Neuropsychologia 136 (2020) 107258

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward

Rachel Ryskin[a,*], Roger P. Levy[a], Evelina Fedorenko[a,b,c]

[a] Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, USA
[b] McGovern Institute for Brain Research, Massachusetts Institute of Technology, USA
[c] Department of Psychiatry, Massachusetts General Hospital, USA

ABSTRACT

Most current accounts of language comprehension agree on a role for prediction, but they disagree on the importance of domain-general executive resources in predictive behavior. In this opinion piece, we briefly review the evidence for linguistic prediction, and the findings that have been used to argue that prediction draws on domain-general executive resources. The most compelling evidence is an apparent reduction in predictive behavior during language comprehension in populations with lower executive resources, such as children, older adults, and second language (L2) learners. We propose that these between-population differences can be explained without invoking executive resources. Instead, differences in the quantity and kind of language experience that these populations bring to bear may affect the probability of engaging in predictive behavior, or simply make prediction effects more difficult to detect in paradigms designed for young adult native speakers. Thus, domain-specific prediction mechanisms remain a viable possibility. We discuss ways to further test accounts of linguistic prediction that do vs. do not require domain-general executive resources, using behavioral, computational, and brain imaging approaches.

Prediction is ubiquitous in human cognition (Clark, 2013; James, 1893): we anticipate trajectories of moving objects, upcoming notes in a melody, and others' emotional reactions. In recent years, evidence has accumulated that language comprehension is similarly not a passive experience in which we receive linguistic input and then process it. Instead, listeners and readers appear to actively predict upcoming material based on what they know about language and the world (Dell and Chang, 2013; Federmeier, 2007; Kuperberg and Jaeger, 2016; Lupyan and Clark, 2015; Pickering and Garrod, 2013). When incoming information conforms to these predictions, processing is facilitated, and violations of these predictions incur processing costs.

The link between linguistic predictability and behavioral or electrophysiological outcomes has been extensively investigated. However, the underlying mechanisms remain a topic of debate (Hasson et al., 2018; Hauk, 2016). In particular, does linguistic prediction rely on *language-specific* mechanisms – i.e., the mechanisms that store our language knowledge and use those knowledge representations to interpret incoming input, or does it instead, or in addition, require *domain-general* executive resources, like working memory and cognitive control?

In this piece, we briefly introduce linguistic prediction (Section 1). We then review the evidence that has been used to argue for a role of domain-general executive resources in prediction during language comprehension – namely, reduced prediction in populations with limited executive resources (Section 2), and speculate about ways in which such resources could be involved in predictive behavior (Section 3). We then provide a re-interpretation of key patterns discussed in Section 2 with respect to between-population differences in linguistic experience (Section 4). Finally, we discuss experimental and computational approaches that could be used to distinguish between the two possibilities in future work (Section 5) and conclude (Section 6).

1. Prediction in language comprehension

Evidence for a predictive processing view of language comes from several experimental paradigms (for a recent review, see Kuperberg and Jaeger, 2016; cf. Nieuwland, 2019). On-line measures of sentence reading reveal that readers spend less time processing a word (as evidenced by e.g., shorter average fixation duration or lower probability of fixation) when that word is highly predictable given the preceding sentence context (e.g., the word "shark" in "The coast guard warned that someone had seen a *shark* off the north shore of the island.") compared to when it is not predictable (e.g., "The zookeeper explained that the life span of a *shark* is much longer than those of other animals."; Ehrlich and Rayner, 1981). Similarly, measures of eye movements recorded while listeners hear a sentence and observe a visual scene (the visual world eye-tracking paradigm; Tanenhaus et al., 1995) reveal that listeners

* Corresponding author.
E-mail address: ryskin@mit.edu (R. Ryskin).

https://doi.org/10.1016/j.neuropsychologia.2019.107258
Received 4 March 2019; Received in revised form 7 November 2019; Accepted 7 November 2019
Available online 12 November 2019

incrementally look to objects in the scene that are likely to be referred to next (e.g., more looks to a cake after hearing “the boy will eat the …” compared to “the boy will move the …“; Altmann and Kamide, 1999).

Studies using recordings of electrical activity over the scalp (EEG/ERP) have identified multiple components of the signal that are modulated by linguistic context. For instance, the N400—a negative deflection that peaks approximately 400 ms after the onset of a critical word—is reduced when that word is predictable in context (e.g., “He planted string beans in his *garden*.“) relative to when it is not (e.g., “He planted string beans in his car.“; Kutas and Hillyard, 1984; inter alia Dambacher et al., 2006; Dimigen et al., 2011; for a review see Kutas and Federmeier, 2011). Moreover, the N400 is also reduced to incongruous words that share features with the predicted word, relative to a word that belongs to a different category than the predicted word (e.g., “The yard was completely covered with a thick layer of dead leaves. Erica decided it was time to get out the shovel (vs. hammer).” [“rake” is predicted]; Federmeier and Kutas, 1999). The N400 is thus thought to index the process of retrieving the meaning of the critical word from multimodal, long-term memory and its amplitude is therefore tied to how much of the nearby semantic space has been pre-activated by the context. Further, a later positive deflection appears to index situations where a strong prediction was generated and then disconfirmed (e.g., “The groom took the bride’s hand and placed the ring on her *dresser*.” [“finger” is predicted]; Federmeier et al., 2010; Federmeier et al., 2007), perhaps reflecting the neural activity related to the listener updating their model of the language.

Recent computational accounts of language comprehension have formalized processing difficulty in terms of surprisal (Hale, 2001; Levy, 2008): the negative log probability of the word, w_i, given the preceding context, where the context encompasses the previous words w_1 through w_{i-1}, in the sentence and any context outside of the sentence, C.

$$surprisal(w_i) = -\log P(w_i \mid w_1, \ldots, w_{i-1}, C)$$

The surprisal of a word is inversely related to its predictability. When a word is highly probable given its context, surprisal is low. When a word is not likely given the context, surprisal is high.

Reading time differences among words (on a log scale), as well as N400 amplitudes, are well approximated by word-by-word surprisal values estimated from large corpora (Frank et al., 2015; Luke and Christianson, 2016; Smith and Levy, 2013), linking the mechanism of prediction directly to the statistical properties of the language. Of course, many factors beyond the conditional probability of a word given the preceding words likely affect the prediction that a person may generate, including who they are talking to, all manner of sensory-perceptual input, and world knowledge (e.g., Heller et al., 2008; Kamide et al., 2003; Van Berkum et al., 2008), but these are more challenging to quantify. As a result, most prior work on linguistic prediction has construed predictability as the probability of a word given preceding linguistic input alone – often estimated using n-gram frequencies or cloze task responses.

The precise mechanism of prediction, and how it differs from other comprehension-related processes, like bottom-up integration of incoming elements, remains an area of active debate in the literature (Ferreira and Chantavarin, 2018; Kuperberg and Jaeger, 2016; Mantegna et al., 2019; Nieuwland et al., 2019; Pickering and Gambi, 2018). Several accounts ground prediction in a process of forward-simulating through the language production system (Dell and Chang, 2013; Federmeier, 2007; Pickering and Garrod, 2013). But these and other accounts vary along at least two key dimensions, which – although independent – sometimes co-vary across proposals.

One dimension concerns the *ubiquity* of predictive processing during language comprehension. Some argue that prediction is a core component of language comprehension, and so all humans always predict upcoming linguistic events (Fitz and Chang, 2019; Kuperberg and Jaeger, 2016; Rabovsky et al., 2018), in line with general predictive processing accounts of cognitive and neural functioning (Clark, 2013; Friston, 2010; Keller and Mrsic-Flogel, 2018; Rao and Ballard, 1999). In contrast, others postulate that prediction is an optional component of language processing (e.g., Huettig and Mani, 2016; Pickering and Gambi, 2018), which may take time to mature over the course of human development (Gambi et al., 2018; Pinker, 2009; Rabagliati et al., 2016; cf. Chang et al., 2006; Chang et al., 2012; Dell and Chang, 2013; Elman, 1990; Fitz and Chang, 2019; Ramscar et al., 2013). Whether or not a comprehender engages in predictive processing may be determined by some utility function based on a trade-off between the cognitive effort required for prediction vs. the resulting processing benefits, properties of the context (e.g., how much evidence the context contains for a predictable next word), and the amount of cognitive resources available, inter alia.

The second dimension concerns the *nature* of the predicted information. In particular, proposals vary with respect to the granularity of predictions: do comprehenders predict one most likely continuation (Van Petten and Luka, 2012) or multiple possible continuations, weighted by their likelihood (e.g., Fitz and Chang, 2019; Kuperberg and Jaeger, 2016; Levy, 2008)?

And to what extent do predictions feed back down from higher-level (e.g., semantic/syntactic) features to lower-level (orthographic/phonological) ones (DeLong et al., 2005; Nicenboim et al., 2019; Nieuwland et al., 2019; Van Berkum et al., 2005; Wicha et al., 2004; Yan et al., 2017)?

Understanding whether and how domain-general executive resources (e.g., Friedman and Miyake, 2017) affect predictive linguistic behavior can importantly constrain the possibilities above and is an important step towards uncovering the essential computations and representations that support predictive processing in language comprehension.

2. Evidence consistent with a role of domain-general executive resources in linguistic prediction

Domain-general executive functions encompass a wide range of cognitive processes, from working memory maintenance and updating, to inhibitory control, to set shifting (e.g., Friedman and Miyake, 2017). These processes have been implicated in goal-directed behavior, broadly construed (Duncan, 2010a). Over the years, many have argued for the importance of executive processes in language comprehension (e.g., Just and Carpenter, 1992; Nozari et al., 2016a), including their potential role in core linguistic processes, like inhibiting irrelevant meanings or parses (e.g., Novick et al., 2005). Here we ask whether executive resources are critical for *predictive* language processing.

The most compelling evidence for the role of domain-general executive resources in linguistic prediction comes from an apparent reduction in prediction[1] in populations with limited executive resources, including children (Friedrich and Friederici, 2005; Gambi et al., 2018; Mani and Huettig, 2012), older adults (Dagerman et al., 2006; Dave et al., 2018; Federmeier and Kutas, 2019; Federmeier et al., 2010; Federmeier et al., 2002; Payne and Federmeier, 2018; Wlotko and Federmeier, 2012; see Payne and Silcox, 2019 for a review), and second language (L2) learners (Grüter et al., 2012; Lew-Williams and Fernald, 2010; Martin et al., 2013; Mitsugi and Macwhinney, 2016). Executive functions do not reach full maturity until early adulthood (Davidson

[1] Empirical evidence of “prediction” may originate both from measures of pre-stimulus (anticipatory) activity and from activity during/after the (not) predictable stimulus. Though the distinction between these signals may be crucial to the understanding of the mechanisms of prediction, they have typically been discussed together in the literature and used interchangeably as indices of predictive processes. Because both indices can be fruitfully leveraged to investigate the key question we are interested in here – whether executive resources play a crucial role in linguistic prediction – we follow the prior literature in including both types of effects in our discussion.

et al., 2006; De Luca and Leventer, 2010), and they begin to decline soon thereafter (Hartshorne and Germine, 2015; Park et al., 2002; Salthouse, 2009). So, reduced linguistic prediction in both children and older adults has been argued to result from this lower amount of available executive resources (e.g., Huettig and Mani, 2016; Pickering and Gambi, 2018). And L2 learners may find language processing more effortful thus loading on the same pool of executive resources that would typically be used for prediction (Linck et al., 2014). These between-population differences (see Table 1 for a summary of some representative findings) provide the most compelling evidence to date for a core role of executive resources in linguistic prediction.

Some investigations of individual differences among native-speaking young adults have also been used to argue for the role of domain-general executive resources in linguistic predictions. In particular, many behavioral, and some electrophysiological, investigations (e.g., Caplan and Waters, 1999; Just and Carpenter, 1992; King and Just, 1991; Misyak and Christiansen, 2012; Payne et al., 2014; Swets et al., 2007; Van Dyke et al., 2014) have found that readers who perform well on tasks that measure executive resources (e.g., reading span) are better at understanding sentences with high surprisal (often discussed in terms of syntactic complexity or memory demands but these are typically correlated with surprisal) or exhibit greater differences in ERP (e.g., P600) response magnitude for predictable vs. unpredictable words (Kim et al., 2018; Nakano et al., 2009; Tanner and Van Hell, 2014). However, much of this evidence suffers from limitations, including poor psychometric properties of the comprehension measures (see James et al., 2018 for relevant discussion) and lack of agreement on the right instruments for assessing executive functions (e.g., Stroop vs. an anti-saccade task). Further, correlations between two measures x and y provide only an indirect route to understanding underlying processes, as they may reflect a common cause (e.g., motivation or general intelligence, factor *g;* Duncan, 2010a; Spearman, 1929). As a result, we here focus on the claims that have come from the between-population comparisons.

3. How could domain-general executive resources affect linguistic prediction?

Broadly speaking, executive resources could affect i) the likelihood of engaging in predictive behavior (cf. proposals whereby prediction is ubiquitous), and/or ii) the nature and/or quality of the predictions. To our knowledge, no specific account of a mechanism by which executive functions might support predictive processing during language comprehension has been proposed in the literature, although most theoretical papers have argued for the lower likelihood of engaging in predictive behavior in populations with limited executive resources rather than for changes in the nature/quality of the predictions (cf. Borovsky et al., 2012; Kaan, 2014). Here, we sketch three specific hypotheses, which are all instantiations of the broad Hypothesis (Hypothesis 1 in Fig. 1, left panel) whereby domain-general executive resources play a role in linguistic prediction.[2] In Hypotheses 1a and 1b, linguistic predictions are *generated* by language-specific mechanisms, and executive resources play a *critical supporting role*, and in Hypothesis 1c, executive resources play a *core role.*

Hypothesis 1a. *Executive resources are needed to maintain the context in working memory.*

To generate precise and accurate predictions about upcoming linguistic events, a comprehender needs a faithful representation of the preceding linguistic (and non-linguistic) context. Executive resources (in particular, a domain-general working memory store) could be critical for maintaining these contextual representations and integrating diverse kinds of information (e.g., the identity of the interlocutor, sensory-perceptual input, world knowledge). In this way, the amount of executive resources would determine the quality of the input that the language model uses to generate predictions (e.g., Futrell and Levy, 2017). When executive resources are plentiful, the representation of the context in working memory is detailed, accurate, and plausibly spanning a longer time-window, allowing for specific, and more likely, correct, predictions to be made, thus leading to facilitation when the predicted continuation is encountered. On the other hand, when executive resources are limited, the memory representation of the context may be of poor quality, leading to less specific and sometimes inaccurate predictions, and thus frequent difficulty.

Hypothesis 1b. *Executive resources are needed to generate/maintain predictions in working memory.*

Executive resources could also be important for keeping the generated predictions active in working memory as linguistic input continues to unfold, especially if our linguistic mechanisms generate multiple possible continuations. When executive resources are plentiful, we can maintain several likely continuations or features thereof active in working memory (weighted by their probability), thus maximizing the chances of the input matching those predictions and leading to facilitation when there is a match. On the other hand, when executive resources are limited, the representation of the generated prediction may be of poor quality (e.g., perhaps only a single likely continuation is maintained, or only a subset of the relevant features, or perhaps even the weighting of the continuations/features is affected), decreasing the probability of the input matching the predictions and leading to processing difficulty.

Hypothesis 1c. *Linguistic prediction is implemented in domain-general inhibitory and selection mechanisms.*

This hypothesis construes prediction as inhibition of low-probability continuations and selection of high-probability continuations. In particular, selection and inhibitory control acting in tandem (Mirman et al., 2011; Nozari et al., 2016a, 2016b) may be essential to circumscribe activation to the likely continuation(s) or features thereof. When executive resources are plentiful, (the features of) high-probability continuations are active and (the features of) low-probability continuations are efficiently inhibited, leading to facilitation when the predicted continuation is encountered and difficulty when a non-predicted continuation is encountered. On the other hand, when executive resources are limited, the selection and inhibition of (the features of) high- and low-probability continuations respectively may not always be successful, leading to less specific and less accurate predictions.

4. Reinterpreting the existing evidence for the role of domain-general executive resources in linguistic prediction

As noted above, taken at face value, the reduction or absence of prediction in populations with lower executive function abilities may appear to provide compelling support the for the role of executive resources in linguistic prediction. However, we here argue that this evidence can be accounted for without alluding to executive resources. The key insight is that young native-speaking adults differ from children, older adults, and L2 learners not only in the availability of executive resources, but also in the amount and kind of linguistic experience. We outline two specific hypotheses, which are both instantiations of the broad hypothesis (Hypothesis 2 in Fig. 1) whereby domain-general executive resources do not play a role in linguistic prediction. According to Hypothesis 2a, the previously reported population differences are

[2] In what follows we do not restrict the use of "prediction" to any particular level of representation, though the majority of existing literature has focused on lexical prediction. A discussion of whether/how executive resources may be differentially involved in prediction at different levels of representation is beyond the scope of this piece. In the absence of evidence to the contrary or a more precise understanding of how predictions are computed in the first place, it seems most parsimonious to start with the assumption that whatever mechanism may or may not change the process of linguistic prediction over the lifespan operates similarly across levels of representation.

Table 1

Examples of studies of predictive language processing across populations. This set of studies is not meant to be exhaustive. Instead, we tried to showcase the kinds of linguistic prediction effects that have been investigated across cohorts. In some of these studies, the authors did not explicitly link linguistic prediction to executive functions. (L2 = Second language).

Population	Paper	Method	Indices of predictive language processing	Population comparison
Children (3–10 years) vs. young adults (18–28 years)	Borovsky et al. (2012)	Visual World Paradigm eye-tracking	More anticipatory fixations to target picture when it is plausible given the agent and action verb than when it is not.	Children and adults anticipate the targets but the effect is smaller in children with low vocabulary.
Children (2–5 years) vs. young adults (18–24 years)	Gambi et al. (2018)	Visual World Paradigm eye-tracking	More anticipatory fixations to target picture after a disambiguating determiner (a/an) than after one that doesn't disambiguate the subsequent referent.	Adults anticipate the targets, but the effect is smaller or absent in children.
Children (1 year vs. 19 months) vs. adults	Friedrich and Friederici (2005)	Event-related potentials (looking at pictures while listening to labels)	Larger N400 responses when listening to a word that's incongruous with the picture than when it matches the picture.	Adults and 19-month-olds show an N400 difference but not 12-month-olds.
Children (2 years)	Mani and Huettig (2012)	Visual World Paradigm eye-tracking	More anticipatory fixations to target picture when it is plausible given the action verb than when it is neutral.	Children with high productive vocabulary anticipate the targets. Those with low vocabulary don't. (No adult control group.)
L2 learners of English (native language: Spanish)	Lew-Williams and Fernald (2010)	Visual World Paradigm eye-tracking	More anticipatory fixations to target picture after a gender-disambiguating determiner (in Spanish) than after one that doesn't disambiguate the subsequent referent.	Unlike native speakers, L2 learners are not able to take advantage of gender information to anticipate the target.
L2 learners of English (native language: Spanish)	Grüter et al. (2012)	Visual World Paradigm eye-tracking	More anticipatory fixations to target picture after a gender-disambiguating determiner (in Spanish) than after one that doesn't disambiguate the subsequent referent.	Unlike native speakers, L2 learners are not able to take advantage of gender information to anticipate the target.
L2 learners of English (native language: Spanish) vs. native speakers	Martin et al. (2013)	Event-related potentials during reading	Larger N400 responses to determiners predictive of words that are unexpected in context.	Native speakers, but not L2 learners showed effects of prediction on the N400.
L2 learners of Japanese (native language: English) vs. native speakers	Mitsugi and Macwhinney (2016)	Visual World Paradigm eye-tracking	More anticipatory fixations to target picture when it fits the syntactic context (case-marking information) compared to when the context is predictive of a syntactic alternative.	Native speakers, but not L2 learners showed a preference for the target consistent with the case-marking cue.
Older adults (mean age: 73 years) vs. young adults (under 30)	Dagerman et al. (2006)	Cross-modal naming	Speakers are faster to name ambiguous words when the preceding sentence context is supportive compared to un-supportive.	Young adults are faster in supportive contexts but older adults are not.
Older adults (64–79 years) vs. young adults (18–33 years)	Dave et al. (2018)	Event-related potentials during reading	Larger N400 responses to words that are unexpected in context (vs. expected) and larger Post-N400 Positivity (PNP) when the context is strongly constraining.	N400s are reduced in older adults but not PNPs
Older adults (60–76 years) vs. young adults (18–24 years)	Federmeier et al. (2010)	Event-related potentials during reading	Larger late frontal positivity (FP) in response to a congruent word when a different word was strongly predicted (but disconfirmed by the input).	FPs are reduced in older adults relative to young adults.
Older adults: continuous age range (32–77 years)	Huettig and Mani (2016)	Visual World Paradigm eye-tracking	More anticipatory fixations to target picture which fits the context, a gender-marked determiner, than to distractors which are not consistent with the determiner gender.	Target to distractor fixation ratio is higher in participants with higher working memory, which is (negatively) correlated with age in this sample.

artifactual in nature. In particular, all comprehenders engage in prediction, but they differ in the nature of the predicted information. As a result, *population-specific norms* are needed to detect effects of prediction. And according to Hypothesis 2b, prediction is, in fact, reduced in children, older adults, and L2 learners, but due to differences in the amount and nature of linguistic experience, leading to less utility associated with predictive behavior.[3]

Hypothesis 2a. *Ubiquitous but differing predictions across populations.*

The predictions that comprehenders make reflect the statistics of their prior linguistic experience (see also Cuetos et al., 1996; Fine et al., 2013; Levy, 2008; MacDonald, 2013; Ryskin et al., 2017; Verhagen et al., 2018). Linguistic input differs in quantity and nature between young native-speaking adults, children, older adults, and L2 speakers, leading to different language models used to predict upcoming linguistic events (for a similar argument focused on L2 comprehension specifically see Kaan, 2014).

In many experiments that have investigated linguistic prediction across populations (see Table 1), norms (e.g., cloze task responses) from young adults are used to create stimuli and categorize them into high- and low-predictability conditions (or more fine-grained bins in some cases). The implicit assumption is that, to the extent that any other group of participants (not native-speaking young adults) engages in predictive behavior, the content of their prediction will match those norms. We here argue that the nature of predictions may differ between groups due to differences in their language experience. Naturally, older adults have more years of experience with their native language than young adult native speakers, whereas children and L2 speakers have much less experience. As a result, older adults are more likely to have heard infrequently occurring words, phrases, or structures (Ramscar et al., 2014), whereas children and L2 learners begin by learning higher frequency words (Braginsky et al., 2019; Goodman et al., 2008). In addition, over the many years of an older adult's experience, the language is itself changing (e.g., Biber and Finegan, 1989; Wolk et al., 2013). As languages evolve, some words and structures fall out of fashion (e.g., "davenport"), while others are born (e.g., "to google", "because [Noun]"). Furthermore, the language experience of these

[3] One interesting thing to note is that, by some metrics (e.g., vocabulary knowledge, syntactic judgments, offline text comprehension), our language abilities are stable or continue to improve with age, well into our 70s–80s (Hartshorne and Germine, 2015; Shafto and Tyler, 2014; but see Johnson, 2003; Payne et al., 2014). If it turns out that older adults engage less in predictive behavior than younger adults, this would suggest that linguistic prediction only benefits the ease of processing incoming words in real time, but does not impact our ability to extract information from linguistic signals (see Huettig and Mani, 2016; Pickering and Gambi, 2018 for proposals along these lines).

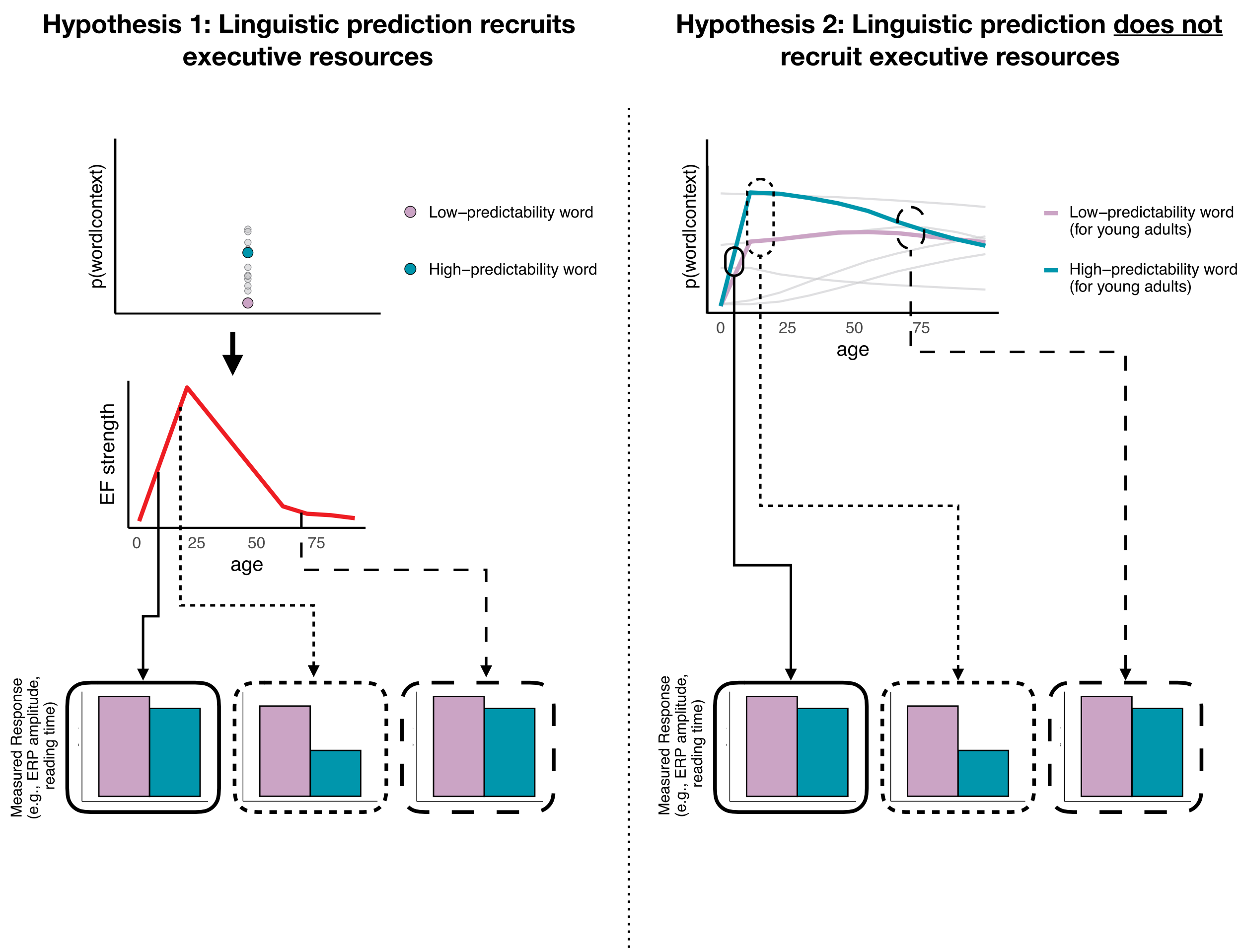

Fig. 1. A schematic illustration of the two main hypotheses that can give rise to the same data pattern: one where domain-general executive resources play a role in linguistic prediction, and one where they don't. ***Left***: The top panel shows the probabilities assigned to two words, one high- and one low-predictability, by the comprehender's language model (grey dots are other words), and the middle panel shows the strength of executive functions (EF) over the lifespan. Because the amount of executive resources affects the ability to generate and maintain linguistic predictions (see Hypotheses 1a-c for specific proposals), responses to the low-predictability and high-predictability word differ more, when EF is high (i.e., around age 20) compared to when EF is low (e.g., around ages 5 or 75, or in L2 speakers [not depicted for clarity]). ***Right***: The top panel shows the probabilities of words in context across the lifespan. The high-predictability and low-predictability words are defined in terms of probabilities assigned by a young adult at the time of data collection (grey lines are other words). In this Hypothesis (see Hypotheses 2a-b for specific proposals), generating and maintaining linguistic predictions does not require executive resources. However, due to variation in the respective probabilities of words over the lifespan, responses to the low-predictability (for a young adult) and high-predictability (for a young adult) word also differ more around age 20 than around ages 5 or 75 (L2 speakers can be thought of as having ~5 years of "age" in terms of cumulative non-native language exposure, though other features of L2 linguistic experience beyond cumulative exposure, e.g., first language transfer, plausibly also play a role). On the bottom, the same pattern is shown as in the left panel but with a different explanatory mechanism.

populations likely differs in a number of other ways (see Wulff et al., 2019 for a review of how the lexicon changes over the lifespan). For instance, the typical source of recent language input (e.g., news outlets vs. social media vs. children's books; Montag et al., 2015; Montag and MacDonald, 2015), the quantity of recent input (e.g., older adults may be more likely to live alone and have fewer interactions, while L2 speakers are perhaps spending most of their time experiencing a different language) and sensory processing (e.g., vision and hearing may be impaired in old age) may all differ between these groups and have downstream consequences for the content of their linguistic predictions.[4] The existing evidence for reduced prediction in children, older adults, and L2 learners is therefore plausibly explained by the use of young adult language norms which do not carve up the predictability space in the appropriate way to detect prediction in the other three populations.[5]

We are not the first to consider the possibility that different populations may have different linguistic expectations. Federmeier et al. (2002) noted this issue with respect to older adults but did not observe significant differences when comparing cloze task responses of 20 older adults to those of 132 young adults. On the other hand, inspection of the Hamberger et al. (1996) cloze norms from 30 young adults and 100 older adults reveals many completions that are produced only by the young cohort or only by the old cohort. Similarly, Lahar et al. (2004) reported some consistency in cloze responses across age groups (young, middle, young-old, old-old) but stronger correlations between age groups that were closer together in age. Further, in a sentence completion paradigm, L2 speakers of English were found to have different subcategorization biases (Dussias et al., 2010) than native English speakers (Garnsey et al., 1997), and children's grammaticality judgments diverge from those of adult, native speakers (Ambridge et al., 2008). Thus, the possibility that differential linguistic experience leads to different kinds of predictions during language comprehension – which do not necessarily match the normative patterns from young adults – deserves further investigation.

Hypothesis 2b. *Reduced prediction in populations with either too little or "too much" linguistic experience.*

Prior reports of reduced prediction in children, older adults, and L2 learners have emphasized that generating prediction is costly (Pickering and Gambi, 2018). (The exact nature of this cost has not been fleshed out in the literature, but it would presumably depend on the nature of the postulated mechanism for how predictions are generated and maintained; see e.g., Hypotheses 1a-c above.) As a result, in populations where executive resources are immature, declining, or taxed, the *costs* may outweigh the benefits of engaging in predictive behavior. However, an alternative explanation is that the cost of generating predictions is negligible and/or does not draw on domain-general executive resources,[6] but the *benefits* are insufficient to justify prediction in some populations. If predictions confer little to no processing advantage, engaging in predictive behavior would not be rational (Anderson, 1990). (Though note that prediction error in this case can still serve the purpose of allowing the comprehender to learn and update their language model.)

One possibility is that there is some optimal amount of language experience at which our predictions are correct (elicit small prediction error signals) frequently enough for them to be useful in facilitating the processing of incoming words. For example, children and L2 learners have relatively small vocabularies. As a result, their predictions would necessarily be wrong whenever they encounter an unfamiliar word, which would happen a lot early on during language acquisition. In other words, prediction may not substantively facilitate processing in these populations leading them to engage in it less. In line with this idea, vocabulary is a better predictor than age for predictive processing in children (Borovsky et al., 2012; also see Ylinen et al., 2017) and verbal fluency is positively correlated with the magnitude (or presence) of prediction-related ERP effects in older adults (Federmeier et al., 2010). On the other hand, vocabulary keeps increasing with age (e.g., Hartshorne and Germine, 2015; Verhaeghen, 2003). Integrating over language experience from several decades and potentially more idiosyncratic sources (as described in Hypothesis 2a) may lead older adults to generate a very different set of predictions than young adults. This may lead to frequently incorrect predictions in certain language settings (e.g., doctor's visits, interacting with younger family members, psychology research experiments) where the distributions may be more reflective of the younger person's linguistic distribution, and thus to predictive behavior being less useful than for young adults.

The precise mechanisms for the switching on and off of predictive behavior would need to be worked out, but we can speculate that – starting with some non-zero probability of making a prediction – the probability of generating a prediction *p* increases when the previously generated prediction *p-1* turned out to be correct (this increase can be driven via e.g., the activity of dopamine neurons; e.g., Schultz et al., 1997; Stauffer et al., 2014; Steinberg et al., 2013), and decreases when *p-1* turned out to be incorrect. In children and L2 learners, a high number of incorrect predictions would lead to an overall low probability of making a prediction (although this probability never gets to 0; otherwise, a separate mechanism would be required for initiating the process again at a later age). By adolescence/young adulthood, linguistic predictions are correct sufficiently frequently, so the probability of making a prediction is approaching 1. As individuals age, the proportion of incorrect predictions is increasing again, leading to a decrease in the probability of future predictions. This additional process may also be subject to individual differences, such that some people's predictive mechanisms are more strongly affected by the ratio of correct to incorrect predictions.

Adding a mechanism for probabilistically switching prediction on and off makes this proposal more complex than Hypothesis 2a. Indeed, of the five hypotheses described above, Hypothesis 2a provides the simplest account of all the relevant phenomena. However, although

[4] Could differences in how well context is encoded fully explain differences in prediction, without appealing to differences in linguistic experience? For example, older adults may have more difficulty hearing speech, especially in noisy situations, or seeing small fonts, and may therefore miss some critical parts of the context. As a result, they may predict a different continuation than a young adult who encoded the context faithfully. Such differences are unlikely to fully account for the relevant prior findings because older adult participants are typically screened for auditory and visual acuity. Moreover, children and L2 learners are not likely to differ from young adults in their ability to perceive the context, arguing against a general low-level perceptual mechanism as an explanation of reduced prediction in older adults, children, and L2 learners.

[5] It is worth noting that some decrease in the difference between high- and low-predictability items across populations is expected simply due to regression to the mean. Maximizing the difference between responses in one group will lead to smaller differences in any other group (for those same stimuli). However, it seems unlikely that regression to the mean would be the sole factor, given that norming and testing are typically done across different samples and yet prediction effects have been robustly observed. Therefore, differences in the quantity and sources of the linguistic input may play an important role in explaining prediction discrepancies between populations.

[6] Another plausible source of prediction cost is forward-modeling through the production system, which has been proposed as a key element of prediction (Pickering and Gambi, 2018; Federmeier, 2007; Dell and Chang, 2013). If these production costs are thought to be at the level of neural implementation (e.g., speed of synaptic transmission), then—as with perceptual acuity (discussed in footnote 4)—this explanation is pertinent to older adults and, perhaps, very young children, but not to L2 speakers. Alternatively, these production costs could reflect the higher-level processes of accessing/assembling utterances (e.g., children/L2 speakers may be lacking certain representations, whereas in older adults they may be difficult to access).

parsimony favors Hypothesis 2a (Myung, 2000), additional systematic empirical work is needed to rigorously evaluate all of the hypotheses above.

5. A path forward: how to distinguish between proposals that do vs. do not require domain-general executive resources for linguistic prediction?

5.1. Behavioral and electrophysiological investigations

The first key step in moving forward would be to assess prediction in children, older adults, and L2 learners using population-specific norms. In particular, prediction in each relevant population (young adults, children, older adults, and L2 learners) would be measured basing the predictability manipulation in the experimental materials on norms produced by the target population vs. by one of the other three groups. The use of young adult norms should replicate prior findings of reduced prediction (as indexed by eye-movements, ERPs, or other measures) in children, older adults, and L2 learners. The critical question is whether the use of population-specific norms (i.e., using norms from older adults in testing prediction in older adults) would eradicate these effects. The use of norms from children, older adults, or L2 learners to test prediction in young adults should also lead to an apparent reduction in prediction effects in this population. If these predictions of Hypothesis 2a hold, that would provide compelling evidence for a non-executive-resource-based explanation of between-population differences.[7]

Another approach to evaluating the role of executive resources in linguistic prediction would be to manipulate the availability of such resources *within individuals* over time using a resource-depletion paradigm (Muraven and Baumeister, 2000; but see Xu et al., 2014), where an individual performs a linguistic prediction task before and after performing a demanding executive function task, or a dual-task paradigm (Kahneman, 1973), where an individual performs a linguistic prediction task in parallel with an executive function task. Observing reduced effects of prediction following executive-resource depletion or in the presence of a secondary task would suggest that executive resources are indeed important for some aspect(s) of linguistic prediction. To meaningfully interpret the results of such an experiment, it would be critical to compare the linguistic prediction task to two other critical tasks: one where executive resource depletion is expected to affect performance (e.g., another executive function task), and one that should not be affected by executive resource depletion (e.g., a face perception task). To argue for the role of executive resources in linguistic prediction, one would need to demonstrate a reduction in linguistic predictive behavior following executive resource depletion, similar to impaired performance on the executive function task, and different from unimpaired performance on the control (face perception) task.

If the role of executive resources in linguistic prediction is supported by future empirical investigations, it will be important to articulate specific testable accounts of how exactly executive processes support predictive behavior (e.g., see Hypotheses 1a-c above). If different accounts favor different kinds of executive processes (e.g., working memory updating vs. inhibitory control), then more targeted individual-differences investigations may be able to tease them apart.

5.2. Computational modeling

The effects of language experience vs. executive resources on linguistic prediction can also be fruitfully investigated with a computational approach. A model trained to predict the next word in a sentence (e.g., n-gram, RNN, incremental grammar-based parser) has the training set as its language experience and presumably no "executive resources." Thus training computational models on different amounts[8] and kinds of data (e.g., texts from different time periods or genres) and comparing how the models assign relative probabilities to words from the same test set is analogous to using norms from one population to assess prediction in other populations. For example, a model trained on a corpus of texts written between 1920 and 1950 can be used to generate a probability distribution over words given a context (e.g., She likes sugar in her …). The most probable continuation, w1 (e.g., tea), and an improbable continuation, w2 (e.g., socks), can be selected and the difference between their surprisal values, Δ_{S1}, can be computed. Next a model trained on a corpus of texts written between 1970 and 2000 can generate a distribution over potential continuations given the same context and a Δ_{S2} can be computed for the same pair of words (e.g., tea and socks). It may be the case that on average, over many contexts and word pairs, Δ_{S1} and Δ_{S2} will not differ. On the other hand, Δ_{S2} may be on average smaller than Δ_{S1}, perhaps because the relative probabilities of continuations change over time (e.g., *coffee* replaces *tea* as the highest probability continuation in the modern corpus).[9] The latter result would be similar to what is observed in behavioral and ERP experiments: larger predictability effects when high-predictability and low-predictability stimuli are defined by language users with a similar linguistic background (e.g., young adult participants in an experiment based on norms from young adults) than when the comprehender and source of stimuli differ in linguistic experience (e.g., older adult participants in an experiment based on norms from young adults). Convergence in the results from this kind of a computational investigation and experiments like the ones outlined above could provide compelling evidence regarding what executive resources need to be posited to account for apparent differences in linguistic prediction between populations.

5.3. Brain imaging

We now know that the human brain is comprised of a number of large-scale functionally distinct networks (e.g., Alexander-Bloch et al., 2013; Bernard et al., 2012; Chen et al., 2012; Fox et al., 2005; Hagmann et al., 2008; Konopka et al., 2012; Power et al., 2011; Raznahan et al., 2011; Seeley et al., 2009; Toro et al., 2008; van den Heuvel and Sporns, 2011; Yeo et al., 2011; inter alia). Given that different networks have been associated with distinct kinds of cognitive processes, understanding which network(s) give rise to linguistic prediction effects can help decipher the underlying mechanisms (Mather et al., 2013). In particular, language comprehension has been linked to at least two brain networks:

[7] We are grateful to an anonymous reviewer for pointing out another avenue for experimentally dissociating language experience and executive function. This could be accomplished by teaching participants—from all four populations—an artificial language and then testing them with sentences from this novel language which either end with a predictable "word" or one that is unpredictable given the context. One challenge with this approach, apart from the lack of ecological validity and potentially additional executive demands associated with learning novel stimuli, is to ensure that the (artificial) language experience is indeed equated across the groups. A potential solution may be to allow participants to learn to some high accuracy criterion before testing their predictions.

[8] More training data/experience typically leads to better accuracy. However, this assumes that the training data and the test data are generated from the same underlying (in this case, language) distribution. This assumption may not be valid for humans belonging to different groups (e.g., young vs. old), if we think of the training data as all the language that an adult has experienced up until the point where they enter the lab for an experiment. Often the test set is generated by aggregating the distributions reported by young adult participants (i.e., norming). An older adult participant certainly has more years of language experience (i.e., a "larger" training set) than a young adult participant entering the lab, but the young adult may have a model much more similar to that which generated the test set, yielding potentially higher accuracy, on this particular prediction task, despite an overall smaller amount of input.

[9] In part, Δ_{S2} is likely to be smaller than Δ_{S1} simply because of regression to the mean (see footnote 5).

a language-specific fronto-temporal network that selectively supports lexico-semantic and syntactic processes (Fedorenko et al., 2011; Fedorenko et al., 2010; Menenti et al., 2011; Silbert et al., 2014), and a domain-general fronto-parietal network that supports executive processes, like attention, working memory, and cognitive control, sometimes referred to as the Multiple Demand (MD) network (e.g., Duncan, 2010b, 2013).

If linguistic prediction is localized to the domain-general MD network or to both the language and the MD network, that would give credence to accounts whereby prediction draws on executive resources. Given that linguistic surprisal has been linked to processing difficulty (e.g., Frank et al., 2015; Smith and Levy, 2013), such a finding would be consistent with the signature increase in the activity of MD brain regions for cognitively effortful tasks (Duncan and Owen, 2000; Fedorenko et al., 2013; Hugdahl et al., 2015; Peelle et al., 2010), and would provide further support for claims that this network encodes predictive signals across domains and relays them as feedback to other, domain-specific, brain regions (Chao et al., 2018; Cristescu et al., 2006; Egner et al., 2008; Strange et al., 2005; Strijkers et al., 2019; Wacongne et al., 2011). If, on the other hand, linguistic predictive processing is localized to the language-specific network, which plausibly stores our linguistic knowledge representations (Fedorenko, 2014; Fedorenko et al., 2018), that would suggest that prediction is carried out by mechanisms that selectively support language comprehension, and does not draw on executive resources. This result would be particularly (though not exclusively) consistent with prediction being ubiquitous within and across individuals, and would align with a growing body of cognitive neuroscience research supporting prediction as a "canonical computation" (Keller and Mrsic-Flogel, 2018) locally implemented in domain-specific circuits (Alink et al., 2010; Bastos et al., 2012; Bubic et al., 2010; Montague et al., 1996; Rao and Ballard, 1999; Singer et al., 2018; Wacongne et al., 2011). Because reliance of linguistic prediction on executive resources is a pillar argument for accounts whereby prediction is not obligatory, finding that linguistic prediction does not engage the MD network would present a fundamental challenge to those accounts as they are currently construed (Huettig and Mani, 2016; Pickering and Gambi, 2018). However, as laid out in Section 4, prediction varying across the lifespan of an individual, or between populations, can also be consistent with a language-experience-based account of linguistic prediction.

Only a handful of studies have used fMRI to probe prediction during language processing. So far, no clear and consistent picture has emerged, perhaps in part due to the variable ways in which prediction has been operationalized across studies. Some studies have actually failed to observe effects of linguistic predictability on brain activation (e.g., Schuster et al., 2016). Others have observed reliable effects of prediction, but the implicated brain regions have differed across studies (e.g., Lopopolo et al., 2017; Willems et al., 2016). For example, using a semantic priming paradigm, Weber et al. (2016) found that the facilitation effect was increased in situations of high predictive validity, and this effect was localized to left inferior frontal gyrus (IFG) and left posterior superior/middle temporal gyrus (post-S/MTG) – two important regions of the fronto-temporal language-selective network (e.g., Fedorenko and Thompson-Schill, 2014). Similarly, Matchin et al. (2017) found that IFG and posterior superior temporal sulcus (pSTS) were more strongly activated by sentences (e.g., the poet will recite a verse) than sequences of two-word phrases (e.g., the fencer the baby their bill) (see also Pallier et al., 2011, for the same empirical finding), and interpreted these results with respect to predictive processing. Some studies that have used naturalistic comprehension tasks have also implicated regions in the fronto-temporal language network. For example, Henderson et al. (2016) used a word-by-word reading task and localized surprisal effects to left IFG, but did not observe effects in posterior temporal cortex. In contrast, Willems et al. (2016) observed surprisal effects in posterior temporal cortex during a naturalistic story listening task, but not in left IFG. Using MEG, Eisenhauer et al. (2019) localized lexico-semantic prediction effects, measured by priming-based facilitation, to the left temporal pole. Others have implicated regions traditionally considered to belong to the fronto-parietal executive network in linguistic prediction. For example, functional (task-related) connectivity analyses have suggested a role for anterior cingulate cortex (ACC) – a region thought to be involved in monitoring changes in statistical contingencies between stimuli and stimulus-response mappings (Behrens et al., 2007; Botvinick et al., 2004). And Strijkers et al. (2019), based on an MEG finding of early sensitivity to a contextual linguistic manipulation, have argued for a role of a domain-general network in the prefrontal cortex in predictive top-down activation.

However, the bulk of this evidence does not clearly speak to the mechanism underlying linguistic prediction. First, artificial task-based paradigms often conflate multiple processes, which may in turn engage multiple distinct mechanisms simultaneously. For example, sentences and "Jabberwocky" sentences (Bonhage et al., 2015), or sentences and sequences of phrases (Matchin et al., 2017) differ in many ways other than differential prediction demands (e.g., semantic composition, memory demands). Leveraging a well-studied, formal operationalization of predictability (i.e., surprisal) may more directly inform our understanding of the neural basis of prediction (Brennan et al., 2016; Hale et al., 2018; Henderson et al., 2016; Shain et al., 2019). Second, most of these studies have used whole-brain analyses and/or anatomically defined regions of interest (e.g., Lopopolo et al., 2017; Willems et al., 2016), which preclude inferences about underlying processes due to the lack of a consistent mapping between function and anatomy across people, especially in the higher-order association cortex (e.g., Frost and Goebel, 2012; Vázquez-Rodríguez et al., 2019), and the resulting low functional resolution (e.g., Nieto-Castañón and Fedorenko, 2012). Sensitivity is also low in whole-brain analyses (e.g., Nieto-Castañón and Fedorenko, 2012; Saxe et al., 2006), with the consequence of a high chance of Type II error.

An alternative approach would be to use independent functional localizers (e.g., Brett et al., 2002; Saxe et al., 2006; Fedorenko et al., 2010) to target the language-specific and the domain-general MD networks, and then probe the responses of these functionally defined areas to a critical linguistic task where prediction demands vary between conditions or over time (in naturalistic stimuli). Recent neuroimaging evidence using this functional localization approach has suggested that the domain-general MD network is not critically engaged during typical language processing. For instance, during naturalistic reading of stories containing both high and low surprisal words and structures—which on most accounts should engage predictive processing—the neural signal in the language network closely "tracks" linguistic input, as evidenced by high inter-subject correlations (Hasson et al., 2008), whereas the neural signal in the MD network does not (Blank and Fedorenko, 2017; see also Lerner et al., 2011; for other evidence of the lack of the MD network's engagement in language processing under passive listening/reading conditions, see Campbell an Tyler, 2018; Diachek et al., 2019). Finally, Shain et al. (2019, this issue) specifically examined responses to predictive processing in the language and the MD networks in a large set of participants listening to naturalistic stories (Futrell et al., 2017). They found robust effects of predictability, controlling for unigram frequency, in the language network, but not in the MD network, arguing against the role of executive resources in linguistic prediction.

Multivariate analytic approaches in fMRI, MEG, EEG, and ECoG are likely to yield other promising avenues of investigation. For example, decoding methods (e.g., Haxby et al., 2001; Baker and van Gerven, 2018; Norman et al., 2006) allow the researcher to ask not only what stimuli/tasks elicit strong responses in a particular brain region but also what information is contained in the neural signal. This methodology can be fruitfully leveraged to understand how linguistic information is represented in the brain, including the language model in general, the particular contextual information used to generate predictions, and the predicted information. A number of studies have established that linguistic meanings, for both single words and sentences, can be reliably decoded from neural activity (Anderson et al., 2017; Huth et al., 2016; Kivisaari et al., 2019; Pereira et al., 2018). Wang et al. (2018) recently applied representational similarity analysis (RSA; Kriegeskorte et al., 2008) to show that the patterns of neural activity in MEG elicited by prediction of the same word based on sentential context—before that word is presented—are more similar than patterns corresponding to different predicted words. One caveat with this type of design is that neural similarity may reflect not the similarity in the meanings of the *predicted words* but rather in the similarity of the meanings of the preceding *context*, which is bound to correlate with the predicted target word. As brain decoding models are refined and become more generalizable (e.g., Pereira et al., 2018), researchers can begin to use them to track what mental constructs are most strongly represented at different time points during the processing of a linguistic input and what alternatives are being considered, providing a window into the prediction generation process.

6. Conclusions

The precise mechanisms of linguistic prediction deserve further investigation. In this piece, we have argued that the strongest evidence for the role of executive resources in linguistic prediction (i.e., an apparent reduction or absence of predictability effects in populations with lower executive resources; Pickering and Gambi, 2018) can be accounted for in terms of language-specific predictive mechanisms. However, the specific hypotheses outlined remain to be evaluated and tested against the executive-resource-based accounts in future behavioral, computational modeling, and brain imaging studies, as discussed in Section 5. With respect to the latter, we have suggested that brain imaging investigations can meaningfully inform the nature of linguistic predictive behavior, as long as we can interpret the loci of the observed effects as indexing particular mental computations/sets of computations (e.g., Mather et al., 2013). In particular, a functional localization approach where brain regions are targeted with independent tasks, thus avoiding the problem of reverse inference from anatomy to function (Poldrack, 2006, 2011), is promising for understanding prediction mechanisms. A recent study by Shain and colleagues (2019, this issue) took this approach and found robust evidence for prediction in the language-specific fronto-temporal network (e.g., Fedorenko et al., 2011), but not in the fronto-parietal network that has been linked to executive functions (Duncan, 2010b). These results align with a construal of linguistic prediction rooted in our experience with language, and provide one illustration of how functional brain imaging can inform human cognitive architecture (Mather et al., 2013). We have also suggested that decoding approaches applied to brain imaging data (e.g., Haxby et al., 2001; Baker and van Gerven, 2018; Norman et al., 2006) hold substantial promise for probing the representations that underlie linguistic predictive processing. These methods will both spur, and be informed by, the development of more precise theories of prediction. Elucidating the basic component processes (domain-general vs. language-specific computations) is a first step toward this end.

Acknowledgments

We would like to acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT, and its support team (Steve Shannon and Atsushi Takahashi). We also thank Craig Chambers and Jonathan Peelle for helpful discussions about some of the ideas presented here, and two anonymous reviewers for constructive comments and suggestions. RR was supported by NIH F32 DC015163. RPL acknowledges support from NSF-BCS 1551866 and a Newton Brain Science Research Seed Award. EF was supported by NIH awards R00-HD057522, R01-DC016607, and R01-DC016950, and by research support provided by the BCS Department at MIT and by McGovern Institute for Brain Research.

References

Alexander-Bloch, A., Giedd, J.N., Bullmore, E., 2013. Imaging structural co-variance between human brain regions. Nat. Rev. Neurosci. 14 (5), 322–336. https://doi.org/10.1038/nrn3465.

Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., Muckli, L., 2010. Stimulus predictability reduces responses in primary visual cortex. J. Neurosci. 30 (8), 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010.

Altmann, G.T.M., Kamide, Y., 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition 73 (3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1.

Ambridge, B., Pine, J.M., Rowland, C.F., Young, C.R., 2008. The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. Cognition 106 (1), 87–129. https://doi.org/10.1016/j.cognition.2006.12.015.

Anderson, A.J., Binder, J.R., Fernandino, L., Humphries, C.J., Conant, L.L., Aguilar, M., et al., 2017. Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. Cerebr. Cortex 27 (9), 4379–4395. https://doi.org/10.1093/cercor/bhw240.

Anderson, J.R., 1990. The Adaptive Character of Thought. https://doi.org/10.4324/9780203771730.

Baker, C.I., van Gerven, M., 2018. New advances in encoding and decoding of brain signals. Neuroimage 180, 1–3. https://doi.org/10.1016/j.neuroimage.2018.06.064.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J., 2012. Canonical microcircuits for predictive coding. Neuron 76 (4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038.

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., Rushworth, M.F.S., 2007. Learning the value of information in an uncertain world. Nat. Neurosci. 10 (9), 1214–1221. https://doi.org/10.1038/nn1954.

Bernard, A., Lubbers, L.S., Tanis, K.Q., Luo, R., Podtelezhnikov, A.A., Finney, E.M., et al., 2012. Transcriptional architecture of the primate neocortex. Neuron 73 (6), 1083–1099. https://doi.org/10.1016/j.neuron.2012.03.002.

Biber, D., Finegan, E., 1989. Drift and the evolution of English style: a history of three genres. Language 65 (3), 487–517. https://doi.org/10.2307/415220.

Blank, I.A., Fedorenko, E., 2017. Domain-general brain regions do not track linguistic input as closely as language-selective regions. J. Neurosci. 37 (41), 9999–10011. https://doi.org/10.1523/JNEUROSCI.3642-16.2017.

Bonhage, C.E., Mueller, J.L., Friederici, A.D., Fiebach, C.J., 2015. Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. Cortex 68, 33–47. https://doi.org/10.1016/j.cortex.2015.04.011.

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. J. Exp. Child Psychol. 112 (4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005.

Botvinick, M.M., Cohen, J.D., Carter, C.S., 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn. Sci. 8 (12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003.

Braginsky, M., Yurovsky, D., Marchman, V.A., Frank, M.C., 2019. Consistency and variability in children's word learning across languages. Open Mind 3, 52–67. https://doi.org/10.1162/opmi_a_00026.

Brett, M., Johnsrude, I., Owen, A., 2002. The problem of functional localization in the human brain. Nat. Rev. Neurosci 3 (243-249). https://doi.org/10.1038/nrn756.

Brennan, J.R., Stabler, E.P., Van Wagenen, S.E., Luh, W.-M., Hale, J.T., 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. Brain Lang. 157–158, 81–94. https://doi.org/10.1016/j.bandl.2016.04.008.

Bubic, A., Von Cramon, D.Y., Schubotz, R.I., 2010. Prediction, cognition and the brain. Front. Hum. Neurosci. 4 https://doi.org/10.3389/fnhum.2010.00025.

Campbell, K.L., Tyler, L.K., 2018. Language-related domain-specific and domain-general systems in the human brain. Curr. Opin. Behav. Sci. 21, 132–137. https://doi.org/10.1016/j.cobeha.2018.04.008.

Caplan, D., Waters, G.S., 1999. Verbal working memory and sentence comprehension. Behav. Brain Sci. 22 (1), 77–94. https://doi.org/10.1017/S0140525X99001788.

Chang, F., Dell, G.S., Bock, K., 2006. Becoming syntactic. Psychol. Rev. 113 (2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234.

Chang, F., Janciauskas, M., Fitz, H., 2012. Language adaptation and learning: Getting explicit about implicit learning: Language Adaptation. Lang. Linguist. Compass 6 (5), 259–278. https://doi.org/10.1002/lnc3.337.

Chao, Z.C., Takaura, K., Wang, L., Fujii, N., Dehaene, S., 2018. Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. Neuron 100 (5), 1252–1266. https://doi.org/10.1016/j.neuron.2018.10.004 e3.

Chen, C.-H., Gutierrez, E.D., Thompson, W., Panizzon, M.S., Jernigan, T.L., Eyler, L.T., et al., 2012. Hierarchical genetic organization of human cortical surface area. Science 335 (6076), 1634–1636. https://doi.org/10.1126/science.1215330.

Clark, Andy, 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36 (03), 1–73. https://doi.org/10.1017/S0140525X12002221.

Cristescu, T.C., Devlin, J.T., Nobre, A.C., 2006. Orienting attention to semantic categories. Neuroimage 33 (4), 1178–1187. https://doi.org/10.1016/j.neuroimage.2006.08.017.

Cuetos, F., Mitchell, D.C., Corley, M.M.B., 1996. Parsing in different languages. In: Language Processing in Spanish. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, pp. 145–187.

Dagerman, K.S., MacDonald, M.C., Harm, M.W., 2006. Aging and the use of context in ambiguity resolution: complex changes from simple slowing. Cogn. Sci. 30 (2), 311–345. https://doi.org/10.1207/s15516709cog0000_46.

Dambacher, M., Kliegl, R., Hofmann, M., Jacobs, A.M., 2006. Frequency and predictability effects on event-related potentials during reading. Brain Res. 1084 (1), 89–103. https://doi.org/10.1016/j.brainres.2006.02.010.

Dave, S., Brothers, T.A., Traxler, M.J., Ferreira, F., Henderson, J.M., Swaab, T.Y., 2018. Electrophysiological evidence for preserved primacy of lexical prediction in aging. Neuropsychologia 117, 135–147. https://doi.org/10.1016/j.neuropsychologia.2018.05.023.

Davidson, M.C., Amso, D., Anderson, L.C., Diamond, A., 2006. Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. Neuropsychologia 44 (11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006.

De Luca, C.R., Leventer, R., 2010, October 18. Developmental Trajectories of Executive Functions Across the Lifespan. https://doi.org/10.4324/9780203837863-12.

Dell, G.S., Chang, F., 2013. The P-chain: relating sentence production and its disorders to comprehension and acquisition. Philos. Trans. R. Soc. Biol. Sci. 369 (1634), 20120394. https://doi.org/10.1098/rstb.2012.0394.

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

Diachek, E., Blank, I., Siegelman, M., Fedorenko, E., 2019. The domain-general multiple demand (MD) network does not support core aspects of language comprehension: a large-scale fMRI investigation. BioRxiv, 744094. https://doi.org/10.1101/744094.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A.M., Kliegl, R., 2011. Coregistration of eye movements and EEG in natural reading: analyses and review. J. Exp. Psychol. Gen. 140 (4), 552–572. https://doi.org/10.1037/a0023885.

Duncan, J., 2010. How Intelligence Happens. Yale University Press.

Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn. Sci. 14 (4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004.

Duncan, J., 2013. The structure of cognition: attentional episodes in mind and brain. Neuron 80 (1), 35–50. https://doi.org/10.1016/j.neuron.2013.09.015.

Duncan, J., Owen, A.M., 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci. 23 (10), 475–483. https://doi.org/10.1016/S0166-2236(00)01633-7.

Dussias, P.E., Marful, A., Gerfen, C., Molina, M.T.B., 2010. Usage frequencies of complement-taking verbs in Spanish and English: data from Spanish monolinguals and Spanish—English bilinguals. Behav. Res. Methods 42 (4), 1004–1011. https://doi.org/10.3758/BRM.42.4.1004.

Egner, T., Monti, J.M.P., Trittschuh, E.H., Wieneke, C.A., Hirsch, J., Mesulam, M.-M., 2008. Neural integration of top-down spatial and feature-based information in visual search. J. Neurosci. 28 (24), 6141–6151. https://doi.org/10.1523/JNEUROSCI.1262-08.2008.

Ehrlich, S.F., Rayner, K., 1981. Contextual effects on word perception and eye movements during reading. J. Verb. Learn. Verb. Behav. 20 (6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6.

Elman, J.L., 1990. Finding Structure in Time. Cognit. Sci. 14 (2), 179–211. https://doi.org/10.1207/s15516709cog1402_1.

Eisenhauer, S., Fiebach, C.J., Gagl, B., 2019. Context-based facilitation in visual word recognition: evidence for visual and lexical but not pre-lexical contributions. Eneuro 6 (2). https://doi.org/10.1523/ENEURO.0321-18.2019. ENEURO.0321-18.2019.

Federmeier, K.D., 2007. Thinking ahead: the role and roots of prediction in language comprehension. Psychophysiology 44 (4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x.

Federmeier, K.D., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. J. Mem. Lang. 41 (4), 469–495. https://doi.org/10.1006/jmla.1999.2660.

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

Federmeier, K.D., Kutas, M., Schul, R., 2010. Age-related and individual differences in the use of prediction during language comprehension. Brain Lang. 115 (3), 149–161. https://doi.org/10.1016/j.bandl.2010.07.006.

Federmeier, K.D., McLennon, D.B., Ochoa, E.D., 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, vol. 14.

Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., Kutas, M., 2007. Multiple effects of sentential constraint on word processing. Brain Res. 1146, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101.

Fedorenko, E., 2014. The role of domain-general cognitive control in language comprehension. Front. Psychol. 5 https://doi.org/10.3389/fpsyg.2014.00335.

Fedorenko, E., Behr, M.K., Kanwisher, N., 2011. Functional specificity for high-level linguistic processing in the human brain. Proc. Natl. Acad. Sci. 108 (39), 16428–16433. https://doi.org/10.1073/pnas.1112937108.

Fedorenko, E., Duncan, J., Kanwisher, N., 2013. Broad domain generality in focal regions of frontal and parietal cortex. Proc. Natl. Acad. Sci. 110 (41), 16616–16621. https://doi.org/10.1073/pnas.1315235110.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. J. Neurophysiol. 104 (2), 1177–1194. https://doi.org/10.1152/jn.00032.2010.

Fedorenko, E., Mineroff, Z., Siegelman, M., Blank, I., 2018. Word meanings and sentence structure recruit the same set of fronto-temporal regions during comprehension. BioRxiv, 477851. https://doi.org/10.1101/477851.

Fedorenko, E., Thompson-Schill, S.L., 2014. Reworking the language network. Trends Cogn. Sci. 18 (3), 120–126. https://doi.org/10.1016/j.tics.2013.12.006.

Ferreira, F., Chantavarin, S., 2018. Integration and Prediction in Language Processing: A Synthesis of Old and New. Current Directions in Psychological Science, 0963721418794491. https://doi.org/10.1177/0963721418794491.

Fine, A.B., Jaeger, T.F., Farmer, T.A., Qian, T., 2013. Rapid expectation adaptation during syntactic comprehension. PLoS One 8 (10), e77661. https://doi.org/10.1371/journal.pone.0077661.

Fitz, H., Chang, F., 2019. Language ERPs reflect learning through prediction error propagation. Cogn. Psychol. 111, 15–52. https://doi.org/10.1016/j.cogpsych.2019.03.002.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Essen, D.C.V., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. 102 (27), 9673–9678. https://doi.org/10.1073/pnas.0504136102.

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 Lang. 140, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006.

Friedman, N.P., Miyake, A., 2017. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. Cortex 86, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023.

Friedrich, M., Friederici, A.D., 2005. Phonotactic knowledge and lexical-semantic processing in one-year-olds: brain responses to words and nonsense words in picture contexts. J. Cogn. Neurosci. 17 (11), 1785–1802. https://doi.org/10.1162/089892905774589172.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11 (2), 127–138. https://doi.org/10.1038/nrn2787.

Frost, M.A., Goebel, R., 2012. Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. Neuroimage 59 (2), 1369–1381. https://doi.org/10.1016/j.neuroimage.2011.08.035.

Futrell, R., Gibson, E., Tily, H.J., Blank, I., Vishnevetsky, A., Piantadosi, S.T., Fedorenko, E., 2017. The Natural Stories Corpus. LREC, pp. 76–82.

Futrell, R., Levy, R., 2017. Noisy-context surprisal as a human sentence processing cost model. Long Papers. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, vol. 1, pp. 688–698. Retrieved from. https://www.aclweb.org/anthology/E17-1065.

Gambi, C., Gorrie, F., Pickering, M.J., Rabagliati, H., 2018. The development of linguistic prediction: predictions of sound and meaning in 2- to 5-year-olds. J. Exp. Child Psychol. 173, 351–370. https://doi.org/10.1016/j.jecp.2018.04.012.

Garnsey, S.M., Pearlmutter, N.J., Myers, E., Lotocky, M.A., 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. J. Mem. Lang. 37 (1), 58–93. https://doi.org/10.1006/jmla.1997.2512.

Goodman, J.C., Dale, P.S., Li, P., 2008. Does frequency count? Parental input and the acquisition of vocabulary. J. Child Lang. 35 (3), 515–531. https://doi.org/10.1017/S0305000907008641.

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

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. PLoS Biol. 6 (7), e159. https://doi.org/10.1371/journal.pbio.0060159.

Hale, J., 2001. A Probabilistic Earley Parser as a Psycholinguistic Model, pp. 1–8. https://doi.org/10.3115/1073336.1073357.

Hale, J., Dyer, C., Kuncoro, A., Brennan, J.R., 2018. Finding Syntax in Human Encephalography with Beam Search [Cs]. Retrieved from. ArXiv:1806.04127. http://arxiv.org/abs/1806.04127.

Hamberger, M.J., Friedman, D., Rosen, J., 1996. Completion norms collected from younger and older adults for 198 sentence contexts. Behav. Res. Methods Instrum. Comput. 28 (1), 102–108. https://doi.org/10.3758/BF03203644.

Hartshorne, J.K., Germine, L.T., 2015. When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. Psychol. Sci. 26 (4), 433–443. https://doi.org/10.1177/0956797614567339.

Hasson, U., Egidi, G., Marelli, M., Willems, R.M., 2018. Grounding the neurobiology of language in first principles: the necessity of non-language-centric explanations for language comprehension. Cognition 180, 135–157. https://doi.org/10.1016/j.cognition.2018.06.018.

Hasson, U., Furman, O., Clark, D., Dudai, Y., Davachi, L., 2008. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57 (3), 452–462. https://doi.org/10.1016/j.neuron.2007.12.009.

Hauk, O., 2016. Preface to special issue "prediction in language comprehension and production. Lang., Cogn. Neurosci. 31 (1), 1–3. https://doi.org/10.1080/23273798.2015.1102300.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430. https://doi.org/10.1126/science.1063736.

Heller, D., Grodner, D., Tanenhaus, M.K., 2008. The Role of Perspective in Identifying Domains of Reference Q, vol. 6.

Henderson, J.M., Choi, W., Lowder, M.W., Ferreira, F., 2016. Language structure in the brain: a fixation-related fMRI study of syntactic surprisal in reading. Neuroimage 132, 293–300. https://doi.org/10.1016/j.neuroimage.2016.02.050.

Huettig, F., Mani, N., 2016. Is prediction necessary to understand language? Probably not. Lang., Cogn. Neurosci. 31 (1), 19–31. https://doi.org/10.1080/23273798.2015.1072223.

Hugdahl, K., Raichle, M.E., Mitra, A., Specht, K., 2015. On the existence of a generalized non-specific task-dependent network. Front. Hum. Neurosci. 9 https://doi.org/10.3389/fnhum.2015.00430.

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532 (7600), 453–458. https://doi.org/10.1038/nature17637.

James, A.N., Fraundorf, S.H., Lee, E.-K., Watson, D.G., 2018. Individual differences in syntactic processing: is there evidence for reader-text interactions? J. Mem. Lang. 102, 155–181. https://doi.org/10.1016/j.jml.2018.05.006.

James, W., 1893. Psychology. Henry Holt.

Johnson, R.E., 2003. Aging and the remembering of text. Dev. Rev. 23 (3), 261–346. https://doi.org/10.1016/S0273-2297(03)00009-1.

Just, M.A., Carpenter, P.A., 1992. A capacity theory of comprehension: individual differences in working memory. Psychol. Rev. 99 (1), 122–149.

Kaan, E., 2014. Predictive sentence processing in L2 and L1: what is different? Linguist. Approaches Biling. 4 (2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa.

Kahneman, D., 1973. Attention and Effort. Prentice-Hall, New Jersey.

Kamide, Y., Altmann, G.T.M., Haywood, S.L., 2003. The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. J. Mem. Lang. 49 (1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8.

Keller, G.B., Mrsic-Flogel, T.D., 2018. Predictive processing: a canonical cortical computation. Neuron 100 (2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003.

Kim, A.E., Oines, L., Miyake, A., 2018. Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: an ERP investigation. J. Exp. Psychol. Learn. Mem. Cogn. 44 (3), 406–420. https://doi.org/10.1037/xlm0000457.

King, J., Just, M.A., 1991. Individual differences in syntactic processing: the role of working memory. J. Mem. Lang. 30 (5), 580–602. https://doi.org/10.1016/0749-596X(91)90027-H.

Kivisaari, S.L., van Vliet, M., Hulten, A., Lindh-Knuutila, T., Faisal, A., Salmelin, R., 2019. Reconstructing meaning from bits of information. Nat. Commun. 10 (1), 927.

Konopka, G., Friedrich, T., Davis-Turak, J., Winden, K., Oldham, M.C., Gao, F., et al., 2012. Human-specific transcriptional networks in the brain. Neuron 75 (4), 601–617. https://doi.org/10.1016/j.neuron.2012.05.034.

Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis—connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2 https://doi.org/10.3389/neuro.06.004.2008.

Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction in language comprehension? Lang., Cogn. Neurosci. 31 (1), 32–59. https://doi.org/10.1080/23273798.2015.1102299.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62 (1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.

Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307 (5947), 161. https://doi.org/10.1038/307161a0.

Lahar, C.J., Tun, P.A., Wingfield, A., 2004. Sentence–final word completion norms for young, middle-aged, and older adults. J. Gerontol.: Ser. Bibliogr. 59 (1), P7–P10. https://doi.org/10.1093/geronb/59.1.P7.

Lerner, Y., Honey, C.J., Silbert, L.J., Hasson, U., 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J. Neurosci. 31 (8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011.

Levy, R., 2008. Expectation-based syntactic comprehension. Cognition 106 (3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006.

Lew-Williams, C., Fernald, A., 2010. Real-time processing of gender-marked articles by native and non-native Spanish speakers. J. Mem. Lang. 63 (4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003.

Linck, J.A., Osthus, P., Koeth, J.T., Bunting, M.F., 2014. Working memory and second language comprehension and production: a meta-analysis. Psychon. Bull. Rev. 21 (4), 861–883. https://doi.org/10.3758/s13423-013-0565-2.

Lopopolo, A., Frank, S.L., van den Bosch, A., Willems, R.M., 2017. Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain. PLoS One 12 (5), e0177794. https://doi.org/10.1371/journal.pone.0177794.

Luke, S.G., Christianson, K., 2016. Limits on lexical prediction during reading. Cogn. Psychol. 88, 22–60. https://doi.org/10.1016/j.cogpsych.2016.06.002.

Lupyan, G., Clark, A., 2015. Words and the world: predictive coding and the language-Perception-Cognition interface. Curr. Dir. Psychol. Sci. 24 (4), 279–284. https://doi.org/10.1177/0963721415570732.

MacDonald, M.C., 2013. How language production shapes language form and comprehension. Front. Psychol. 4 https://doi.org/10.3389/fpsyg.2013.00226.

Mani, N., Huettig, F., 2012. Prediction during language processing is a piece of cake—but only for skilled producers. J. Exp. Psychol. Hum. Percept. Perform. 38 (4), 843–847. https://doi.org/10.1037/a0029284.

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

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

Matchin, W., Hammerly, C., Lau, E., 2017. The role of the IFG and pSTS in syntactic prediction: evidence from a parametric study of hierarchical structure in fMRI. Cortex 88, 106–123. https://doi.org/10.1016/j.cortex.2016.12.010.

Mather, M., Cacioppo, J.T., Kanwisher, N., 2013. How fMRI can inform cognitive theories. Perspect. Psychol. Sci. 8 (1), 108–113. https://doi.org/10.1177/1745691612469037.

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. Psychol. Sci. 22 (9), 1173–1182. https://doi.org/10.1177/0956797611418347.

Mirman, D., Yee, E., Blumstein, S.E., Magnuson, J.S., 2011. Theories of spoken word recognition deficits in Aphasia: evidence from eye-tracking and computational modeling. Brain Lang. 117 (2), 53–68. https://doi.org/10.1016/j.bandl.2011.01.004.

Misyak, J.B., Christiansen, M.H., 2012. Statistical learning and language: an individual differences study. Lang. Learn. 62 (1), 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x.

Mitsugi, S., Macwhinney, B., 2016. The use of case marking for predictive processing in second language Japanese*. Biling. Lang. Cognit. 19 (1), 19–35. https://doi.org/10.1017/S1366728914000881.

Montag, J.L., Jones, M.N., Smith, L.B., 2015. The words children hear: picture books and the statistics for language learning. Psychol. Sci. 26 (9), 1489–1496. https://doi.org/10.1177/0956797615594361.

Montag, J.L., MacDonald, M.C., 2015. Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. J. Exp. Psychol. Gen. 144 (2), 447–468. https://doi.org/10.1037/xge0000054.

Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. 16 (5), 1936–1947. https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996.

Muraven, M., Baumeister, R.F., 2000. Self-regulation and depletion of limited resources: does self-control resemble a muscle? Psychol. Bull. 126 (2), 247–259. https://doi.org/10.1037/0033-2909.126.2.247.

Myung, I.J., 2000. The importance of complexity in model selection. J. Math. Psychol. 44 (1), 190–204. https://doi.org/10.1006/jmps.1999.1283.

Nakano, H., Saron, C., Swaab, T.Y., 2009. Speech and span: working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. J. Cogn. Neurosci. 22 (12), 2886–2898. https://doi.org/10.1162/jocn.2009.21400.

Nicenboim, B., Vasishth, S., Rösler, F., 2019. Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly. https://doi.org/10.31234/osf.io/2atrh available data [Preprint].

Nieto-Castañón, A., Fedorenko, E., 2012. Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. Neuroimage 63 (3), 1646–1669. https://doi.org/10.1016/j.neuroimage.2012.06.065.

Nieuwland, M., 2019. Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. Neurosci. Biobehav. Rev. 96, 367–400. https://doi.org/10.1016/j.neubiorev.2018.11.019.

Nieuwland, M., Barr, D., Bartolozzi, F., Busch-Moreno, S., Donaldson, D., Ferguson, H.J., et al., 2019. Dissociable Effects of Prediction and Integration during Language Comprehension: Evidence from a Large-Scale Study Using Brain Potentials. Philosophical Transactions of the Royal Society B: Biological Sciences. https://doi.org/10.1101/267815. Retrieved from.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10 (9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005.

Novick, J.M., Trueswell, J.C., Thompson-Schill, S.L., 2005. Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. Cogni. Affect. Behav. Neurosci 5 (3), 263–281.

Nozari, N., Mirman, D., Thompson-Schill, S.L., 2016. The ventrolateral prefrontal cortex facilitates processing of sentential context to locate referents. Brain Lang. 157–158, 1–13. https://doi.org/10.1016/j.bandl.2016.04.006.

Nozari, N., Trueswell, J.C., Thompson-Schill, S.L., 2016. The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. Psychon. Bull. Rev. 23 (6), 1942–1953. https://doi.org/10.3758/s13423-016-1068-8.

Pallier, C., Devauchelle, A.-D., Dehaene, S., 2011. Cortical representation of the constituent structure of sentences. Proc. Natl. Acad. Sci. 108 (6), 2522–2527. https://doi.org/10.1073/pnas.1018711108.

Park, D.C., Lautenschlager, G., Hedden, T., Davidson, N.S., Smith, A.D., Smith, P.K., 2002. Models of visuospatial and verbal memory across the adult life span. Psychol. Aging 17 (2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299.

Payne, B.R., Federmeier, K.D., 2018. Contextual constraints on lexico-semantic processing in aging: evidence from single-word event-related brain potentials. Brain Res. 1687, 117–128. https://doi.org/10.1016/j.brainres.2018.02.021.

Payne, B.R., Grison, S., Gao, X., Christianson, K., Morrow, D.G., Stine-Morrow, E.A.L., 2014. Aging and individual differences in binding during sentence understanding: evidence from temporary and global syntactic attachment ambiguities. Cognition 130 (2), 157–173. https://doi.org/10.1016/j.cognition.2013.10.005.

Payne, B.R., Silcox, J.W., 2019. Chapter Seven—aging, context processing, and comprehension. In: Federmeier, K.D. (Ed.), Psychology of Learning and Motivation, vol. 71, pp. 215–264. https://doi.org/10.1016/bs.plm.2019.07.001.

Peelle, J.E., Troiani, V., Wingfield, A., Grossman, M., 2010. Neural processing during older adults' comprehension of spoken sentences: age differences in resource allocation and connectivity. Cerebr. Cortex 20 (4), 773–782. https://doi.org/10.1093/cercor/bhp142.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S.J., Kanwisher, N., et al., 2018. Toward a universal decoder of linguistic meaning from brain activation. Nat. Commun. 9 (1) https://doi.org/10.1038/s41467-018-03068-4.

Pickering, M.J., Gambi, C., 2018. Predicting while comprehending language: a theory and review. Psychol. Bull. 144 (10), 1002–1044. https://doi.org/10.1037/bul0000158.

Pickering, M.J., Garrod, S., 2013. An integrated theory of language production and comprehension. Behav. Brain Sci. 36 (04), 329–347. https://doi.org/10.1017/S0140525X12001495.
Pinker, S., 2009. Language Learnability and Language Development, with New Commentary by the Author: with New Commentary by the Author. Harvard University Press.
Poldrack, R.A., 2006. Can cognitive processes be inferred from neuroimaging data? Trends Cogn. Sci. 10 (2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004.
Poldrack, R.A., 2011. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron 72 (5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., et al., 2011. Functional network organization of the human brain. Neuron 72 (4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006.
Rabagliati, H., Gambi, C., Pickering, M.J., 2016. Learning to predict or predicting to learn? Lang., Cogn. Neurosci. 31 (1), 94–105. https://doi.org/10.1080/23273798.2015.1077979.
Rabovsky, M., Hansen, S.S., McClelland, J.L., 2018. Modelling the N400 brain potential as change in a probabilistic representation of meaning. Nature Human Behav. 2 (9), 693. https://doi.org/10.1038/s41562-018-0406-4.
Ramscar, M., Dye, M., McCauley, S.M., 2013. Error and expectation in language learning: the curious absence of "mouses" in adult speech. Language 89 (4), 760–793.
Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., Baayen, H., 2014. The myth of cognitive decline: non-linear dynamics of lifelong learning. Top. Cogn. Sci. 6 (1), 5–42. https://doi.org/10.1111/tops.12078.
Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2 (1), 79–87. https://doi.org/10.1038/4580.
Raznahan, A., Lerch, J.P., Lee, N., Greenstein, D., Wallace, G.L., Stockman, M., et al., 2011. Patterns of coordinated anatomical change in human cortical development: a longitudinal neuroimaging study of maturational coupling. Neuron 72 (5), 873–884. https://doi.org/10.1016/j.neuron.2011.09.028.
Ryskin, R.A., Qi, Z., Duff, M.C., Brown-Schmidt, S., 2017. Verb biases are shaped through lifelong learning. J. Exp. Psychol. Learn. Mem. Cogn. 43 (5), 781–794. https://doi.org/10.1037/xlm0000341.
Salthouse, T.A., 2009. When does age-related cognitive decline begin? Neurobiol. Aging 30 (4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023.
Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. Neuroimage 30 (4), 1088–1096. https://doi.org/10.1016/j.neuroimage.2005.12.062.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275 (5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593.
Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., Richlan, F., 2016. Words in context: the effects of length, frequency, and predictability on brain responses during natural reading. Cerebr. Cortex 26 (10), 3889–3904. https://doi.org/10.1093/cercor/bhw184.
Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L., Greicius, M.D., 2009. Neurodegenerative diseases target large-scale human brain networks. Neuron 62 (1), 42–52. https://doi.org/10.1016/j.neuron.2009.03.024.
Shafto, M.A., Tyler, L.K., 2014. Language in the aging brain: the network dynamics of cognitive decline and preservation. Science 346 (6209), 583–587. https://doi.org/10.1126/science.1254404.
Shain, C., Blank, I., van Schijndel, M., Schuler, W., Fedorenko, E., 2019. FMRI Reveals Language-specific Predictive Coding during Naturalistic Sentence Comprehension. CUNY Conference on Human Sentence Processing, Boulder, CO.
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. Proc. Natl. Acad. Sci. 111 (43), E4687–E4696. https://doi.org/10.1073/pnas.1323812111.
Singer, Y., Teramoto, Y., Willmore, B.D., Schnupp, J.W., King, A.J., Harper, N.S., 2018. Sensory cortex is optimized for prediction of future input. ELife 7. https://doi.org/10.7554/eLife.31557.
Smith, N.J., Levy, R., 2013. The effect of word predictability on reading time is logarithmic. Cognition 128 (3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013.
Spearman, C., 1929. The uniqueness of "G.". J. Educ. Psychol. 20 (3), 212–216. https://doi.org/10.1037/h0072998.
Stauffer, W.R., Lak, A., Schultz, W., 2014. Dopamine reward prediction error responses reflect marginal utility. Curr. Biol. 24 (21), 2491–2500. https://doi.org/10.1016/j.cub.2014.08.064.
Steinberg, E.E., Keiflin, R., Boivin, J.R., Witten, I.B., Deisseroth, K., Janak, P.H., 2013. A causal link between prediction errors, dopamine neurons and learning. Nat. Neurosci. 16 (7), 966–973. https://doi.org/10.1038/nn.3413.
Strange, B.A., Duggins, A., Penny, W., Dolan, R.J., Friston, K.J., 2005. Information theory, novelty and hippocampal responses: unpredicted or unpredictable? Neural Netw. 18 (3), 225–230. https://doi.org/10.1016/j.neunet.2004.12.004.
Strijkers, K., Chanoine, V., Munding, D., Dubarry, A.-S., Trébuchon, A., Badier, J.-M., Alario, F.-X., 2019. Grammatical class modulates the (left) inferior frontal gyrus within 100 milliseconds when syntactic context is predictive. Sci. Rep. 9 (1), 4830. https://doi.org/10.1038/s41598-019-41376-x.
Swets, B., Desmet, T., Hambrick, D.Z., Ferreira, F., 2007. The role of working memory in syntactic ambiguity resolution: a psychometric approach. J. Exp. Psychol. Gen. 136 (1), 64–81. https://doi.org/10.1037/0096-3445.136.1.64.
Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C., 1995. Integration of visual and linguistic information in spoken language comprehension. Science 268 (5217), 1632–1634. https://doi.org/10.1126/science.7777863.
Tanner, D., Van Hell, J.G., 2014. ERPs reveal individual differences in morphosyntactic processing. Neuropsychologia 56, 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002.
Toro, R., Fox, P.T., Paus, T., 2008. Functional coactivation map of the human brain. Cerebr. Cortex 18 (11), 2553–2559. https://doi.org/10.1093/cercor/bhn014.
Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. J. Exp. Psychol. Learn. Mem. Cogn. 31 (3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443.
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. J. Cogn. Neurosci. 20 (4), 580–591. https://doi.org/10.1162/jocn.2008.20054.
van den Heuvel, M.P., Sporns, O., 2011. Rich-club organization of the human connectome. J. Neurosci. 31 (44), 15775–15786. https://doi.org/10.1523/JNEUROSCI.3539-11.2011.
Van Dyke, J.A., Johns, C.L., Kukona, A., 2014. Low working memory capacity is only spuriously related to poor reading comprehension. Cognition 131 (3), 373–403. https://doi.org/10.1016/j.cognition.2014.01.007.
Van Petten, C., Luka, B.J., 2012. Prediction during language comprehension: benefits, costs, and ERP components. Int. J. Psychophysiol. 83 (2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015.
Vázquez-Rodríguez, B., Suárez, L.E., Markello, R.D., Shafiei, G., Paquola, C., Hagmann, P., et al., 2019. Gradients of structure–function tethering across neocortex. Proc. Natl. Acad. Sci. 116 (42), 21219–21227. https://doi.org/10.1073/pnas.1903403116.
Verhaeghen, P., 2003. Aging and vocabulary score: a meta-analysis. Psychol. Aging 18 (2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332.
Verhagen, V., Mos, M., Backus, A., Schilperoord, J., 2018. Predictive language processing revealing usage-based variation. Lang. Cogn. 10 (2), 329–373. https://doi.org/10.1017/langcog.2018.4.
Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., Dehaene, S., 2011. Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc. Natl. Acad. Sci. 108 (51), 20754–20759. https://doi.org/10.1073/pnas.1117807108.
Wang, L., Kuperberg, G., Jensen, O., 2018. Specific lexico-semantic predictions are associated with unique spatial and temporal patterns of neural activity. ELife 7, e39061. https://doi.org/10.7554/eLife.39061.
Weber, K., Lau, E.F., Stillerman, B., Kuperberg, G.R., 2016. The yin and the yang of prediction: an fMRI study of semantic predictive processing. PLoS One 11 (3), e0148637. https://doi.org/10.1371/journal.pone.0148637.
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. J. Cogn. Neurosci. 16 (7), 1272–1288. https://doi.org/10.1162/0898929041920487.
Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., van den Bosch, A., 2016. Prediction during natural language comprehension. Cerebr. Cortex 26 (6), 2506–2516. https://doi.org/10.1093/cercor/bhv075.
Wlotko, E.W., Federmeier, K.D., 2012. Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension: context use in young and older adults. Psychophysiology 49 (6), 770–785. https://doi.org/10.1111/j.1469-8986.2012.01366.x.
Wolk, C., Bresnan, J., Rosenbach, A., Szmrecsanyi, B., 2013. Dative and genitive variability in Late Modern English: exploring cross-constructional variation and change. Diachronica 30 (3), 382–419. https://doi.org/10.1075/dia.30.3.04wol.
Wulff, D.U., De Deyne, S., Jones, M.N., Mata, R., 2019. New perspectives on the aging lexicon. Trends Cogn. Sci. 23 (8), 686–698. https://doi.org/10.1016/j.tics.2019.05.003.
Xu, X., Demos, K.E., Leahey, T.M., Hart, C.N., Trautvetter, J., Coward, P., et al., 2014. Failure to replicate depletion of self-control. PLoS One 9 (10), e109950. https://doi.org/10.1371/journal.pone.0109950.
Yan, S., Kuperberg, G.R., Jaeger, T.F., 2017. Prediction (or not) during language processing. A commentary on Nieuwland et al. (2017) and DeLong et al. (2005). BioRxiv, 143750. https://doi.org/10.1101/143750.
Yeo, T.B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (3), 1125–1165. https://doi.org/10.1152/jn.00338.2011.
Ylinen, S., Bosseler, A., Junttila, K., Huotilainen, M., 2017. Predictive coding accelerates word recognition and learning in the early stages of language development. Dev. Sci. 20 (6), e12472 https://doi.org/10.1111/desc.12472.