of prediction.

presentation rates, but form-related words led to a reduced N400 (relative to unrelated words) only at the slower presentation rate. These findings suggest that comprehenders did not have the time (or resources) to predict form at the faster presentation rate. They therefore support the optionality of prediction-by-production. Note that Laszlo and Federmeier (2009) did find evidence for form-based prediction at a 500ms SOA. However, it is difficult to make direct comparisons between studies using different items: Even if the Cloze values for the target word are similar, the target may become predictable earlier in one study than another, which may in turn allow more rapid form-based predictions.

Does the electrophysiological evidence support prediction-by-production? (3.1.3) We have now reviewed a large number of ERP studies (both here and in section 2.1) that either demonstrate prediction or can be interpreted in terms of prediction. We argue these studies support our model and specifically our proposal that the most important mechanism for prediction during comprehension is prediction-by-production. First, the vast majority of studies show that comprehenders predict continuations that are constrained by the linguistic context so far and fit with the derived intention. We discussed a single study (Metusalem et al., 2012) that is not compatible with this conclusion, as it showed comprehenders predicted a word which was not a plausible continuation for the sentence context (*build a….jacket*). This study therefore provides evidence for prediction-by-association. But the bulk of the evidence supports prediction-by-production.

Second, many of the studies suggest that comprehenders predict semantics and, on the basis of a predicted semantic representation, activate representations at other processing levels which follow on from semantics during language production. Some studies show that predicted semantics can lead to the activation of syntax (Van Berkum et al., 2005; Wicha et al., 2004; Otten et al., 2007; Otten & Van Berkum, 2008) and others provide more or less definitive evidence that predicted semantics can lead to activation of form (DeLong et al., 2005; Ito et al., 2017; Kim & Lai, 2012; Dikker & Pylkkänen, 2011, 2013; Fruchter et al., 2015; Laszlo & Federmeier, 2009). Moreover, predicted syntax can also lead to the activation of form (Dikker et al., 2010; and perhaps Boylan et al., 2014, Kim & Gilley, 2013). Taken together, these findings support prediction-by-production, as they show that comprehenders' predictions are compatible with a directional flow of information, proceeding from semantics, to syntax, and then to form.

Finally, Ito et al.'s (2016) finding that semantic prediction occurred at a faster SOA compared to prediction of form supports prediction-by-production as it suggests that predictions involving later production stages take longer than earlier production stages. In sum, electrophysiological and neuroscientific evidence supports prediction-by-production.

Eye-Tracking (and Other Behavioral) Studies of Reading (3.2)

We now consider the evidence for prediction in studies of reading. We first discuss how parafoveal preview relates to prediction of upcoming words (3.2.1). We then argue that the way in which readers plan eye movements strongly suggests that they predict processing difficulty before lexical access occurs (3.2.2), and propose that eye-tracking evidence for prediction is consistent with prediction-by-production (3.2.3). Finally, we address the broader language comprehension literature that suggests that readers predict syntactic and semantic properties of sentences (3.2.4).

Prediction and preview. (3.2.1) It is well known that readers preview upcoming text: They take in spaces between words well ahead of the point of fixation, and to a lesser extent information about letter shape, orthographic and phonological regularity, and lexical information (e.g., Hohenstein & Kliegl, 2014). These preview effects of course do not demonstrate prediction, as previewed information is part of the input. Readers also skip more predictable words (typically after they fixate toward the end of the previous word) more often than less predictable words (Balota, Pollatsek, & Rayner, 1985; Rayner et al., 2011; Rayner & Well, 1996). This effect must be due to processing before the target word is fixated, but it could be due to preview of the target word. In fact, in Balota et al., the context was constant across conditions and the target word varied; thus, the reader must (logically) have taken in some information about the target word while fixating a previous word and so the skipping effect may be due to integration of previewed information. In Rayner et al., the context varied and the target word was constant. In this study, readers could have predicted the target word more in the more predictive (70% Cloze or higher) context than the less predictive (17% Cloze or lower) context. However, the effect could equally well be due to integration of previewed information about the target word with the context. Moreover, Drieghe, Rayner, and Pollatsek (2005) used a contingent-change paradigm in which a different word or non-word initially appeared in the location of a predictable target word (64% Cloze). This stimulus changed into the target word when an eye movement left the pre-target word. Participants skipped the target word more often when the target word did not change than when it did. This finding indicates that participants process predictable words, at least to some extent, before they fixate them, and hence means that word skipping, by itself, does not demonstrate prediction. To do so, it would be necessary to show that skipping was more likely for more predictable than less predictable words that cannot be previewed. To our knowledge, this effect has not been found.

Predictability, frequency, and the familiarity check. (3.2.2) Very roughly, readers of English and languages with similar orthographies tend to fixate on most words for roughly 200-250ms, and then typically perform a rapid saccadic eye movement to the next word (though some words are fixated more than once). They primarily take in information about the fixated word, and are affected very rapidly by its characteristics (e.g., font, spelling, frequency, or contextual plausibility; Rayner, 1998). As already noted (section 1.1), readers fixate less predictable words for longer than more predictable words (Ehrlich & Rayner, 1981; Rayner, Slattery, Drieghe, & Liversedge, 2011; Rayner & Well, 1996). Moreover, their first fixations on infrequent words are longer than on frequent words (e.g., Inhoff & Rayner, 1986). It might appear that such first-fixation effects could be due to integration rather than prediction because the effect is measured on the predictable word itself (and this word may even have been processed during the previous fixation, in studies that allow preview) and first fixations are longer than lexical access time.

But there is a problem with this explanation: Readers need 175-200ms to program their saccades (Rayner, Slowiaczek, Clifton, & Bertera, 1983). Such planning therefore takes place early in a fixation. Now, the two most comprehensive and implemented models of eye movement behavior during reading (E-Z Reader, e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998; and SWIFT, e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005) assume that word frequency affects saccade programming. In other words, the reader must have access to frequency information well before they could have extracted that information bottom-up. It therefore appears that frequency affects the decision to move the eye well before readers get to the point in lexical access at which frequency exerts an influence.

In particular, the E-Z Reader model provides a formalized account of this process. Readers use high spatial frequency information to perform a rapid "familiarity check" (*L1 stage*). They then use the result of this check, together with low spatial frequency information about word boundaries, to plan a saccade. At this point, the plan is labile (i.e., subject to change). The reader then continues to lexical access (*L2 stage*) and contextual integration, and sometimes uses the result of this integration to change the saccadic plan. In other words, the saccade typically depends on the familiarity check, but can be affected by subsequent processing (e.g., if the sentence becomes ungrammatical or implausible).

The familiarity check could be explained without prediction if apparent frequency effects were actually due to form properties that correlate with frequency, such as orthographic familiarity. White (2008) found some evidence of a small effect of orthographic familiarity on first-fixation time (and later measures), suggesting some direct effect of orthography on saccade planning. But she found robust frequency effects on first-fixation time when orthographic familiarity was controlled. So early frequency effects cannot be explained by properties of the word that are independent of lexical access. Readers must therefore regularly plan their movement before lexical access, but in a way that appears to be affected by frequency. For this to be possible, they must predict word frequency.

There is another way in which frequency effects in reading suggest a role for prediction. It is known that predictability and frequency independently affect average fixation times (e.g., Rayner, Ashby, Pollatsek, & Reichle, 2004; Kretzschmar, Schlesewsky, & Staub, 2015).[13] They also independently affect distributions of fixations: They both influence the central tendency of the distribution of fixations (i.e., the μ-parameter in an ex-Gaussian distribution) but only frequency influences the distribution's right tail (i.e., the τ-parameter). This latter finding occurs because low frequency words sometimes lead to abnormally long fixations (e.g., Staub, White, Drieghe, Hollway, & Rayner, 2010), but unpredictable words do not (e.g., Staub, 2011). To the extent that frequency effects at least partly reflect processing that takes place after encountering a word, the finding that predictability effects are independent of frequency effects provides some support for the claim that the effects of predictability reflect processing that takes place before encountering a word – in other words, prediction.[14]

Parallel lexical predictions. (3.2.3) Traditionally, most researchers assume that readers can predict only a single word (*serial prediction*). If so, there should be a linear relationship between reading time and degree of predictability (see Schotter, Lee, Reiderman, & Rayner, 2015). But in fact fixation times differ much more between low-Cloze (4%) and medium-Cloze (41%) contexts than between medium- and high-Cloze (86%) contexts (e.g., Rayner & Well, 1996). Moreover, Smith and Levy (2013) estimated word predictability in a corpus using trigram probabilities extracted from a much larger corpus, and compared these results with reading times. After controlling for factors such as word length and frequency, they found a logarithmic relationship between reading time and predictability on both the target word and the following word, for predictability values from 1 down to 10^{-6}. (They found similar results using self-paced reading, except that the effects were delayed, occurring on the following three words but not the target word.)

Smith and Levy's (2013) findings are compatible with Rayner and Well (1996), but in addition they found that reading times were slower for extremely unlikely words than very unlikely words – in itself a quite remarkable result. Their results are incompatible with serial

---

[13] Note that Kretzschmar et al. (2015) found effects of predictability but no effects of frequency on the N400 in an ERP experiment that was parallel to their eye-tracking study. This difference (which is consistent with an extensive literature) means that the different methods are sensitive to different processes.

[14] A large self-paced reading study also found support for this conclusion (Brothers, Swaab, & Traxler, 2017). Participants read predictable and unpredictable target words in the context of experimental lists that contained lower or higher proportions of predictable words. The target words also varied in frequency. While the frequency effect was unaffected, the predictability effect was larger when the proportion of sentences completed by predictable words was higher, suggesting that predictability effects may be generated by a different process than the one that is responsible for frequency effects.

prediction and they argued that comprehenders predict a very large set of words (i.e., at least including words that would occur once in 100,000).

These findings are not unambiguously due to prediction (rather than integration), but are in any case compatible with prediction-by-production. Although a reader using prediction-by-production would ultimately predict the semantic representation of a single word (just as a speaker eventually selects a single lemma), she would first activate a whole network of semantic representations, which in turn lead to the activation of many lemmas. Therefore, prediction-by-production is compatible with parallel activation of concepts and lemmas.

Prediction during sentence processing. (3.2.4) Traditional theories of how comprehenders syntactically analyze and interpret sentences have been framed in terms of incremental processing and its limits (e.g., Frazier, 1987; MacDonald, Pearlmutter, & Seidenberg, 1994). More recent accounts make use of the notions of prediction and predictability (e.g., Levy, 2008; see Section 1.1.1). Many important findings that are explained in terms of prediction are equally compatible with integration (e.g., Chow et al., 2016), but some provide clearer evidence for prediction.

Gibson's (1998) influential *Syntactic Prediction Locality Theory* of linguistic complexity proposed that "the longer a predicted category must be kept in memory before the prediction is satisfied, the greater is the cost for maintaining that prediction" (abstract, p. 1). For example, Chen, Gibson, and Wolf (2005) found that readers had particular difficulty with the underlined phrase in *The realization that the implication that the company planned the layoff was not just a rumor caused a panic*. On Gibson's account, the difficulty occurs because readers have predicted that two verbs are still required. In support of this proposal, the same phrase was easier to process in otherwise similar sentences in which no verbs were required (*The employee realized that the boss implied that the company planned the layoff* …). Similar findings occur for Japanese, a language with different orders of verbs and noun phrases (Nakatani & Gibson, 2010). These results are compatible with our proposal that prediction-by-production is resource intensive (section 2.4).

Boland, Tanenhaus, Garnsey, and Carlson (1995) had participants read a sentence word-by-word and simultaneously judge if it made sense (i.e., they pressed "yes" or "no" as each word appeared). In one experiment, participants often judged sentences such as *Which car salesmen did Harriet distribute the science exam papers to in class?* as not making sense before the preposition *to* (in contrast to equivalent plausible sentences). As verbs such as *distribute* require a prepositional object including the preposition *to*, participants must have predicted this preposition in order to determine *car salesmen* was an implausible recipient of *distribute*. Similarly, participants did not judge *Which movie did the girl remind* … as less plausible than *Which child did the girl remind* … (whereas they did judge *Which stone did the assistant watch* … as less plausible than *Which star did the assistant watch* …). They appear to have considered an interpretation in which *remind* introduces a clause (e.g., *which movie did the girl remind them to watch?*). To do this, they must have predicted a clause containing a verb for which *movie* is a plausible object (see also Altmann, 1999). Pickering and Traxler (2001) found similar effects in normal reading, thus ruling out the possibility that such prediction depends on the stop-making-sense paradigm. In conclusion, these results appear to require prediction of a specific word (*to*) or types of words (contextually plausible transitive verbs such as *watch* after *which movie did the girl remind* …).

Staub and Clifton (2006) found that reading *or the subway* was faster following *the team took either the train* … than following *the team took the train* …. They argued that without *either*, readers assumed that *the train* was the complete object of *took* and therefore had to reanalyze when they encountered *or the subway*. But *either* led them to predict a conjoined object (i.e., *or* followed by another noun phrase). An integration account (i.e., that *either* makes a conjoined object predictable but not predicted) is unlikely because it would require reanalysis to occur both with and without *either*, but for the reanalysis with *either* to be unproblematic. Even if comprehenders predicted meaning (i.e., disjunction) or the word *or*, they presumably predicted syntactically as well.

A particular concern about investigations of the prediction of syntactic structure is that the interpretation of results as predictive or not often depends on syntactic assumptions. To give one example, Traxler and Pickering (1996) found that readers experienced difficulty with *That's the garage with which the heartless killer shot the hapless man yesterday afternoon* at the verb *shot*. According to theories based on transformational grammar, readers associate the filler *with which* with an empty category after *man* before they can interpret the sentence. So the effect at *shot* means that they predicted this empty category (Gibson & Hickok, 1993). But in a linguistic theory without empty categories, the filler is directly associated with the verb and no prediction is necessary (Pickering & Barry, 1991). Even if a study demonstrates some form of prediction, its characterization may depend on linguistic assumptions (e.g., in studies of ellipsis; Yoshida, Dickey, & Sturt, 2013). In sum, it appears that syntactic structure is predicted during sentence processing, but it is difficult to draw specific conclusions in the absence of clear agreement about the nature of parsing.[15]

Prediction Involving Non-Linguistic Contexts (3.3)

Comprehenders can quickly integrate non-linguistic and linguistic information, for example rapidly experiencing anomaly when performing sentence-picture matching (e.g., H. H. Clark & Chase, 1972) or when hearing *Every evening I drink some wine before I go to sleep* uttered in a child's voice (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). In this example, the non-linguistic context may have been used integratively, but we proposed (section 2.3) that people can also use it predictively. For example (section 2.6 and Fig. 2), people could predict *gunman* instead of *robber* when they hear *When news broke about the break-in, John arrived and arrested the* … and at the same time see a man with a gun. In fact, much research has investigated linguistic prediction in situations that combine linguistic and non-linguistic contexts, where the non-linguistic context provides "scaffolding" that may facilitate prediction and help the comprehender determine the derived intention.

[15] Wright and Garrett (1984) arguably provided behavioral evidence for syntactic prediction. They presented participants with a sentence fragment word by word, and then had them make a lexical decision to a target word. Decisions were faster when the context (e.g., *The crowd near the church indicates that an important funeral*) was syntactically congruent with the target word (*translates*) than when it was incongruent (*translation*), even though the target word was always semantically incongruent. Clearly, a verb continuation is syntactically predictable after this context, and so comprehenders may have predicted a verb phrase and then have processed a compatible upcoming word (*translates*) easily. Although the facilitatory effect occurred on the predictable word itself, there is some reason to believe that the effect is due to prediction (rather than integration). This is because there is extensive evidence that syntax is not processed strictly before semantics (e.g., Marslen-Wilson, 1973; Trueswell et al., 1994). Therefore, difficulty at the semantic integration stage should have led to just as much disruption for the syntactically congruent but semantically incongruent target word as for the syntactically and semantically incongruent target word. Instead, the fact that processing of syntactically congruent words was easier despite the fact they were semantically incongruent suggests facilitation at the syntactic level occurred very early (i.e., before any semantic processing took place), and this in turn suggests that readers had predicted a verb before processing *translates*.

Most relevant studies use the visual world paradigm and provide strong evidence for prediction-by-production. We have already reviewed many of these studies in Section 2 (e.g., Altmann & Kamide, 1999; Kamide et al., 2003; Chambers & San Juan, 2008), but did not include all the evidence in that section in the interest of readability. We instead review it in full here. The review is organized into two parts. The first part (3.3.1) demonstrates that many different aspects of the linguistic context constrain comprehenders' prediction via covert imitation. The second part focuses on how comprehenders use non-linguistic context to derive the speaker's underlying intention (3.3.2). [16]

Evidence from the visual-world paradigm supports covert imitation. (3.3.1)

In section 2.3.1, we pointed out that in Kamide et al.'s (2003a) Japanese experiment, participants looked more at a likely theme (the hamburger) when they had heard *customer* in the dative (and thus typically a recipient) than when had heard it in the accusative, thus showing that their predictions were constrained by covert imitation of the context (see also Hintz et al., 2017; Arai et al., 2007 in Section 2.3.3). Similarly, Kamide, Scheepers, and Altmann (2003b) found that participants looked at a cabbage after the German sentence *the hare-NOM eat shortly* … ("The hare will shortly eat …") but at a fox after *the hare-ACC eat shortly* … ("… will shortly eat the hare"). Participants' predictions therefore depended on the thematic role (agent or patient) that they had ascribed to the hare, and indicate that they must have covertly imitated the context.

In addition, Boland (2005) found that people predicted potential arguments more often than potential adjuncts – a finding that suggests that covert imitation of the linguistic context constrain which predictions are made. Kaiser and Trueswell (2004) presented participants with a linguistic context that described one exemplar of a category, such as a sitting nurse. In Finnish, an object-verb-subject order is used when the subject is discourse-new. When participants then heard *the doctor-OBJ glances at the nurse-SUBJ* (in contrast to *the doctor-SUBJ glances at the nurse-OBJ*) they preferentially looked at a different exemplar, in this case a standing nurse, before they could recognize *the nurse*. Thus, a combination of information about discourse status and case marking can drive predictions, which is again compatible with predictions being constrained by covert imitation of the linguistic context.

Moreover, Weber, Grice, and Crocker (2006) had participants listen to German sentences such as *The cat*-AMB *chases possibly the bird-ACC/the dog-NOM* ("The cat possibly chases the bird"/"The dog possibly chases the cat"), in which *the cat* is ambiguous until the last noun phrase makes its role clear. At the adverb, participants tended to launch predictive eye-movements towards the plausible object (bird) when the sentence had the appropriate intonation for subject-verb-object order. However, when the sentence had the appropriate intonation for object-verb-subject order, this preference disappeared. This study indicates that prosody is another aspect of the linguistic context used by comprehenders to constrain their predictions (see also Hirose & Mazuka, 2015, who found that listeners used stress on a noun to predict whether it is the first noun of a Japanese compound).

Finally, comprehenders also take into account what is implied by the linguistic context (e.g., via scalar implicatures) to constrain their predictions. Kim, Gunlogson, Tanehaus, and Runner (2015) had participants hear discourses such as *Neill has some apples and pears. Jeff only has some apples*, and found that they looked at the apples or pears before hearing the final word to a greater extent than when *only* was removed. Additionally, in

[16] In theory, the visual world paradigm can also provide evidence about the time course of prediction. If eye movements consistent with prediction of semantics occur before eye movements consistent with prediction of form, it would provide additional evidence for prediction-by-production.

Kurumada, Brown, Bibyk, Pontillo, and Tanenhaus (2014), participants heard sentences such as *It looks like a zebra* with the focus (emphasis) on *zebra*, which implies that it is a zebra, or on *looks*, which implies that it is not a zebra. Before they heard *zebra*, they were more likely to fixate an unfamiliar animal that resembled a zebra when the focus was on *looks* than when it was on *zebra*.

Visual-word studies show how comprehenders use non-linguistic context to compute the derived intention. (3.3.2) Visual-world studies are also informative about the role of non-linguistic context in constraining the derived intention (see Fig. 2, arrow from non-linguistic context to derived intention). The non-linguistic context includes both the shared visual context and shared background knowledge. As an example of how the shared visual context may affect the derived intention, Knoeferle, Crocker, Scheepers, and Pickering (2005) presented participants with pictures of a princess washing a pirate and a fencer painting that princess. After hearing *The princess-AMB washes*, they tended to look at the pirate; after hearing *The princess-AMB paints*, they tended to look at the fencer. They therefore combined linguistic and visual information to interpret the princess as the agent of the washing-event or the patient of the painting-event, and used their interpretation to look at the other relevant entity before it was mentioned, presumably because they predicted its mention.

Similarly, Knoeferle and Crocker (2006, 2007) had participants listen to object-verb-subject sentences in German, such as *The pilot-ACC spies-on soon the…*, while looking at a scene containing the patient (pilot) and two other characters (a wizard and a detective). In the critical condition, one of the characters was a prototypical agent for the action described by the verb (detective), but the other character (wizard) was the actual agent (i.e., performed the action). Comprehenders looked at the actual agent shortly after they heard the verb (and before hearing *wizard*), rather than the prototypical agent (Knoeferle & Crocker, 2006; Experiment 2). (Instead, they looked at the prototypical agent when they listened to the same utterance while watching a display that did not contain the depicted agent). It is possible that comprehenders were incrementally interpreting the verb (*spied-on*) and simply looked at the picture that depicted the relevant event (spying), but we suggest that comprehenders assumed the speaker would refer to the visual context, and therefore used the visual context to assist in deriving the communicative intention.

Further, Kaiser and Trueswell's (2004) Finnish study suggests that people predict reference and not merely lexical information (as *the nurse* could refer to one of two exemplars). Thus, people do not just predict the words that they are likely to hear but also what entities they believe speakers are likely to refer to (i.e., they predict the referent of the speaker's communicative act; see Van Berkum, 2013). Similarly, Altmann and Kamide (2007) showed that comprehenders do not merely predict words but also their likely reference (and also demonstrated that predictions can make use of verb tense): Participants tended to look at an empty glass following *The man has drunk…* but at a full glass following *The man will drink…*

Moreover, predictions appear to make use of a mental (situation) model, which forms part of the derived intention, and incorporates shared background knowledge about the world. For example, Altmann and Kamide (2009) showed that predictive eye movements can target locations that are consistent with such a mental representation, even if they do not correspond to the actual location of the object. After hearing *The woman will put the glass on the table. The woman will pick up the bottle and pour …*, participants looked at the table rather than the actual location of the glass. This did not happen when the first sentence was *The woman was too lazy to put the glass on the table*.

Similarly, Lowder and Ferreira (2016) had participants who heard *The meat was pretty bland, so the chef reached for some salt uh I mean …* were more likely to look at a related alternative (pepper) than participants who heard *The meat was pretty bland, so the chef reached for some salt and also ….* Another group of listeners interpreted the disfluent utterance in much the same way as an utterance which included an explicit negation (…*so the chef reached for not some salt but rather…*). Therefore, participants appeared to use the presence of a disfluency to infer the speaker's intention and predict that she would now mention a different but related entity. Thus, predictions can draw on shared knowledge about the process of speaking itself.

The above findings do not demonstrate that the derived intention is linked to the speaker rather than the comprehender. In order to show that comprehenders maintain a representation of the speaker's intention that is distinct from their own, and that they (at least sometimes) apply self-other adjustments to compensate for differences between themselves and the speaker (see dashed "Self-Other Adjustments" arrow to the derived intention in Fig. 2), we turn to studies that manipulated whether the comprehender shared the same level of knowledge or ability with the speaker. Recall that Chambers and San Juan (2008) showed that listeners took into account their beliefs about what the speaker could or could not see when interpreting instructions to move objects around a grid (see section 1.3.2). In addition, Arnold, Hudson Kam, and Tanenhaus (2007) found that comprehenders looked at an unfamiliar object rather than a familiar object more after the disfluent *Click on thee …uh … red* than the fluent *Click on the red …*(see also Heller, Arnold, Klein, & Tanenhaus, 2015). However, the effect did not occur when they were told that the speaker had object agnosia (and hence had a plausible reason to be disfluent before a familiar object). Thus, the effect cannot be due to an association between disfluencies and difficult-to-name objects.

Similarly, Bosker, Quené, Sanders, and de Jong (2014) found that participants tended to look at an object with a low-frequency name after hearing a disfluency produced by a native speaker. Importantly, participants did not tend to look at such an object when the disfluency was produced by a non-native speaker (as indicated by a clear foreign accent). As in Chambers and San Juan (2008), these results imply that the comprehender can adjust for differences between herself and her beliefs about the speaker (i.e., having agnosia or being non-native), in accordance with prediction-by-production. Together, these studies suggest that comprehenders derive the speaker's intention and in doing so take into account characteristics of the speaker.

In conclusion, the evidence about prediction involving non-linguistic contexts is particularly strong. Comprehenders make extensive predictions based on different aspects of their knowledge of the language (e.g., grammar, meaning), predict reference and not merely words, and take into account their assumptions about speaker intention. These findings therefore allow us to draw two main conclusions. First, they indicate that the comprehender derives the intention by integrating linguistic processing (including using covert imitation) and non-linguistic context (see Fig. 2). Second, they support the prediction-by-production account in which the comprehender derives the speaker's intention and then uses that intention to predict upcoming language.

Speech (3.4)

As mentioned in Section 1.1, many experiments show that a predictive context influences perception of ambiguous sounds (e.g., Ganong, 1980) or degraded speech (Miller, Heise, & Lichten, 1951; Samuel, 1981). In these studies, context sometimes refers to the lexical item a sound appears in, and sometimes to the larger (i.e., sentential or discourse) context. As other contextual facilitation effects (e.g., shorter reading times for words in high-cloze sentences), these findings could be due to prediction of the target sound, but they could also be due to easier integration of the target sound.

In the speech-comprehension literature, the proposal that context effects are integrative is characteristic of feedforward or bottom-up accounts (e.g., Norris et al., 2000; McQueen, Cutler, & Norris, 2003; Norris, McQueen, & Cutler, 2003; McQueen, Norris, & Cutler, 2006; McQueen, Jesse, & Norris, 2009). These accounts propose that the activation of sound-based representations is initially based only on acoustic processing of the target sound, and is unaffected by context. Activation then flows from sound-based to lexical and semantic representations, and it is only at this later stage that activation of these higher-level representations can be influenced by context. In contrast, interactive accounts propose that context effects are due to prediction of sound-based representations (e.g., Elman & McClelland, 1988; McClelland, Mirman, & Holt, 2006). These accounts propose that lexical and semantic representations can potentially affect the earliest stages of acoustic processing.

As testified by a long-lasting debate, distinguishing between these two types of account is extremely difficult. Recent studies provide some evidence that comprehenders can indeed predict upcoming speech sounds, and that they do so using prediction-by-production, though these studies' reliance on inference from activation of specific brain areas makes them dependent on assumptions about the localization of specific aspects of comprehension. Below, we review these findings, first in relation to ambiguous speech and speech in noise (3.4.1), and then in relation to cases when individual speech sounds are replaced by noise or silence (3.4.2). Finally, we consider the evidence for prediction of sounds due to coarticulation (3.4.3) and for motor activation during speech comprehension (3.4.4). To foreshadow, we argue that the evidence from speech in noise, and to a lesser extent from speech sounds replaced by noise or silence and from coarticulation, indicates that sounds are predicted during speech; there is also some evidence that such predictions involve production mechanisms, particularly from studies that show motor activation. Of course, most of the evidence we review uses ambiguous, noisy, or otherwise manipulated speech, and it is an open question to what extent this evidence is representative of comprehension under less adverse conditions.

Ambiguous speech and speech in noise. (3.4.1) When comprehenders hear a sound that is acoustically "mid-way" between two phonemes, they tend to categorize it in line with the context. For example, recall that the same ambiguous fricative is categorized more often as /s/ after *tremendou-*, but as /ʃ/ after *repleni-* (Ganong, 1980; Samuel, 2001). Similarly, the same ambiguous velar stop at the beginning of *-oat* is categorized more often as /g/ when it follows a context that makes *goat* highly predictable (e.g., *The busy dairyman hurried to milk the…*), and more often as /k/ when the context makes *coat* highly predictable (e.g., *The careful laundress had to dry-clean the…*; Borsky, Tuller, & Shapiro, 1998). In addition, words in noise are more likely to be accurately identified if they are more predictable than if they are less predictable (e.g., Bradlow & Alexander, 2007; see Mattys, Davis, Bradlow, & Scott, 2012). But as we argued above, these contextual effects could reflect integration rather than prediction.

Elman and McClelland (1988) reported a finding that is potentially informative about prediction. When they hear a sound that is ambiguous between /t/ and /k/, comprehenders tend to report /k/ more often following /s/ than following /ʃ/ (probably as a result of compensation for co-articulation; Mann & Repp, 1981). Interestingly, Elman and McClelland showed that comprehenders' categorization of the ambiguous sound is similarly influenced by a context sound that is itself ambiguous (mid-way between /s/ and /ʃ/) as long as it is embedded in a biasing lexical context (e.g., *progre-*). They argued that the lexical node *progress* was activated by the fragment *progre-*; activation then flowed from *progress* to its component phonemes, including /s/ but not /ʃ/. The predicted /s/ phoneme, in turn, biased perception of the ambiguous /t/-/k/ sound. However, it is unclear whether these findings actually demonstrate prediction of /s/: The ambiguous /t/-/k/ sound may activate both /s/ and

/ʃ/, and later /s/ is selected because it integrates more easily with the context. Therefore, Elman and McClelland's study does not demonstrate prediction. To demonstrate prediction of sound using ambiguous stimuli, it would be necessary to show that processing at the lexical level temporally precedes and causally affects processing at the phonological level. [17]

In an fMRI study, Guediche, Salvata, and Blumstein (2013) had participants hear words with unambiguous or ambiguous initial consonants (e.g., consistent with *coat*/*goat*, as in Borsky et al., 1998), following sentence contexts either where *coat* and *goat* were equally predictable or where only one of them was predictable. There was an interaction between predictability and perceptual ambiguity in the superior temporal gyrus (STG), which is associated with acoustic processing, and in the middle temporal gyrus (MTG), which is associated with lexico-semantic processing (e.g., Friederici, 2012). The predictability effect in an acoustic area is compatible with prediction (see Obleser & Kotz, 2010, for a related finding using degraded sentences), but does not demonstrate prediction because this study did not indicate whether activation in MTG preceded activation in STG (which would support prediction) or *vice versa*. Moreover, when Davis, Ford, Kherif, and Johnsrude (2011) used time-resolved fMRI (an analysis technique that provides temporal as well as spatial information) to compare clear and degraded (i.e., with added signal-correlated noise) versions of both coherent (e.g., *The child left all of his lunch at home*) and anomalous sentences (e.g., *The thing felt all of his speech at line*), they found no evidence for prediction. Although an interaction between degradation and sentence coherence was present in both STG and the left inferior frontal gyrus (LIFG, which is associated with lexical processing; Friederici, 2012; Indefrey & Levelt, 2004), it emerged earlier in the former than the latter (i.e., contrary to the timecourse that would support prediction).

In contrast, Sohoglu, Peelle, Carlyon, and Davis (2012) found earlier activation in LIFG than in STG using combined EEG and MEG recording. They used noise-vocoded words (i.e., synthesized by using the speech amplitude envelope to modulate noise, across different frequency bands). Spoken words were preceded by a written stimulus that was neutral (i.e., a row of *X*s), or a word matching or mismatching the upcoming spoken stimulus. In the matching condition (compared to the neutral and mismatching conditions), brain activity in LIFG increased *before* activity in STG decreased. Moreover, the LIFG effect occurred as early as 90-130ms after the onset of the spoken stimulus. These findings provide strong evidence for prediction-by-production. Specifically, they suggest that predicted lexical representations led to prediction of phonological representations.

Gow and Olson (2015) used a design similar to Guediche et al. (2013) while recording combined MEG and EEG. They applied Granger causation analysis to identify the patterns of activity in other brain areas that could (statistically) explain activity observed over time in posterior STG (pSTG) during the comprehension of the ambiguous stimulus words (100-500ms post-stimulus). They found that several areas implicated in lexical and semantic processing (including part of the LIFG and left anterior MTG) could explain activity related to acoustic processing in pSTG, suggesting that semantic and lexical representations directly

[17] Even if it could be demonstrated that Elman and McClelland's (1988) findings are due to prediction, it is not clear whether it would be prediction-by-production (i.e., with activation spreading from lexical nodes to phonemes), or rather due to activation spreading within the phonological level from one phoneme to another. The latter could reflect listeners' knowledge of associations (i.e., transitional probabilities) between sounds (see Pitt & McQueen, 1998; Samuel & Pitt, 2003; Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McQueen et al., 2006 for discussion), and would thus constitute a form of prediction-by-association.

influence acoustic processing. Interestingly, the posterior MTG (pMTG), an area which maps from lexico-syntactic and semantic representations to word forms, influenced pSTG via the supramarginal gyrus (SMG). These findings suggest a flow of activation from higher-order to sound-based areas, which is again consistent with prediction-by-production. However, McQueen, Eisner, and Norris (2016) noted that there is substantial disagreement over the role of SMG in speech comprehension: This area may mediate articulatory-based and sound-based representations sublexically, rather than via a shared lexical entry, and so it is difficult to interpret the flow of activation in Gow and Olson's study.

Finally, Gagnepain, Henson, and Davis (2012) provided evidence for the prediction of specific sounds. Participants learned novel words (e.g., *formubo*) that were very similar to existing ones (i.e., *formula*). They were then tested on the following day, to allow the newly learned items to consolidate overnight. The purpose was to create a new uniqueness point (after /ˈfɔːmjʊ-/ for *formula*) by introducing a competitor in the participants' mental lexicon. Participants were also tested on novel words they had not learned and words they had learned but not consolidated (which presumably would not compete with existing words). Crucially, consolidated words elicited less activity in STG before the new uniqueness point, and they elicited more activity in STG after the new uniqueness point. This pattern is consistent with prediction. Before the new uniqueness point (at /ˈfɔːm-/), listeners activated two lexical entries (*formula* and *formubo*) that both pre-activated the upcoming diphthong /jʊ/ and therefore predicted these sounds more strongly, thus facilitating their processing in STG. In contrast, after the new uniqueness point (i.e., after /ˈfɔːmjʊ-/), listeners were more likely to predict the wrong sounds, and the resulting mismatch between predictions and sensory input led to increased activity in STG. In sum, findings from Gagnepain et al. and Sohoglu et al. (2012) strongly suggest that listeners predict upcoming speech sounds based on activated lexical representations, which is consistent with prediction-by-production.

Speech replaced by noise or silence. (3.4.2) When a single speech sound is replaced by white noise or a cough, listeners often fail to notice the sound is missing if the spectral characteristics of the replacement are sufficiently similar to those of the replaced sound (e.g., if the replaced sound is a fricative). This is known as the phoneme restoration effect (R. M. Warren, 1970). Although the fact that restoration is more likely for predictable than unpredictable words (Samuel, 1981, Experiment 3) could suggest that listeners predict the missing sound, the behavioral evidence for phoneme restoration is also consistent with an ease-of-integration explanation (although cf. Repp, 1992), because it relies on offline categorization responses made after listeners have heard the end of the word.

But in an fMRI study, Shahin, Bishop, and Miller (2009) found more activation in areas related to language production (LIFG, left pre-supplementary motor area [pre-SMA], and bilateral insula) during successful restoration than when listening to an intact stimulus (with the LIFG and insula showing more activation for words than pseudo-words). Leonard, Baud, Sjerps, and Chang (2016) used electrocorticography (ECoG) to compare brain activity (in STG) during perception of replaced and intact stimuli at electrodes that are known to discriminate between specific pairs of sounds (e.g., between /k/ in /fæktr/ (factor) and /s/ in /fæstr/ (faster)). They found that activation elicited within 150ms of the onset of the replaced sound (i.e., noise, represented by # in /fæ#tr/) closely matched the activation elicited by actually hearing the reported sound (as in the intact stimuli). Crucially, they also found that activation in a separate area (left frontal cortex) peaking 130ms before the onset of noise could be used to categorize what sound the participant would later report having perceived. These effects were similar whether the participants heard the words embedded in sentential contexts or in isolation and suggest that participants predicted the missing sound on the basis of having predicted a particular lexical item (either as a likely completion to a sentential context or because they had recently heard the intact version of that lexical item).[18] Thus, neuroscientific evidence on phoneme restoration of sounds replaced by noise supports the conclusion that prediction-by-production is implicated in the filling-in of sounds replaced by noise.

We now discuss ERP studies that investigated how words with missing phonemes are processed in real time. Sivonen and colleagues (Sivonen, Maess, Lattner, & Friederici, 2006; Sivonen, Maess, & Friederici, 2006) had participants listen to high-Cloze (80%) and low-Cloze (0%) sentences where the initial phoneme of the final word was either present or replaced by a cough or a silent gap. The N400 effect for manipulated words occurred later, but was no larger than for intact words (see also Groppe et al., 2010 for a replication of this finding using tones). This finding suggests that lexical access could proceed on the basis of partial acoustic information. In addition, manipulated words in high-Cloze sentences generated an N400-like effect compared to intact words, but only when the replaced sound was short (i.e., a plosive). According to the authors, such short gaps or coughs did not afford sufficient time for listeners to generate predictions about the target word, therefore causing additional difficulty with lexical retrieval. Such an interpretation is compatible with prediction of words, but does not demonstrate that the component sounds of the predicted word can be predicted as well.

Mattys, Pleydell-Pearce, Melhorn, and Whitecross (2005) investigated the brain's response when participants detected silent gaps while listening to highly predictable and weakly predictable words with an early or late uniqueness point. The amplitude of the N1 (a negative deflection peaking soon after 100ms and reflecting early auditory processing) elicited by the gap was largest for highly predictable words with late uniqueness points. One interpretation of this finding is that such words afford strong predictions for upcoming sounds (see Gagnepain et al., 2012). When the listeners encounter the silence, these predictions are disconfirmed, resulting in a larger N1. However, it is also possible that detecting gaps is easier when the word is easier to process because it is more predictable – an explanation that is compatible with an integration account (Mattys et al., 2005). More compellingly, Bendixen, Scharinger, Strauss, and Obleser (2014) showed larger Mismatch Negativity (MMN) responses to an omitted word-final sound when the word was highly predictable. The MMN is elicited without conscious attention, so it is unlikely that this finding can be explained by the fact that it is easier to detect gaps in more predictable words. Importantly, though, these findings do not show comprehenders predicted specific sounds but rather that they predicted the occurrence of sound (vs. silence). The latter possibility is of course consistent with other evidence that comprehenders are extremely sensitive to the timing of speech, for example that they show a smaller N400 in response to semantically anomalous words that occur in a regular than an irregular metrical context (Rothermich, Schmidt-Kassow, & Kotz, 2012).

Prediction due to coarticulation. (3.4.3) Listeners use co-articulatory information to quickly direct their attention to visual referents. In a visual-world study, Dahan, Magnuson, Tanenhaus, and Hogan (2001) had participants listen to instructions such as *Click on the net*, and found that they took longer to fixate the corresponding object when the initial syllable of the target word (*net*) had been cross-spliced from a different word (e.g., *neck*), thus showing that misleading co-articulatory cues slow down speech comprehension. More importantly, Salverda, Kleinschmidt, and Tanenhaus (2014) showed that listeners begin directing their attention towards the referent of a consonant-initial noun as soon as they hear co-articulatory

[18] Note that in this study the same replaced stimulus was heard as either intact word (e.g., as either *factor* or *faster*) at least 25% of the time, and so the effect cannot be due to residual acoustic information about the replaced sound carried by the surrounding sounds.

cues in the final vowel of the preceding word *the* (see Mahr, McMillan, Saffran, Weismer, & Edwards, 2015 for a similar study with toddlers). Gow and McMurray (2007) showed that listeners hearing *green boat* are quicker to look at a boat (rather than another green object) when the final nasal in the adjective is (appropriately) assimilated to a labial place of articulation than when it is not; this effect occurs very rapidly, starting from 140ms after the onset of *boat*. Finally, when observing a signed utterance that ended in a semantically unpredictable sign, users of German sign language showed an enhanced N400 effect whose onset began before the onset of the sign itself, during the transition from the previous sign (Hosemann, Herrmann, Steinbach, Bornkessel-Schlesewsky, & Schlesewsky, 2013); the transition is similar to the later portion of a vowel immediately preceding a consonant in the sense that, like the vowel, it carries information about the nature of the upcoming sign.

In sum, comprehenders make immediate use of co-articulation information. Does this mean that they predict upcoming speech sounds? Listeners who hear a nasal consonant with labial features may not predict a labial consonant. Rather, when they later encounter /b/, they may recognize this phoneme and integrate it with the context more easily when it follows a nasal with labial features, and this may in turn speed up looks towards the correct visual referent. But it is also possible that listeners use their knowledge about place assimilation to predict a labial consonant (or activation may spread from labial features to all labial phonemes). They then look towards the visual referent whose name begins with a labial consonant, before the labial consonant itself occurs. This would be of course be consistent with ERP and MEG evidence for the activation of form (e.g., Dikker et al., 2010; Kim & Lai, 2012). Thus, findings that show rapid use of co-articulation information to guide speech comprehension do not demonstrate prediction, but they do provide some support for it.

Motor activation during speech comprehension. (3.4.4) We have noted (section 2.2) that the language production system appears to be activated during comprehension (e.g., Fadiga et al., 2002; Pulvermüller et al., 2006), and particularly under adverse conditions (e.g., Adank, 2012). There is some evidence that such activation is predictive. D'Ausilio, Jarmolowska, Busan, Bufalari, and Craighero (2011) repeatedly exposed participants to a pseudoword (*birro* or *biffo*) while recording MEPs from their tongue. On most trials, they heard a prime pseudoword (e.g., *birro*) and 1s later the same pseudoword, either pronounced appropriately (i.e., the same as the prime) or inappropriately (with *bi* containing coarticulation cues appropriate for *biffo*). Specifically, they heard *bi*, then after 100ms received TMS to the tongue, and after 300-350ms heard the double consonant (e.g., *rro*). Upon TMS stimulation, they found immediate (within 8-11ms) activation of tongue muscles (associated with the articulation of *rr*) when they heard the appropriately but not inappropriately pronounced target. Thus, the appropriate articulators are active within just over 100ms of the offset of *bi*. This result could be due to predictive activation of *birro* (i.e., because the participant predicted that the prime will be repeated), which in turn led to prediction of the upcoming double-consonant while listening to *bi*, specifically via pre-activation of the articulators; though it is possible that perception of *bi* (when co-articulated with *rro*) activated the tongue muscles bottom-up.

D'Ausilio et al.'s (2011) findings provide some evidence for prediction of upcoming sounds and specifically for prediction-by-production. Interestingly, these effects are consistent with the facilitatory role of visual speech (i.e., observing the speaker's articulatory movements) on speech comprehension; visual speech might specifically support predictions of upcoming sounds, because articulatory movements sometimes precede the corresponding sounds by more than 100ms (see Skipper, Nusbaum, & Small, 2006; p. 252).

Predicting in Dialogue (3.5)

In dialogue, interlocutors may predict the content of what their partners are going to say, something which also takes place in passive comprehension (i.e., monologue). But they may also predict when their partner is likely to finish speaking, something which may help them to respond in a timely manner. We consider these two types of prediction and discuss how they may be related.

Predicting content in dialogue may be similar to predicting content in monologue. At present, we have no reason to believe that the mechanisms are different (e.g., Figs 1 and 2 should still hold). But dialogue is unlike most monologue, for example having many brief, fragmentary contributions rather than complete sentences, and so addressees may tend to predict different units from comprehenders in monologue. It also tends to be more repetitive than monologue (see Pickering & Garrod, 2004) and hence more predictable, and so addressees may particularly rely on prediction (cf. Brothers, Swaab, & Traxler, 2017). Note that the speaker might also predict the addressee's response as well as the addressee predicting the rest of the speaker's utterance (though we know of no studies on this issue).

Moreover, the addressee may have her production system "ready," because she may contribute during the speaker's utterance (providing "backchannel feedback") or later (e.g., with a response). This might facilitate prediction-by-production – a claim that is compatible with the evidence that activating the production system facilitates the processing of predictable utterances: Hintz, Meyer, and Huettig (2016) found sentences with predictable words were read faster than sentences with unpredictable words, but only when participants also named pictures following sentence contexts on other trials during the experiment. These effects may be due to prediction or integration, but they suggest that people use production mechanisms in comprehending predictable sentences, and those mechanisms can be primed when comprehenders use them in production. However, there are no direct comparisons of prediction in monologue and dialogue.

A few studies have investigated neural coupling (how activation in listeners' brains correlates with activation in the speaker's brain) using fMRI. Stephens, Silbert, and Hasson (2010) found that listeners whose brain activity precedes correlated activity in the speaker (i.e., exhibit stronger predictive coupling) also comprehend better. Dikker, Silbert, Hasson, and Zevin (2014) found increased coupling (in posterior STG) for more versus less predictable picture descriptions but there was no evidence that this effect was due to predictions by the listener. To our knowledge, no study has specifically investigated whether increased coupling in areas related to language production is related to enhanced prediction of the speaker's utterance.

In contrast, there is substantial research about when addressees predict their partners will finish speaking. Observational researchers assume that interlocutors regularly predict each other's turn-endings (e.g., Sacks et al., 1974), largely because most turn-transitions are very short. Stivers et al. (2009) found that turn-transitions across ten languages had a mean ranging between 0ms and 500ms and a mode ranging between 0ms and 200ms (see also De Ruiter, Mitterer, & Enfield, 2006, who reported 45% of transitions between −250ms and +250ms, for a corpus of Dutch telephone conversations). It takes at least 600ms to produce a single word (e.g., Indefrey & Levelt, 2004) and longer for a multi-word utterance (e.g., M. Smith & Wheeldon, 1999), so addressees cannot regularly wait till the speaker has finished before preparing their turn. This is even the case for prepared utterances, which take over 500ms (Ferreira, 1991).

An appealing explanation for short turn-transitions is that the addressee predicts when the speaker is likely to end and prepares a response in advance (e.g., Levinson, 2016). However, it is possible that listeners react to early cues present in the speaker's turn (e.g., aspects of prosody or speaker gaze) rather than predicting turn ends (Duncan, 1972; Heldner

& Edlund, 2010). Alternatively, they could respond on the basis of a point at which the utterance might have ended but did not (a so-called transition relevant place), for example in tag questions (e.g., *What do you want to do, Alex?*).

Clearer evidence for turn-end prediction comes from Magyari, Bastiaansen, De Ruiter, and Levinson (2014), who had participants listen to turns extracted from the corpus of Dutch telephone conversations used by De Ruiter et al. (2006). To assess the predictability of turns, they were cut at several points, and a norming group of participants provided completions. They were more likely to correctly complete some turns (predictable) than others (unpredictable), starting from 600ms before the actual end of the turn. Experimental participants were instructed to press a button exactly at turn end, and were encouraged to predict when this moment would occur. These listeners responded very close to turn end and, importantly, earlier when listening to predictable (-70ms) than unpredictable (+140ms) turns. Moreover, listeners' EEG recordings showed an earlier (starting at least 1250ms before turn end) power decrease in the beta frequency range for predictable than unpredictable turns, and this effect was localized to brain areas involved in directing attention to a moment in time and in syntactic and lexical processing. These findings thus suggest a role for lexical and syntactic information in the prediction of turn endings.

In De Ruiter et al. (2006), participants judged turn endings for unedited conversational turns and for turns without prosodic information (i.e., with flattened pitch) or without lexical information (i.e., low-pass filtered). The lexically edited turns led to earlier (and hence more inaccurate) responses, but the prosodically edited turns did not, suggesting that people predict turn-endings based on the content of what they hear. Magyari and De Ruiter (2012) had another group of participants provide completions to fragments of these turns (cut off at various points). They found that De Ruiter et al.'s participants made better turn-end judgments both for turns that the new participants were more likely to complete with the turn's actual ending (i.e., the same words) and for turns that the new participants were more likely to complete with the same number of words as the turn's actual ending. These findings suggest that people predict the content and length of endings.

Other studies suggest that prosody may also be implicated in turn-end predictions. Bögels and Torreira (2015) showed that listeners judged the end of questions such as *So, are you a student?* to be later when the question had been extracted from a longer question (*So, are you a student at Radboud University?*) than when it had not. As the words are the same, they concluded that participants must use prosodic cues to predict turn-endings. These cues may be particularly important when turns contain a transition-relevant place (just after *student* in this case), so that only prosody can inform listeners whether the speaker has finished speaking or will continue. Similarly, Lammertink, Casillas, Benders, Post, and Fikkert (2015) showed that adult (and child) listeners were most likely to switch their gaze from the current to the next speaker when both the prosody and the syntax suggested that the turn was complete, presumably as a result of prediction. However, there was a stronger effect when only syntax was complete than when only prosody was complete.

Overall, it is clear that people can predict when utterances will end using aspects of those utterances' linguistic content, including syntax and prosody. We cannot be certain whether such turn-end predictions rely mainly on prediction-by-production or on prediction-by-association. However, using prediction-by-production may have additional benefits for addressees, as it could help them prepare a response: If the speaker asks *What type of water would you like, still or sparkling?*, an addressee who predicts *sparkling* using production mechanisms would be able to produce a response more quickly, and might find it especially easy to produce the word *sparkling*.

As discussed above, listeners can use content predictions to drive predictions about the timing of turn-endings. But how can content predictions be transformed into timing predictions? Listeners might extract coarse timing estimates (e.g., number of words) or finer timing estimates (e.g., number of syllables or phonemes) from their predictions of content. However, the duration of words and syllables varies greatly with speech rate. Therefore, it may be that the listener tracks the speech rate of the current speaker (Garrod & Pickering, 2015; Wilson & Wilson, 2005), and thereby predicts upcoming timing (e.g., Dilley & Pitt, 2010). In fact, MEG evidence suggests that oscillatory entrainment to the speech signal in left auditory cortex is driven by oscillations in areas that include pre-motor cortex, suggesting that speech-rate tracking may be production-based (Park, Ince, Schyns, Thut, & Gross, 2015).

In conclusion, prediction may be particularly important as a means of facilitating smooth interactions (e.g., Scott, McGettigan, & Eisner, 2009) – it is hard to see how dialogue could function without prediction. Dialogue highlights the benefit of predicting the timing of utterances as well as their content. It may lead to enhanced activation of the production system (compared with monologue) so that addressees can be ready to respond when appropriate. As we have argued that prediction-by-production is central to most prediction during language comprehension, we propose that it should be enhanced during dialogue, and that forms of dialogue that require extensive response preparation (e.g., involving interrogatives) may further enhance prediction. Studies directly comparing monologue and dialogue would be valuable in testing these proposals.

Prediction in Different Populations (3.6)

Most studies consider prediction in young adult native speakers. In this section, we review studies that have instead considered prediction (using various methods) in different populations, specifically older adults and children. This literature is not extensive, but provides some evidence about the extent to which prediction is affected by knowledge and resources. In addition, we have already discussed the literature for non-native speakers in section 2.4.3, where we suggested that non-native speakers' predictions may be similar to native speakers' predictions at the semantic level but non-native speakers may be less likely to predict syntax than native speakers (Foucart et al., 2014, 2016; Mitsugi & MacWhinney, 2016) and even less likely to predict form (Ito et al., 2016; Martin et al., 2013). These findings support the optionality of prediction and also support prediction-by-production, with later stages being less likely to occur in non-native speakers as a consequence of their poorer proficiency. In that section, we also discussed the effects of reading skills on prediction, where we suggested that prediction may be more likely for skilled adult readers than readers with low literacy (Mishra et al., 2012) or dyslexia (Huettig & Brouwer, 2015). Note that there has been some interest in prediction during aphasic comprehension (e.g., Mack, Ji, & Thompson, 2013; T. Warren, Dickey, & Lei, 2016), but the limited evidence and the underlying differences across aphasics make any conclusions premature. Below we consider older adults (2.6.1) and children (2.6.2).

Older adults. (3.6.1) Some ERP evidence suggests that older adults predict less than younger adults. Federmeier et al. (2002) conducted a version of Federmeier and Kutas (1999) using auditory presentation, and found similar results for younger adults, but a smaller reduction for words related to the predictable word (e.g., *pines* for *palms*) in older adults, and (unlike Federmeier & Kutas, 1999) this effect occurred only when the context was weakly (rather than strongly) predictive. Interestingly, these age-related differences were driven by older adults with lower verbal fluency and vocabulary size, which is consistent with the resource-intensive nature of prediction-by-production (see also Federmeier et al., 2010, but cf. Wlotko et al., 2012). Finally, we noted in section 2.4.3 that DeLong et al. (2012) could not replicate DeLong et al. (2005)'s evidence for prediction of form with older adults (but see footnote 5). In summary, older adults appear to predict less than younger adults, both at the semantic and at the form level.[19]

Children. (3.6.2) More studies have investigated children's ability to predict, in part because of a theoretical proposal that prediction may underlie language learning (see F. Chang et al., 2006; Rabagliati, Gambi, & Pickering, 2015). Using a similar method to Altmann and Kamide (1999), Nation, Marshall, and Altmann (2003) found that 10-11 year olds looked at the target object (e.g., a cake) in a predictive context (*Jane watched her mother eat a cake*) well before the onset of the target noun (*cake*). In contrast to studies with adults (Huettig & Brouwer, 2015; Mishra et al., 2012), they found no relationship between reading skill and prediction. However, Mani and Huettig (2014) did find that 8 year olds who read words better predicted more.

As noted in section 2.3.1, Borovsky et al. (2012) had 3-10 year old children (and adults) listen to sentences such as *The pirate will chase the ship* while viewing pictures of a ship, treasure, a cat, and a bone. Similarly to Kamide et al. (2003), they found that listeners looked most at the predictable entity (a ship) and also more at entities related to the subject (treasure) or verb (a cat), in comparison to an unrelated entity (a bone). They also found a correlation between children's comprehension vocabulary and extent of prediction (see also Borovsky & Creel, 2014). Borovsky, Sweeney, Elman, and Fernald (2014) extended Borovsky et al.'s (2012) findings to novel events (e.g., *The monkey rides the bus*) that participants learned about from story books before they took part in the visual-world task. Interestingly, 3-to-4-year olds did not look at the most predictable entity (a bus) more than at an entity related to the verb (a car), whereas older children and adults did. This suggests that younger children may find it difficult to combine information from the subject and the verb to constrain their predictions when the event being described is novel (rather than already known), and hence that their predictions are affected by either knowledge or resource limitations.

Using two-object displays, Mani and Huettig (2012) showed prediction in two-year-olds, with their production vocabulary (rather than their comprehension vocabulary) correlating with the extent of prediction. Mani, Daum, and Huettig (2016) showed two-year-olds predict an object more when it is more strongly associated with the verb: For example, on hearing *read* toddlers looked more at a book (strongly associated) rather than a letter (weakly associated). Finally, Bobb, Huettig, and Mani (2016) showed that 30-month-olds predict shape-related information (similarly to adults, as shown by Rommers et al., 2015). Overall, thus, there is good evidence that children predict meaning from a very young age.

[19] Further ERP work has compared younger to older adults' comprehension abilities, but it is unclear whether the differences it uncovered are specifically related to prediction. Federmeier et al. (2007) had young participants read more (85% Cloze) or less (27% Cloze) predictive contexts followed by predictable or unpredictable (but plausible) target words; they found an N400 effect after both types of contexts, which was greater following the more predictive contexts. Wlotko, Federmeier, and Kutas (2012) found the same pattern with older adults (72 year olds), but the N400 effect was reduced (and somewhat delayed), and in fact there was no significant difference between predictable and unpredictable words following less predictive contexts. These effects could be due to integration, but they are compatible with the idea that older adults take advantage of contextual predictability less than young adults (cf. Wlotko & Federmeier, 2012a; Federmeier et al., 2010). In addition, Wlotko et al. (2012) found that the older adults did not show a late pre-frontal positivity in response to unpredictable words in high-Cloze contexts. This effect did occur for younger adults in Federmeier et al. (2007) and has been interpreted as evidence of effects of a disconfirmed prediction (see Section 4.4).

Similarly, it is clear that young children predict syntax. Using a method similar to Arai et al. (2007), Thothatiri and Snedeker (2008) found that 3- and 4-year-olds predict that upcoming referents will be mentioned in the order implied by the syntactic structure (PO or DO) they have just comprehended (i.e., therefore showing effects of both syntactic priming and syntactic prediction). In addition, Lukyanenko and Fisher (2016) showed that 3- and, to a lesser extent, 2.5-year-olds predictively look at a picture of multiple objects when they hear *Where are the* ... They argued that the children used the syntactic number of the verb to predict the number of an upcoming subject noun; thus, they use a syntactic relation (i.e., syntactic agreement) to guide their predictions (see also Melançon & Shi, 2015). Finally, Gambi, Pickering, and Rabagliati (2016) used a method similar to Kukona et al. (2011) to show that 3-to-5-year olds predict entities that are both semantically associated and syntactically predictable, but not merely semantically associated (e.g., looking at a robber but not a policeman after *Pingu will arrest the*…). These findings are consistent with prediction-by-production but not prediction-by-association.

There is little evidence about whether young children predict form. Mahr et al. (2015) showed that 2-year-olds look more quickly to a referent when its name is preceded by a determiner carrying informative coarticulation cues. These results are similar to those of Salverda et al. (2014) for adults (see section 2.4.3) and, like those findings, do not demonstrate that they use the determiner to predict the noun. Children's looks to the referent are overall much slower than adults' and so an integration explanation is especially likely for them.

Evidence from prediction in different population supports the optionality of prediction. (3.6.3) Overall, predictive abilities develop early and can be present in non-native speakers and older adults. However, prediction appears to be less pronounced in such populations than in native-speaking young adults – people who are fast and skilled language users with extensive processing resources – and in fact may not always occur. Given that such groups can comprehend language, these findings suggest that prediction is an aid to comprehension rather than a necessary component of it.

We have proposed (sections 2.4-2.6) that optionality characterizes prediction-by-production in particular, but not prediction-by-association, because the latter is an integral component of every act of comprehension and is largely resource-free. This proposal means that prediction-by-association, unlike prediction-by-production, should be unimpaired in comprehenders with limited resources, but we know of no study that has tested this directly. It also means that predictions that correspond to later stages of production (and particularly predictions of form based on predictions of semantics, such as in De Long et al., 2005) should be more impaired in comprehenders with limited resources, because such predictions should require more time-consuming and resource-intensive computations. At present, the evidence from non-native speakers supports this claim (e.g., Martin et al., 2013), but there is insufficient evidence from other populations to be able to generalize this conclusion.

Discussion (4.)

As we have shown, there is overwhelming evidence that prediction is widespread in language comprehension. Studies using electrophysiology, eye movements, and reaction times demonstrate that it occurs when utterances are encountered in isolation or in non-linguistic contexts, in monologue and dialogue, and in reading and listening. It also occurs at different linguistic levels, from semantics to syntax to form. The conclusion that prediction is widespread holds even if we only consider findings for which an integration explanation is not possible (as we have done in Section 2), but is further reinforced by the systematic review we conducted in Section 3.

Our review provides strong and converging evidence that the most effective means of prediction during comprehension utilizes the system that is used to produce utterances, a system that is both sophisticated and already available to the comprehender. As illustrated in Figure 2, the comprehender derives the intention that would be used to drive production, using a combination of covert imitation (of what the speaker has said so far) and the non-linguistic context. They use the intention to start the process of production, and the output of the production process constitutes the predictions that they make. This process takes place as the comprehender encounters every new word in the utterance (see Fig. 1).

The two strongest forms of evidence for prediction-by-production come from electrophysiology and from the visual-world paradigm, and they are largely complementary to each other. Electrophysiological studies demonstrate that comprehenders predict levels of representation that are computed later in production on the basis of levels of representation that are computed earlier (e.g., grammatical gender from semantics; Wicha et al., 2004), and that such predictions require the involvement of the production system (Martin et al., 2018). Visual-world studies demonstrate that comprehenders use covert imitation (e.g., Kamide et al., 2003) and derive the speaker's intention (Chambers & San Juan, 2008). In addition, studies of speech provide evidence for motor activation during prediction (e.g., Drake & Corley, 2015) and the evidence for turn-end prediction (e.g., Magyari et al., 2014) and early preparation of responses (Bögels et al., 2015) in dialogue strongly suggests that prediction-by-production benefits the smooth and rapid exchange of turns we observe in conversation.

The When and How of Prediction-by-Production (4.1)

Prediction-by-production is widespread but optional. In Figure 2, we distinguished obligatory processes (comprehension and covert imitation) that lead to deriving the speaker's intention from optional processes that generate predictions by running the intention through the production system. In Figure 1, we indicated that the obligatory processes take place continuously. The key open question is what factors determine whether and when the optional processes take place. But this question has not been the focus of research on prediction and the main conclusion we can make at this point is that the optional processes depend on time and resources.

We have identified many cases in which there is no evidence that prediction-by-production occurs. But since competent users of a language almost always manage to comprehend what they encounter, they must be able to do so without using prediction-by-production. We have therefore proposed an account that combines traditional (non-predictive) mechanisms of incremental interpretation (e.g., Marslen-Wilson, 1973) with an optional mechanism that recruits the production system for prediction.

Good evidence that prediction-by-production requires time comes from Ito et al. (2016) and Wlotko and Federmeier (2015), which we discussed in section 3.1.2. In Ito et al. (2016), the N400 effect indexing semantic prediction occurred at presentation rates (SOAs) of 500ms and 700ms, but the N400 effect indexing phonological prediction occurred only at the slower presentation rate. Thus, slower presentation enhances prediction and specifically those aspects of prediction that correspond to the later stages of production. In Wlotko and Federmeier (2015), the N400 effect indexing semantic prediction was itself reduced at the faster presentation rate of 250ms, showing that even semantic prediction takes time.

There is some evidence that prediction-by-production is resource-intensive (section 2.4.2), with predictive eye movements being sensitive to working-memory limitations (Huettig & Janse, 2016) and memory load (Ito et al., 2017). An additional way to investigate this issue is to consider the effects of adverse listening or reading conditions on prediction. We might expect prediction-by-production to be used less under adverse conditions, because the comprehenders' limited resources are more taxed. But it is also possible that prediction-by-production might be engaged more, because it is needed more (and indeed, this possibility would be consistent with enhanced motor activation while listening in adverse conditions; Adank, 2012). As far as we are aware, no research has directly addressed this issue.

We propose that comprehenders predict by production whenever some aspect of the upcoming utterance is predictable, but only if time and resources are available. For example, the comprehender hearing *The boy went out to the park to fly a kite* (Fig. 1) would initiate predictions at several points in the sentence. Assuming sufficient resources, she would predict an event and a verb after *The boy,* but would not predict form. After *The boy went out to the*, she would predict the semantic category of OPEN-SPACE. After *The boy went out to the park to fly a*, she would predict the word *kite*, which includes all of its lexical information (e.g., +FLYABLE, noun, /kaIt/). Resource and time limitations may prevent some of these predictions by effectively stopping the production system, for example allowing her to predict an event but not a verb after *boy* or predict +FLYABLE and noun but not /kaIt/ after *fly a*. Overall, we expect predictions of earlier stages in production to occur more often than predictions of later stages (as suggested by Ito et al., 2016).

In addition, some predictions may be ready "early," well before the (potentially) corresponding input occurs. For example, the comprehender may predict that the sentence will mention *kite* several words before it might occur (e.g., around *park*). Studies have not typically addressed this issue, though we noted that Ito et al. (2018) found that listeners looked at pictures that were phonologically related to the predictable word from 500ms before the word onset. Moreover, evidence that comprehenders predict turn-ends several hundred milliseconds before they occur (Magyari et al., 2014) also supports early prediction. More speculatively, the evidence that addressees prepare answers well before the end of a question and then produce the answer at the appropriate time (Bögels et al., 2015) suggests extensive overlap between comprehension and production in dialogue and is compatible with early prediction-by-production.

Two Components to Prediction-by-Production? (4.2)

Throughout this review, we have assumed that prediction-by-production makes use of production mechanisms that are traditionally assumed in psycholinguistics (e.g., Bock & Levelt, 1994; Dell, 1986; Levelt, 1989). But some theories assume that speakers predict aspects of what they are likely to say before they prepare the representations that underlie the act of speaking (Hickok, 2012; Pickering & Garrod, 2013; Tourville & Guenther, 2011). These theories therefore distinguish between *predicted* and *implemented* representations within production itself (with the implemented representations being those that are traditionally assumed in psycholinguistics). If such theories are correct, then prediction-by-production might make use of the predicted representations as well as the implemented representations. Below we discuss the characteristics of these predicted representations and what role they might serve in prediction-by-production.

Speakers predict what a sound or syllable they are about to articulate will sound like. This allows them to spot deviations from the predicted sound extremely quickly, because all they have to do is compare what it did sound like to what it should have sounded like (according to their prediction). If they match, the earliest auditory response (roughly 100ms) in the EEG and MEG records is reduced (e.g., Heinks-Maldonado, Nagarajan, & Houde, 2006; see also Niziolek, Nagarajan, & Houde, 2013).

According to computational models of speech motor control based on this evidence (Hickok, 2012; see also Tourville & Guenther, 2011), the speaker takes an efference copy of his intention to speak, runs it through a forward model of syllabic production, and rapidly computes the predicted percept of the syllable (i.e., what it sounds like to the speaker), before the syllable is implemented by the articulation system. A forward model is thus a mapping between the (motor) intention to move the articulators in a certain way and the perceptual outcome of actually moving them. Speakers learn forward models by repeatedly performing an action and hence pairing the intention with the percept of the outcome, and the discrepancy between the predicted and actual percept drives error-based learning (Wolpert, 1997). Pickering and Garrod (2013) proposed that forward models are not only involved in the production of speech sounds, but can be computed in relation to every stage of the process of speaking. They directly map production-based representations onto comprehension-based representations and could therefore serve as extremely flexible and fast prediction-by-production mechanisms that speed up the process of comparing predictions to the input by generating predictions that are already in a format comparable to the input.

We have discussed three groups of studies that provide some evidence for the use of forward models in prediction. The first is Dikker et al. (2010), who found that the M100 response generated in visual cortex was enhanced when the visual form of a word was atypical for a word of a predictable syntactic category (section 2.1.3). In other words, the syntactic context led to a prediction of visual form that was compared to the perceived visual form. This requires a mechanism to convert an abstract (non-sensory) prediction into a sensory format. If comprehenders predict by production, they need to map from a production-based to a comprehension-based representation. As illustrated above, this is precisely the function that forward models serve in language production, and it may be that they serve the same function during prediction in language comprehension too.

Second, we noted arguments that readers plan eye-movements before lexical access is completed, but that these plans are nevertheless affected by word frequency (Reichle et al., 1998; Engbert et al., 2005; White, 2008), which suggests that readers predict word frequency (section 3.2.2). Such predictions appear to require forward modeling, as they do not depend on using context to predict a word and hence its frequency (as shown by Fruchter et al., 2015; section 3.1.1), but rather depend on the familiarity of the target word itself. So the prediction must involve a mapping – a forward model – between the target word (specifically, its familiarity as assessed in the L1 stage) and lexical access time. In other words, the reader processes a word form and predicts (based on experience with that form) how long lexical access will take – and then begins saccade planning before lexical access.

Finally, there is much evidence that the cerebellum computes forward models in motor control (i.e., predicting the sensory consequences of movements, including speech movements; Tourville & Guenther, 2011). We have noted that inhibiting the cerebellum disrupts prediction-by-production (Lesage et al., 2012; section 2.2). Additionally, activation in the right cerebellum correlates with adaptation to distorted speech (Guediche, Holt, Laurent, Lim, & Fiez, 2015), which suggests that it plays a role in facilitating comprehension. In fact, many authors have proposed that the cerebellum computes forward models that support prediction during comprehension (see Moberget & Ivry, 2016).

These arguments do not prove that comprehenders use forward models during prediction-by-production. But given the strong evidence for forward models in production and the evidence that the other components of production are implicated in prediction, it would be worthwhile to directly investigate forward modeling in comprehension.

What Role for Prediction-by-Association? (4.3)

In Section 2.5, we argued that some cases of prediction are due to prediction-by-association. In Figure 2, prediction-by-association is treated as an integral (non-optional) component of comprehension – it takes place whether or not the comprehender goes on to predict by production. Much of the evidence for prediction-by-association comes from traditional priming studies, such as associative priming (D. E. Meyer & Schvaneveldt, 1971; Bentin et al., 1985). The spreading-activation account of such studies involves prediction (e.g., Lau et al., 2013), though, as we noted, a non-predictive (integration-based) account may be possible. Additionally, we discussed a few studies in which comprehenders appear to predict-by-association and where a prediction-by-production account is not possible. For example, Kukona et al. (2011) found that comprehenders look at a policeman after hearing *Bill will arrest* – the word *policeman* is associatively related to *arrest*, but is incompatible with covert imitation at this point and thus the comprehender would not predict it by production (see also Methusalem et al., 2012; Kukona et al., 2014; Sauppe, 2016; Kamide et al., 2003; Borovsky et al., 2012).

These examples all involve spreading activation between representations linked in long-term memory. The key question for understanding prediction-by-association is determining the content of these representations and the nature (i.e., number and strength) of the links between representations, along which activation spreads. This amounts to a theory of the organization of semantic memory, and is beyond the scope of this paper. With regard to the content of the representations between which activation spreads, we propose that the starting point for prediction-by-association is the comprehension representations that also feed into the process of covert imitation (Figure 2). Such representations need not be limited to lexical semantics, and in fact our explanation of Methusalem et al. (2012) explicitly assumed that prediction-by-association can have more complex event representations as its starting point. However, we suggest the starting point for prediction-by-association is unlikely to incorporate the non-linguistic context and shared background knowledge, which are instead part of the process of deriving the intention, and thus constitute the initial stages of prediction-by-production.

Prediction-by-association is of course dependent on experience: For example, regularly encountering *queen* and *king* (and their referents) in similar contexts creates an association between them (e.g., Andrews, Vigliocco, & Vinson, 2009), and the strength of this learned association affects prediction. Future studies should investigate how our experience of language and the world shapes our prediction. A recent example of this line of research is Borovsky (2016), who investigated how much experience of novel events (e.g., *the monkey is riding the bus*) is necessary before comprehenders begin to generate predictions based on combining the meaning of agent and the verb (similarly to Borovsky et al., 2014, see section 3.6.2): Findings from two visual-world studies showed that people look at predictable patients before they encounter them, but only after being exposed to the novel events (with the same agents) more than once.

In addition, a large literature addresses the question of how people learn regularities in the order of syntactic categories (e.g., that nouns tend to follow determiners in English) and sounds (so called phonotactic constraints; e.g., in Italian /st/ must be followed by either a vowel or /r/). Infants can acquire such patterns easily (e.g., Mintz, 2003; Saffran, 2003; Mattys & Jusczyk, 2001), and adults and children learn new (artificially created) patterns by simple exposure (e.g., Gómez & Gerken, 2000; Saffran, 2001; Misyak, Christiansen, & Bruce Tomblin, 2010; Pozzan & Trueswell, 2015), including patterns that involve nonadjacent dependencies (e.g., Newport & Aslin, 2004; Gómez, 2002; Misyak & Christiansen, 2007). The nature of such (statistical) learning mechanisms is not well understood, but they are likely not specific to language (e.g., Kirkham, Slemmer, & Johnson, 2002; Thiessen, 2011; Conway, Bauernschmidt, Huang, & Pisoni, 2010). People may use these learned regularities to make predictions, in which case they would be using prediction-by-association. However, the extent to which these regularities drive prediction is not clear (Dale, Duran, & Morehead, 2012). In sum, prediction-by-association potentially plays a ubiquitous role during comprehension. But as we have noted in section 1.5, its role in supporting comprehension is limited because it is also undirected and short-lived.

Effects of Disconfirming Predictions (4.4)

What happens when a prediction is incorrect? This is an important question, but we have not made it a focus of our paper because the answer is still unclear. Here we briefly discuss the little available evidence. In many areas of language comprehension (e.g., garden-paths; Bever, 1970), misanalysis leads to difficulty, and we might similarly expect difficulty following an incorrect prediction. But there is very little evidence that making a wrong prediction causes comprehenders to read more slowly or make more regressive eye-movements, compared to making no prediction at all. In particular, words do not appear to be harder to process when they follow a context that is strongly predictive of a different word versus a context that is not strongly predictive of a different word. For example, Luke and Christianson (2016) found no evidence of prediction costs in a large-scale reading study. They considered all words with Cloze values below 50% and found that their difficulty did not depend on the Cloze value of the most predictable alternative word given that context. Frisson, Harvey, and Staub (2017) found similar reading times for an unpredictable word in a context that predicted a different word versus a context that predicted no specific word (though reading times for the predicted word itself were of course shorter; see Staub, 2015). Finally, Traxler and Foss (2000) found that a predictive context had the same effect on target word naming whether the context also predicted another word to a similar extent or did not. All of these findings suggest that recovering from an incorrect prediction does not tend to cause difficulty.

Moreover, it is possible that comprehending a word should actually be easier rather than harder when it follows a context that strongly predicts a different word versus a context that does not strongly predict a different word. In a timed Cloze task, Staub et al. (2015) found that participants were faster producing a completion with a given Cloze probability in a higher-constraint context than a completion with the same Cloze probability in a lower-constraint context, and showed how this result was compatible with a race-based model in which the activation levels of alternative completions are independent. An equivalent model for prediction during comprehension would therefore actually claim an advantage for a failed prediction rather than a cost.

Note that ERP studies demonstrate effects of having a more predictable alternative word. For example, Federmeier et al. (2007) found that unpredictable words evoked a larger anterior positivity following a context that strongly predicted a different word versus a context that did not strongly predict any word (see Van Petten & Luka, 2012). Brothers, Swaab, and Traxler (2015) also showed a larger anterior positivity when participants reported not having predicted the target word than when they reported having predicted it. Interestingly, this positivity was elicited by unpredicted targets in sentences with comparatively low Cloze values.

In sum, there are effects associated with making the wrong prediction, but they may not reflect costs. Note that most researchers (e.g., Luke & Christianson, 2016; Van Petten & Luka, 2012) assume that wrong predictions should lead to costs only when such predictions are detailed enough to pre-activate a specific lexical item, which would then compete with activation of alternative words. Instead, if comprehenders predict sets of words, then these researchers assume that costs should be much less likely, perhaps because more words would be compatible with what is predicted. In addition, Staub et al. (2015) found that completions were produced faster when they had a closer semantic relationship to the most likely completion (see also Roland, Yun, Koenig, & Mauner, 2012), and this facilitatory effect of semantic relatedness was larger after more predictive sentence contexts (though cf. Kleinman, Runnqvist, & Ferreira, 2015).

Staub et al.'s (2015) conclusions are compatible with prediction-by-production. At the semantic level, comprehenders typically pre-activate a large set of related concepts in parallel, and this broad activation might explain the facilitatory effect of semantic relatedness. But as the comprehender proceeds further through the stages of prediction-by-production, pre-activation becomes increasingly focussed on fewer and fewer alternatives (and ultimately a single alternative). This may result in alternatives losing activation and therefore could cause a processing load if one of these abandoned alternatives ends up being the one that is actually encountered by the comprehender. In sum, the staged nature of prediction-by-production means that there might be both benefits and costs to disconfirmed predictions.

Overall, more detailed research is needed to determine the effects of disconfirmed predictions. At the moment, most evidence seems to suggest that are no costs associated with disconfirmed predictions (though the ERP studies show that the brain registers when a prediction is disconfirmed). This is perhaps surprising but it can be argued to motivate prediction. The benefit of successful prediction is that pre-activation of representations facilitates subsequent bottom-up processing (see section 1.2). If unsuccessful predictions are not costly, then comprehenders who can predict may as well do so.

Methodological Implications (4.5)

Most research suggests that Cloze tests accurately measure predictability and their results are closely related to actual prediction. But why should this be the case? In Cloze tests, participants engage their production systems to complete the sentence fragments. In other words, they simply engage the same processes used in prediction-by-production but actually produce the predicted word. In fact, Staub et al. (2015) found that participants produced high-Cloze words more quickly than low-Cloze words (both when comparing higher- and lower-constraint contexts and when comparing completions to the same context), a finding which is consistent with prediction-by-production.

Researchers often raise concerns that experiments that involve slow presentation rates may not reflect “normal” processing. Our proposals give a clear basis for this intuition. The stages involved in prediction-by-production take a similar amount of time to the equivalent stages in language production (e.g., Indefrey & Levelt, 2004). So a slow presentation rate (e.g., 700ms/word in an ERP study; Ito et al., 2016) or slow speech (e.g., Altmann & Kamide, 1999) allows comprehenders to engage their production systems extensively in making predictions. Comparable processes may not occur in skilled reading or everyday speech comprehension, as they proceed at a faster rate. The results from comprehension at slow presentation rates may thus not be representative of other forms of comprehension, and differences will be particularly apparent in late stages of production, such as phonology, as the comprehender may not have time to make the relevant predictions.

Conclusions (5.)

Comprehenders regularly predict different aspects of what they are likely to encounter – specific words, aspects of meaning, grammar, and sound. To do this, they use general-purpose associative mechanisms, which are ubiquitous but not usually very effective. But by far the most important route to prediction involves the production system, so that comprehenders predict using the mechanisms that they would use if they took over the role of speaker at this point – as they do in natural dialogue and in Cloze tasks. Prediction-by-production is highly accurate and effective (unlike prediction-by-association) but does not occur all the time and is not necessary for successful comprehension. Instead, it is a very important but optional mechanism that helps comprehenders achieve their goals of rapid and robust understanding of speeches, texts, and conversations.

References

Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood (ALE) meta-analyses. *Brain and Language, 122*(1), 42-54.

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science, 21*(12), 1903-1909.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419-439.

Allport, D., & Funnell, E. (1981). Components of the mental lexicon. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 295*(1077), 397-410.

Altmann, G. T. (1999). Thematic role assignment in context. *Journal of Memory and Language, 41*(1), 124-145.

Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: The ‘blank screen paradigm’. *Cognition, 93*(2), B79-B87.

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

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518.

Altmann, G. T., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111*(1), 55-71.

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583-609.

Amsel, B. D., DeLong, K. A., & Kutas, M. (2015). Close, but no garlic: Perceptuomotor and event knowledge activation during language comprehension. *Journal of Memory and Language, 82*, 118-132.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review, 116*(3), 463.

Arai, M., Nakamura, C., & Mazuka, R. (2015). Predicting the unbeaten path through syntactic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 482-500.

Arai, M., Van Gompel, R. P., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology, 54*(3), 218-250.

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If You Say Thee Uh You Are Describing Something Hard: The On-Line Attribution of Disfluency during Reference Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(5), 914-930.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*(3), 364-390.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109(1),* 18-40.

Bendixen, A., Scharinger, M., Strauß, A., & Obleser, J. (2014). Prediction in the service of comprehension: Modulated early brain responses to omitted speech segments. *Cortex, 53*, 9-26.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and clinical Neurophysiology, 60*(4), 343-355.
Bever, T. G. (1970). The cognitive basis for linguistic structures. In R. Hayes (Ed.), *Cognition and language development* (pp. 279-362). New York: Wiley & Sons.
Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition, 42*(1), 23-60.
Bobb, S. C., Huettig, F., & Mani, N. (2016). Predicting visual information during sentence processing: Toddlers activate an object's shape before it is mentioned. *Journal of Experimental Child Psychology, 151*, 51-64.
Bock, K. (1986a). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(4), 575.
Bock, K. (1986b). Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355-387.
Bock, K., & Levelt, W. J. M. (1994). Language Production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945-984). San Diego: Academic Press.
Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics, 52,* 46-57.
Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General, 143*(1), 295.
Boland, J. E. (2005). Visual arguments. *Cognition, 95*(3), 237-274.
Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*(6), 774.
Borovsky, A. (2016). The amount and structure of prior event experience affects anticipatory sentence interpretation. *Language, Cognition and Neuroscience, 32*(2), 190-204.
Borovsky, A., & Creel, S. C. (2014). Children and adults integrate talker and verb information in online processing. *Developmental psychology, 50*(5), 1600-1613.
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.
Borovsky, A., Sweeney, K., Elman, J. L., & Fernald, A. (2014). Real-time interpretation of novel events across childhood. *Journal of Memory and Language, 73,* 1-14.
Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America, 103*(5), 2670-2676.
Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language, 75*, 104-116.
Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339-2349.
Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition, 136*, 135-149.
Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language, 93,* 203-216.

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language, 61(2),* 171-190.
Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1430-1450.
Carminati, M. N., van Gompel, R. P., Scheepers, C., & Arai, M. (2008). Syntactic Priming in Comprehension: The Role of Argument Order and Animacy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1098-1110.
Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences, 18(2),* 90-98.
Chambers, C. G., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition, 108*(1), 26-50.
Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234.
Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language, 52*(1), 144-169.
Chow, W.-Y., Smith, C., Lau, E., & Phillips, C. (2016). A "bag-of-arguments" mechanism for initial verb predictions. *Language, Cognition and Neuroscience, 31*(5), 577-596.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(2), 181-204.
Clark, H. H. (1996). *Using language*. Cambridge University Press: Cambridge, U.K..

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*(3), 472-517.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.
Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*(3), 256-266.
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.
D'Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I., & Craighero, L. (2011). Tongue corticospinal modulation during attended verbal stimuli: Priming and coarticulation effects. *Neuropsychologia, 49*(13), 3670-3676.
Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*(4), 317-367.
Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5-6), 507-534.
Dale, R., Duran, N. D., & Morehead, J. R. (2012). Prediction during statistical learning, and implications for the implicit/explicit divide. *Advances in Cognitive Psychology, 8*(2), 196-209.
Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top‐down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience, 23*(12), 3914-3932.
De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82*(3), 515-535.

Dell, G. S. (1986). A spreading-activaction theory of retrieval in sentence production. *Psychological Review, 93*(3), 283-321. doi: http://dx.doi.org/10.1037/0033-295X.93.3.283
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. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language, 121*(3), 226-239.
DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.
Dikker, S., & Pylkkanen, L. (2011). Before the N400: Effects of lexical‑semantic violations in visual cortex. *Brain and Language, 118*(1), 23-28.
Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language, 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. *Psychological Science, 21*(5), 629-634.
Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition, 110*(3), 293-321.
Dikker, S., Silbert, L. J., Hasson, U., & Zevin, J. D. (2014). On the same wavelength: predictable language enhances speaker‑listener brain-to-brain synchrony in posterior superior temporal gyrus. *Journal of Neuroscience, 34*(18), 6267-6272.
Dilley, L., & Pitt, M. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*(11), 1664-1670. doi:10.1177/0956797610384743
Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 118*(2), 191-211.
Drake, E., & Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition, 43*(8), 1136-1147.
Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal Of Experimental Psychology. Human Perception And Performance, 31*(5), 954-959.
Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology, 23*(2), 283-292.
Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior, 20*(6), 641-655.
Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179-211.
Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27*, 143-165.
Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation during Reading. *Psychological Review, 112*(4), 777-813.
Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of neuroscience, 15*(2), 399-402.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*(32), 12203-12208.
Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (2011). Phonological Typicality Influences Sentence Processing in Predictive Contexts: Reply to Staub, Grant, Clifton, and Rayner (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1318-1325.
Farmer, T. A., Yan, S., Bicknell, K., & Tanenhaus, M. K. (2015). Form-to-expectation matching effects on first-pass eye movement measures during reading. *Journal of Experimental Psychology: Human Perception and Performance, 41*(4), 958.
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. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.
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.
Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133-146.
Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146,* 75-84.
Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210-233.
Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language, 35(5),* 724-755.
Forster, K. I. (1979). Levels of processing and the structure of the language processor *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (Vol. 27, pp. 27-85). Hillsdale, NJ: Erlbaum.
Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1461.
Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition, 19*(01), 213-219.
Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language, 140,* 1-11.
Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance 12: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Lawrence Erlbaum.
Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178-210.
Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences, 16*(5), 262-268.
Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language, 95,* 200-214.
Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360(1456),* 815-836.

Fruchter, J., Linzen, T., Westerlund, M., & Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. *Journal of Cognitive Neuroscience, 27*(10), 1912-1935.
Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology, 22*(7), 615-621.
Gambi, C., & Pickering, M. J. (2017). Models linking production and comprehension. In E. Fernàndez, and H. Cairns (Eds.), *Handbook of Psycholinguistics*. (pp 157-182). Hoboken, N.J.: Wiley-Blackwell.
Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond Associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, *157*, 340-351.
Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110-125.
Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.
Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology, 6*, 751.
Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1-76.
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.
Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes*, *8*(2), 147-161.
Goldrick, M. A., Ferreira, V., & Miozzo, M. (2014). *The Oxford handbook of language production*. Oxford: Oxford University Press.
Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.
Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*(5), 178-186.
Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. *Papers in laboratory phonology, 9*, 173-200.
Gow, D. W., & Olson, B. B. (2015). Lexical mediation of phonotactic frequency effects on spoken word recognition: A Granger causality analysis of MRI-constrained MEG/EEG data. *Journal of Memory and Language, 82*, 41-55.
Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *Journal of Neuroscience, 37(18),* 4848-4858.
Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., & Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain Research, 1361*, 54-66.
Guediche, S., Holt, L. L., Laurent, P., Lim, S.-J., & Fiez, J. A. (2015). Evidence for cerebellar contributions to adaptive plasticity in speech perception. *Cerebral Cortex*, *25*(7), 1867-1877.
Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. *Journal of Cognitive Neuroscience, 25*(5), 706-718.
Hale, J. (2001). *A probabilistic Earley parser as a psycholinguistic model.* Paper presented at the Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49(1),* 43-61.

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport, 17*(13), 1375.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics, 38*(4), 555-568.

Heller, D., Arnold, J. E., Klein, N., & Tanenhaus, M. K. (2015). Inferring difficulty: Flexibility in the real-time processing of disfluency. *Language and Speech, 58*(2), 190-203.

Herrmann, B., Maess, B., Hasting, A. S., & Friederici, A. D. (2009). Localization of the syntactic mismatch negativity in the temporal cortex: an MEG study. *NeuroImage, 48*(3), 590-600.

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

Hintz, F., & Meyer, A. S. (2015). Prediction and production of simple mathematical equations: Evidence from visual world eye-tracking. *PLoS ONE, 10*(7), e0130766.

Hintz, F., Meyer, A. S., & Huettig, F. (2016). Encouraging prediction during production facilitates subsequent comprehension: Evidence from interleaved object naming in sentence context and sentence reading. *The Quarterly Journal of Experimental Psychology, 69*(6), 1056-1063.

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye-movements in the visual world. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 43(9),* 1352-1374,

Hirose, Y., & Mazuka, R. (2015). Predictive processing of novel compounds: Evidence from Japanese. *Cognition, 136*, 350-358.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of experimental psychology. Learning, memory, and cognition, 40*(1), 166-190.

Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German Sign Language. *Neuropsychologia, 51*(11), 2224-2237.

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

Huettig, F., & Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia─evidence from eye‐tracking. *Dyslexia, 21*(2), 97-122.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80-93.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience, 31(1),* 19-31.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology, 2*, 255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144. doi:doi:10.1016/j.cognition.2002.06.001

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*(6), 431-439.

Ito, A., Corley, M., & Pickering, M. J. (2017). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, doi:10.1017/S1366728917000050

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

Ito, A., Martin, A. E., & Nieuwland, M. S. (2016). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, 1-12.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1-11.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 824-843.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354.

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

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

Kempen, G. (2014). Prolegomena to a neurocomputational architecture for human grammatical encoding and decoding. *Neuroinformatics, 12*(1), 111-142.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11(1),* 32-38.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science, 7(2),* 46-49.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89(1)*, 25-41.

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience, 24*(5), 1104-1112.

Kim, A. E., & Gilley, P. M. (2013). Neural mechanisms of rapid sensitivity to syntactic anomaly. *Frontiers in Psychology, 4,* 45.

Kim, C. S., Gunlogson, C., Tanenhaus, M. K., & Runner, J. T. (2015). Context-driven expectations about focus alternatives. *Cognition*, *139*, 28-49.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), B35-B42.

Kleinman, D., Runnqvist, E., & Ferreira, V. S. (2015). Single-word predictions of upcoming language during comprehension: evidence from the cumulative semantic interference task. *Cognitive Psychology, 79*, 68-101.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30*(3), 481-529.

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language, 57*(4), 519-543.

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

Kretzschmar, F., Schlesewsky, M., & Staub, A. (2015). Dissociating word frequency and predictability effects in reading: Evidence from coregistration of eye movements and Eeg. *Journal of experimental psychology. Learning, memory, and cognition, 41*(6), 1648-1662.

Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 326-347.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition, 119*(1), 23-42.

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

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition, 133*(2), 335-342.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Kutas, M., Hillyard, S. A., & Gazzaniga, M. S. (1988). Processing of semantic anomaly by right and left hemispheres of commissurotomy patients: Evidence from event-related brain potentials. *Brain, 111(3),* 553-576.

Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in Chinese: Evidence from ERPs. *Cognition, 166*, 433-446.

Lammertink, I., Casillas, M., Benders, T., Post, B., & Fikkert, P. (2015). Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology, 6*.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language, 61*(3), 326-338.

Lau, E., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*(3), 484-502.

Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920-933.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98*(1), 74-88.

Leonard, M. K., Baud, M. O., Sjerps, M. J., & Chang, E. F. (2016). Perceptual restoration of masked speech in human cortex. *Nature Communications, 7,* doi: 10.1038/ncomms13619

Lerner, G. H. (2002). Turn-sharing: The choral co-production of talk-in-interaction. In C. E. Ford, B. A. Fox, Thompson, S. A. (Eds). *The language of turn and sequence*, pp. 225-256. Oxford University Press: New York.

Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S., & Miall, R. C. (2012). Cerebellar rTMS disrupts predictive language processing. *Current Biology, 22*(18), R794-R795.
Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*(1), 41-104.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-75.
Levinson, S. C. (2016). Turn-taking in human communication⁻origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6-14.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177.
Linzen, T., & Jaeger, T. F. (2016). Uncertainty and Expectation in Sentence Processing: Evidence From Subcategorization Distributions. *Cognitive Science, 40*(6), 1382-1411.
Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(9), 1400-1416.
Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88*, 22-60.
Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition, 146*, 349-370. doi:http://dx.doi.org/10.1016/j.cognition.2015.10.012
MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution [corrected]. *Psychological Review, 101*(4), 676-703.
Mack, J. E., Ji, W., & Thompson, C. K. (2013). Effects of verb meaning on lexical integration in agrammatic aphasia: Evidence from eyetracking. *Journal of neurolinguistics, 26*(6), 619-636.
Maess, B., Mamashli, F., Obleser, J., Helle, L., & Friederici, A. D. (2016). Prediction signatures in the brain: semantic pre-activation during language comprehension. *Frontiers in Human Neuroscience*, doi:10.3389/fnhum.2016.00591
Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science, 27*(2), 285-298.
Magyari, L., Bastiaansen, M., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26*(11), 2530-2539.
Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology, 3*. doi:10.3389/fpsyg.2012.00376
Mahr, T., McMillan, B. T., Saffran, J. R., Weismer, S. E., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition, 142*, 345-350.
Mani, N., Daum, M. M., & Huettig, F. (2016). "Pro-active" in many ways: Developmental evidence for a dynamic pluralistic approach to prediction. *Quarterly Journal of Experimental Psychology, 69*(11), 2189-2201.
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.
Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America, 69*(2), 548-558.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244*(5417), 522-523.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71-102.

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific reports, 8(1),* 1079.

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.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953-978.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*(2), 91-121.

Mattys, S. L., Pleydell-Pearce, C. W., Melhorn, J. F., & Whitecross, S. E. (2005). Detecting silent pauses in speech. A new tool for measuring on-line lexical and semantic processing. *Psychological Science, 16*(12), 958-964.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10*(8), 363-369.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*: Psychology Press.

McQueen, J. M., Cutler, A., & Norris, D. (2003). Flow of information in the spoken word recognition system. *Speech Communication, 41*(1), 257-270.

McQueen, J. M., Eisner, F., & Norris, D. (2016). When brain regions talk to each other during speech processing, what are they talking about? Commentary on Gow and Olson (2015). *Language, Cognition and Neuroscience, 31(7),* 860-863.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language, 61*(1), 1-18.

McQueen, J. M., Norris, D., & Cutler, A. (2006). Are there really interactive processes in speech perception? *Trends in Cognitive Sciences, 10*(12), 533-533.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition, 33*(7), 1174-1184.

Melançon, A., & Shi, R. (2015). Representations of abstract grammatical feature agreement in young children. *Journal of child language, 42*(6), 1379-1393.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science, 22*, 1173–1182.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language, 66*(4), 545-567.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language, 35*(4), 477-496.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227-234.

Miall, R., Antony, J., Goldsmith-Sumner, A., Harding, S., McGovern, C., & Winter, J. (2016). Modulation of linguistic prediction by TDCS of the right lateral cerebellum. *Neuropsychologia, 86*, 103-109.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of test materials. *Journal of Experimental Psychology, 41*(5), 329-335.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117.

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), 1-10.

Misyak, J. B., & Christiansen, M. H. (2007). *Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language.* Paper presented at the Proceedings of the 29th Annual Cognitive Science Society Conference, Austin, TX.

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.

Mitsugi, S., & MacWhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition, 19*(01), 19-35.

Moberget, T., Gullesen, E. H., Andersson, S., Ivry, R. B., & Endestad, T. (2014). Generalized role for the cerebellum in encoding internal models: evidence from semantic processing. *Journal of Neuroscience, 34*(8), 2871-2878.

Moberget, T., & Ivry, R. B. (2016). Cerebellar contributions to motor control and language comprehension: searching for common computational principles. *Annals of the New York Academy of Sciences, 1369(1)*, 154-171.

Nakatani, K., & Gibson, E. (2010). An On ‐ Line Study of Japanese Nesting Complexity. *Cognitive Science, 34*(1), 94-112.

Nation, K., Marshall, C. M., & Altmann, G. T. (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.

Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In A. M. Suprenant (Ed.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 69-93). Washington, DC, US: American Psychological Association.

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language, 25(1)*, 75-92.

Newman, R. L., & Connolly, J. F. (2004). Determining the role of phonology in silent reading using event-related brain potentials. *Cognitive Brain Research, 21*(1), 94-105.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., . . . Huettig, F. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *bioRxiv*. doi:10.1101/111807

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience, 33*(41), 16110-16116.

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357-395.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(03), 299-325.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204-238.

Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., & Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex. *NeuroImage, 128*, 218-226.

Obleser, J., & Kotz, S. A. (2010). Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network. *Cerebral Cortex, 20*(3), 633-640.

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC neuroscience, 8*(1), 89.

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

Otten, M., & Van Berkum, J. J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research, 1291*, 92-101.

Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology, 25*(12), 1649-1653.

Perea, M., & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition, 62*(2), 223-240.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 539.

Pickering, M. J., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6*(3), 229-259.

Pickering, M. J., & Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends in Cognitive Sciences, 18*(9), 451-456.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*(3), 427-459.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-226.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329-392.

Pickering, M. J., & Traxler, M. J. (2001). Strategies for processing unbounded dependencies: lexical information and verb-argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1401-1410.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*(3), 347-370.

Pozzan, L., & Trueswell, J. C. (2015). Revise and resubmit: How real-time parsing limitations influence grammar acquisition. *Cognitive Psychology, 80*, 73-108.

Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research, 1*(2), 73-86.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2006). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience, 17(6)*, 884-892.

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America, 123*(2), 1104-1113.

Rabagliati, H., Gambi, C., & Pickering, M. J. (2015). Learning to predict or predicting to learn? *Language, Cognition, and Neuroscience, 31*(1), 94-105.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95*(3), 385-408.
Rayner, K. (1998). Eye Movements in Reading and Information Processing. *Psychological Bulletin, 124*(3), 372-422.
Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The Effects of Frequency and Predictability on Eye Fixations in Reading: Implications for the EZ Reader Model. *Journal of Experimental Psychology, 30*(4), 720-732.
Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 514-528.
Rayner, K., Slowiaczek, M., Clifton, C., & Bertera, J. (1983). Latency of sequential eye movements: implications for reading. *Journal of Experimental Psychology: Human Perception and Performance, 9*(6), 912-922.
Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*(4), 504-509.
Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125-157.
Repp, B. H. (1992). Perceptual restoration of a "missing" speech sound: Auditory induction or illusion? *Attention, Perception, & Psychophysics, 51*(1), 14-32.
Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). *Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing.* Paper presented at the Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1.
Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology, 2*, 10.3389/fpsyg.2011.00307.
Roland, D., Yun, H., Koenig, J.-P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition, 122*(3), 267-279.
Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia, 85*, 245-255.
Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics, 77*(3), 720-730.
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.
Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: regular meter facilitates semantic sentence processing. *Neuropsychologia, 50*(2), 232-244.
Rugg, M. D. (1985). The Effects of Semantic Priming and Word Repetition on Event‐Related Potentials. *Psychophysiology, 22*(6), 642-647.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.
Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*(2), 149-169.
Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science, 12*(4), 110-114.
Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71*(1), 145-163.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General, 110*(4), 474-494.
Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348-351.
Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language, 48*(2), 416-434.
Sauppe, S. (2016). Verbal Semantics Drives Early Anticipatory Eye Movements during the Comprehension of Verb-Initial Sentences. *Frontiers in Psychology, 7,* 95.
Schank, R. C., & Abelson, R. (1977). *Scripts, Goals, Plans, and Understanding.* Hillsdale,NJ: Erlbaum.
Schotter, E. R., Lee, M., Reiderman, M., & Rayner, K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language, 83*, 118-139.
Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24*(2), 232-252.
Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action: Candidate roles for the motor cortes in speech perception. *Nature Reviews Neuroscience, 10*(4), 295-302.
Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*(2), 109-147.
Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2011). Shared syntax in language production and language comprehension—an fMRI study. *Cerebral Cortex, 22(7),* 1662-1670.
Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Cciences, 7(11),* 489-493.
Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage, 44*(3), 1133-1143.
Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences, 111(43),* E4687-E4696.
Sivonen, P., Maess, B., & Friederici, A. D. (2006). Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context. *Neuroscience letters, 408*(3), 220-225.
Sivonen, P., Maess, B., Lattner, S., & Friederici, A. D. (2006). Phonemic restoration in a sentence context: evidence from early and late ERP effects. *Brain Research, 1121*(1), 177-189.
Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition, 136*, 304-324.
Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: another motor theory of speech perception. In M. A. Arbib (Ed.), *Actio to language via the Mirror Neuron System* (pp. 250-285). Cambridge: Cambridge University Press.
Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning,Memory, and Cognition, 13*(1), 64-75.
Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73*(3), 205-246.

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(3)*, 675-686.
Smith, N. J., & Levy, R. (2011). *Cloze but no cigar: The complex relationship etween cloze, corpus, and subjective probabilities in language processing.* Paper presented at the 33rd Annual Meeting of the Cognitive Science Society, Boston, MA.
Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*(3), 302-319.
Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience, 32*(25), 8443-8453.
Söderström, P., Horne, M., Frid, J., & Roll, M. (2016). Pre-activation negativity (PrAN) in brain potentials to unfolding words. *Frontiers in Human Neuroscience*, doi: 10.3389/fnhum.2016.00512.
Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition, 7*(2), 77-85.
Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review, 18*(2), 371-376.
Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass, 9*(8), 311-327.
Staub, A., & Clifton, C. J. (2006). Syntactic prediction in language comprehension: Evidence from *either. . .or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 425-436.
Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1-17.
Staub, A., Grant, M., Clifton, C., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 806-814.
Staub, A., White, S., Drieghe, D., Hollway, E., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal Of Experimental Psychology. Human Perception And Performance, 36*(5), 1280-1293.
Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences, 107*(32), 14425-14430.
Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*(26), 10587–10592.
Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition, and Neuroscience, 31*(4), 484-503.
Swinney, D. A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of verbal learning and verbal behavior, 18*(6), 645-659.
Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language, 68*(4), 297-314.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1635.

Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journalism Bulletin, 30*(4), 415-433.
Thiessen, E. D. (2011). Domain general constraints on statistical learning. *Child Development*, *82*(2), 462-470.
Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382-392.
Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three-and four-year-old children. *Journal of Memory and Language*, *58*(2), 188-213.
Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*(7), 952-981.
Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1266.
Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454-475.
Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285-318.
Van Berkum, J. J. (2013). Anticipating communication. *Theoretical Linguistics*, *39(1-2),* 75-86.
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, and Cognition, 31*(3), 443-467.
Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580-591.
Van Den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967-985.
Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394.
Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190.
Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, *8*(4), 314-317.
Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, *1106*(1), 150-163.
Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392-393.
Warren, T., Dickey, M. W., & Lei, C.-M. (2016). Structural prediction in aphasia: Evidence from either. *Journal of neurolinguistics*, *39*, 38-48.
Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), B63-B72.

White, S. J. (2008). Eye Movement Control During Reading: Effects of Word Frequency and Orthographic Familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 205-223.
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*(7), 1272-1288.
Willems, R. M., Frank, S. L., Nijhoff, A. D., Hagoort, P., & Van den Bosch, A. (2016). Prediction during natural language comprehension. *Cerebral Cortex*, *26*(6), 2506-2516.
Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957-968.
Wlotko, E. W., Federmeier, K., & Kutas, M. (2012). To predict or not to predict: age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975-988.
Wlotko, E. W., & Federmeier, K. D. (2012a). Age ‐ related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*(6), 770-785.
Wlotko, E. W., & Federmeier, K. D. (2012b). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356-366.
Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20-32.
Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 209-216.
Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition*, *12*(1), 31-45.
Yan, S., Kuperberg, G. R., & Jaeger, F. T. (2017). Prediction (or not) during language processing. A commentary on Nieuwland et al. (2017) and DeLong et al. (2005). *bioRxiv*, doi:10.1101/143750
Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, *28*(3), 272-302.