THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Learning to predict or predicting to learn?

OPEN ACCESS

Learning to predict or predicting to learn?

Hugh Rabagliati, Chiara Gambi and Martin J. Pickering

Department of Psychology, University of Edinburgh

Please address correspondence to:

Hugh Rabagliati

Department of Psychology

University of Edinburgh

Edinburgh

United Kingdom

hugh.rabagliati@ed.ac.uk

Contact details for other authors:

cgambi@exseed.ed.ac.uk

martin.pickering@ed.ac.uk

Abstract

Humans complete complex commonplace tasks, such as understanding sentences, with striking speed and accuracy. This expertise is dependent on anticipation: Predicting upcoming words gets us ahead of the game. But how do we master the game in the first place? To make accurate predictions, children must first learn their language. One possibility is that prediction serves double duty, enabling rapid language learning as well as understanding. Children could master the structures of their language by predicting how speakers will behave and, when those guesses are wrong, revising their linguistic representations. A number of prominent computational models assume that children learn in this way. But is that assumption correct? Here, we lay out the requirements for showing that children use "predictive learning", and review the current evidence for this position. We argue that, despite widespread enthusiasm for the idea, we cannot yet conclude that children "predict to learn".

147 words

Learning to predict or predicting to learn?

Human interactions require a well-timed interplay of behaviours that we seem to accomplish effortlessly, despite the complex and dynamic nature of our environment. When dancers move in pairs, when musicians improvise in a group, and when speakers converse in a pub, each individual has to rapidly make sense of the environment around them and then respond in an appropriate manner. For conversation, these difficulties are particularly acute. In order to keep an interaction going, listeners must make sense of what they hear in terms of its underlying syntax, its semantics and its contribution to the discourse. Then, they have to make a considered response, running the process in reverse, from high-level discourse goals all the way down to motor commands.

Adult humans manage to accomplish these complex processes with ease, which raises two questions. First, what are the mental mechanisms that allow adults to accurately and effortlessly converse? Second, how do these mechanisms develop? That is to say, how do children, who start off innocent of their conversational environment, become savvy, swift interlocutors? A large number of theories have recently converged on the idea that both of these questions have the same answer: Prediction (Bar, 2007; Clark, 2013; Elman, 1990).

Prediction has become key to a number of successful cognitive models of

language comprehension and production (Chang, Dell, & Bock, 2006; Hale, 2001; Levy, 2007; Pickering & Garrod, 2013). Accurate expectations about the world allow us to quickly make sense of incoming stimuli and respond in appropriate ways: If we can guess what others will say before they say it, then we can reduce the computational burden of quickly interpreting their utterance, and produce preliminary plans for what we should say ourselves (Kutas, DeLong, & Smith, 2011; Pickering & Garrod, 2013). This sort of moment-by-moment prediction seems to be pervasive in the mechanisms that we use to understand and produce language (Kutas et al., 2011; Pickering & Garrod, 2013).

In contrast, research on how children acquire the ability to use language has traditionally either ignored or eschewed prediction, on the assumption that learning to predict requires knowing the language in the first place (Pinker, 1984, 1989; Tomasello, 1992, 2003). This is part of a deeper disconnect, in which theories of language learning rarely take account of the moment-by-moment processes by which children actually understand language (Bowerman, 1987). Instead, linguistic development is typically conceived of either as a process of offline hypothesis testing (Pinker, 1995; Wexler & Culicover, 1980) or as fitting a grammar to a large stock of memorized data (Perfors, Tenenbaum, & Regier, 2011).

Is this separation between mechanisms for processing and mechanisms for learning justified? An increasing number of theories and models have argued that it is not (Chang et al., 2006; Elman, 1990; Seidenberg & MacDonald, 1999). Instead, children are supposed to be predictive learners, generating expectations based on the limited grammar that they have learned so far, and then adjusting that grammar when their predictions fail. This possibility cuts to a fundamental question in cognitive science: How much continuity is there between the cognitive architectures found in adults and children? That is, are there common mechanisms for learning about and processing the world around us? In particular, is prediction an ability that emerges only in expert systems, or is it a mechanism that enables the system to become expert in the first place?

This paper attempts to evaluate the current evidence for these two possibilities. Predictive theories make certain assumptions about children's ability to make sense of language on a moment-by-moment basis, and we will compare these assumptions to what we currently know about online language processing in young children. To foreshadow, we will argue that the evidence is not there yet: what we know about the way children process language may point in the direction suggested by these theories, but there is still room for reasonable doubt. In the final part of this paper we will discuss how these models might be extended from explaining how children learn simple sentences, to accounting for

children's developing ability to take part in complex, coordinated language use –
in other words, conversation.

Prediction-based learning

The idea that prediction might help with language learning is not new, but it has
been controversial. For instance, learning via prediction has often been
suggested as a solution to the famed "no negative evidence problem", in which
children can "unlearn" ungrammatical forms (e.g., *mouses, don't giggle me*)
despite not receiving explicit corrections (so-called negative evidence). As one
example, Chomsky (1981, p.9) suggested that prediction-driven learning could
serve as a form of "implicit" negative evidence: "if certain structures or rules fail to
be exemplified in relatively simple expressions, where they would expect to be
found, than a (possibly marked) option is selected excluding them in the
grammar." That is to say, children can make guesses about exactly how
meanings should be expressed in their language, and then update their grammar
based on whether or not their prediction is correct.

Still, many of the most prominent models of language learning have either
ignored or explicitly disavowed this type of implicit feedback. For example,
Pinker's (1984) theory of language acquisition, which remains one of the most
complete and well-specified models that we have, explicitly states that predictive

learning is not necessary for overcoming the learnability problems associated with the lack of negative evidence. Instead, he suggests that children gradually learn to exclude ungrammatical wordforms or sentences based on a Unique Entry Principle, according to which no two forms (e.g., *mouses* and *mice*) will have exactly the same meaning and occupy exactly the same slot in an inflectional or derivational paradigm. Therefore, a child who has encountered *mice* a sufficient number of times will eventually learn that the form *mouses* should not be part of the grammar, as *mice* already provides a plural form for *mouse*. Constraint-satisfaction models (MacWhinney, 1987, 2004) assume that children can recover from overgeneralization because the two alternative forms compete with one another; the adult form is strengthened and eventually "wins over" the regular form simply because it recurs more often in the input. In both types of model children do not need to predict the wrong form to learn the correct one, but it is sufficient for them to comprehend the correct form and realize it has the same meaning as the incorrect one.

A number of other prominent models, such as Tomasello's Verb Island hypothesis (Tomasello, 1992, 2003), also tend to elide the notion of prediction from discussion. Under Tomasello's hypothesis (building on classic ideas by Braine, 1963) the process of language acquisition is a process of abstraction. Children gather large amounts of linguistic data, and then generate a grammar through comparison and analogy across sentences. What models such as

Pinker's and Tomasello's have in common is the assumption that learning occurs off-line (Bowerman, 1987). These models largely ignore what happens while the child attempts to understand or produce a sentence (i.e., during on-line processing). Instead, they are interested in the end result of these processes (e.g., does the child understand the sentence or not? Is the child able to identify lexical items as nouns and verbs?). This is because it is the end result that matters for learning. In a sense, these models conceive of language acquisition as a process occurring at a rather slow pace, with the child's internal grammar changing only when a considerable amount of evidence accumulates against the child's hypotheses about the structure of the language.

Models of predictive learning are very different. In these models the learning mechanism is not based on feature checking or abstraction, but on prediction error. In particular, children can use their current linguistic knowledge to generate predictions about what they will hear next. When these predictions are checked against reality, the resulting discrepancy (the error signal) can be used to update the child's linguistic knowledge, and to enhance future predictions. This is similar in spirit to Chomsky's ideas about implicit negative evidence, but on a much grander scale. Perhaps more importantly, these models make explicit connections between processing and learning. Every time the child processes a sentence, prediction error can be used to tune the child's ability to make successful predictions later on, and therefore every instance of processing is also

a learning instance. Learning therefore takes place on-line (Bowerman, 1987), rather than after the fact.

The actual mechanics of predictive learning vary across different models. The locus classicus for these theories is Elman's (Elman, 1990, 1993) application of a simple recurrent network to the problem of learning simple languages. Briefly, Elman developed and trained a set of simple neural networks that were constructed from a layer of input units, a layer of output units, a hidden layer between the input and output units, and (unusually) a context layer, which received input from the hidden layer and returned its output back to the hidden layer at the next time-point. As such, the context units acted as a form of memory. This allowed the network to learn the structure of sequential input. For instance, Elman provided it with words strung together in simple English-like sentences (*man eat cookie*, *woman smell rock*). The network was trained to predict which word would be said next; during training, this predicted word (the model's output) was compared to the actual next word in the sequence and the resulting error signal was used to adjust connections between the network's units through backpropagation (i.e., the network used prediction-driven learning). Learning to predict words incrementally allowed the network to perform a distributional analysis (i.e., learning co-occurrences between different words), which allowed it to learn representations in its hidden units that, to some degree, are similar to distributionally-defined categories such as noun or intransitive verb (although

they are not as abstract, Chang, 2002; Marcus, 2001). In doing so, it became able to accurately predict the types of words that would likely follow one another (e.g., transitive verbs are followed by nouns).

Later work has extended Elman's results in a number of ways, for example by defining more complex networks that are built from more complex primitive elements. Most prominently, Chang, Dell and Bock's (2006) connectionist model assumes that, during language comprehension, children predict which word they will hear next using not only a representation of the previously heard word, but also two additional pathways containing two types of representation. One pathway contains semantic representations that specify relationships between events (e.g., eating) and lexical concepts (e.g., cake), including the assignment of lexical concepts to thematic roles (e.g., cake → theme). This allows the network to guess the intended message of the speaker, and use that to generate predictions. Another pathway contains a representation of how linguistic items should be sequenced (i.e., how words are ordered in a syntactically correct manner) using an embedded recurrent network. These two routes are combined to generate predictions about the most likely candidates for the next word.

The model's prediction is then compared to the actual word that occurs in the input. The discrepancy (prediction error) is used to adjust the weights in the model's sequencing units as well as in the connections between the sequencing

units and the semantic units, and between words and lexical concepts. The particular structure of this network – its combination of single recurrent network for sequencing and a more symbolic architecture to represent meaning – allows it to learn much more abstract mappings between meaning and linguistic form than Elman's original network. This means that, like children, it can represent fully abstract syntactic categories and, like adults, can generate predictions about upcoming words in novel situations. In sum, when prediction-driven learning is combined with an adequately structured cognitive architecture, it can generate behaviors that are plausibly similar to how children learn language, and how adults process it.

Not all prediction-error based models are connectionist. Ramscar, Dye, and McCauley (2013) proposed a much simpler model, based on ideas from learning theory, that attempts to explain how children master morphology (rather than syntax). In particular, the model attempts to learn which of a scene's semantic features are the best predictors of its appropriate phonological label (e.g., should scenes with multiple entities be given plural labels? Should that label be a regular or irregular plural?). The model assumes that children extract a range of semantic information from a scene (e.g., type of objects, number of objects) and gradually learn which of these features is informative with regard to its label. They do this using prediction. More specifically, children generate expectations about which labels they should hear for which scenes and, when they are wrong,

adjust the associations between meaning and form in accordance with the Rescorla-Wagner learning rule.

The Rescorla-Wagner rule provides a description of how learners should adjust associations between two stimuli. Historically, it was developed to provide descriptions of conditioning behaviour in animals (Rescorla & Wagner, 1972), but its application has been much broader (e.g., to areas of human psychology such as causal reasoning or category learning Gluck & Bower, 1988; Medin, Shanks, & Holyoak, 1996). It is particularly interesting because it provides a precise specification of how the model should learn from both positive evidence for an association between two stimuli, *and* from evidence that an association between two stimuli is absent (i.e., implicit negative evidence). For linguistic stimuli, this means that the *non*-occurrence of an expected event can be as informative as the *occurrence* of an unexpected event: Both cause the learner to update their associations between meaning and form and thereby learn to discriminate which aspects of meaning are informative with regard to the occurrence of a given form. This means that the model can explain notoriously challenging developmental phenomena such as the "no negative evidence" problems. For example, the aspect "plural number of objects" in combination with the aspect "mousiness" will initially lead children to predict both the regular form *mouses* and the irregular *mice* as labels for a scene with more than one mouse in it. However, only *mice* will occur when both meaning aspects are present (and it will not occur when

either of them is not present); therefore the child will learn to associate *mice* to the perception of more than one mouse, and commensurately reduce the association with *mouses*.

These models are not perfect. Elman's fails to learn fully abstract syntax; Chang Dell and Bock's fails to explain certain standard results in adult psycholinguistics (such as the lexical boost, Pickering & Branigan, 1998, although see Chang, Baumann, Pappert, & Fitz, 2014); Ramscar, Dye and McCauley's has no notion of combinatorial linguistic structure. But they all capture the idea that on-line, prediction-driven learning can help children to extract structure that is not clearly flagged in the input. That is to say, the models show how children can solve some of the major problems of language acquisition: Prediction lets them learn the distributional structure of the mapping from semantics to word sequences (i.e., the model can discover syntax) or structure in the mapping from lexical semantics to phonology (i.e., the model can discover morphology). What is the evidence that these predictive mechanisms allow children to master language?

Requirements for prediction-based learning

Based on these models, and on an analysis of the problems children face in language learning, we can ask what sort of predictions children will need to make in order to generate the necessary error signals for predictive learning. One critical component is that children's predictions must be incremental: Children

must be generating expectations about upcoming words based on what they have heard so far. This characteristic is important for performing a distributional syntactic analysis of the input (as demonstrated in the models of Elman 1990, 1993, and Chang et al., 2006). This means that children must be able to construct partial interpretations of sentences, predict based on these, and then update based on the input. For example, if a child hears the sentence *The boy will eat the cake,* it is not sufficient that she only recognize *boy* as the subject and *cake* as the object after having heard the whole sentence. Rather, she must build a partial interpretation in which *boy* is recognized as the subject before the sentence ends, allowing her to predict that the object might be *cake.*

Children must also be predicting using information that will generate an error signal at the appropriate level of analysis. For instance, in order to learn the syntax of their language, children should predict based on the syntactic knowledge that they have acquired so far, and not simply rely on predictions based on simpler representations, such as semantic associations. The difference between these two strategies can be illustrated by the verb *eat,* which is related to both *cake* and *chew,* but does not predict *chew* in *The boy ate the…* This difference can be seen in Chang et al.'s (2006) model, in which expectations about upcoming words are not only driven by semantic units, but also by representations of previously learned linguistic sequences.

Useful predictions must also be highly detailed, containing not just meanings or syntactic categories, but also sounds. For instance, there are certain facts about English that predictive learners can only capture if their predictions incorporate the sounds of words. These include phonological regularities (e.g., which nouns require *a/an*) and also grammatical *irregularities*. In particular, predictions have to include the sounds of upcoming words if they are going to help with the "no negative evidence" problem, as seen in Ramscar's model. To illustrate, if children consistently predict that the upcoming phonology of a word will be *mouses* rather than *mice*, then the resulting error signal when they encounter *mice* should lead them to the correct form. But if, instead, their predictions solely concern the syntax and semantics of upcoming words (i.e., children expect a plural noun with the meaning "multiples of mouse"), then their input (*mice*) will be completely consistent with their predictions (as *mice* and *mouses* are synonyms), and should not result in learning.

Finally, in an optimal system, predictions should be probabilistic and parallel: A system that is able to simultaneously suggest, weigh and check multiple possibilities is considerably more adaptive (fast-learning) than a system that can only check one hypothesis at a time, which will be slow and ponderous (this is an aspect of all the models discussed above). As an example, on hearing *The girl ate the…* a child would assign different probabilities to multiple outcomes: *cake, bread, potato, water,* etc., and then—once the predicted word has occurred—

could update all these probabilities in order to make their language more adult-like. All of the above models have this feature at least at some level of analysis (e.g., in Chang et al.'s model more than one meaning representation can be active in parallel but a single word is predicted at any given time).

Do children's predictions meet these requirements? Interestingly, we know from prior work that *adults'* predictions have all of the required properties: Adults make predictions incrementally, they predict on the basis of several types of cues, and they predict at different linguistic levels (from semantics to phonology). Evidence for incrementality comes from Kamide, Altmann, and Haywood's (2003) demonstration that adults update their predictions based on what has been said so far. They recorded adult listeners' eye movements to a visual scene while listening to sentences such as *The girl will ride the carousel* or *The man will ride the motorbike*. Upon hearing *ride*, listeners were more likely to look towards the agent-related theme, suggesting that they combined semantic information about the agent (e.g., that girls are too small for motorbikes) with semantic information provided by the verb, and anticipated what would be mentioned next (although see the discussion of Borovsky, Elman, & Fernald, 2012 below). Similarly, adult listeners combine intonation with semantic information to constrain predictions of upcoming referents, at least when the input is syntactically ambiguous (Weber, Grice, & Crocker, 2006).

To make these incremental predictions, adults attend to multiple levels of linguistic representation. They generate predictions based on semantic associations (e.g., expecting to hear dog soon after hearing cat, Lau, Holcomb, & Kuperberg, 2013), they generate predictions based on syntactic structure (e.g., expecting to read a noun after reading a determiner, Hale, 2001; Levy, 2008), and they generate predictions based on a combination of the syntax so far and their background semantic knowledge. This final point was demonstrated in a study by Kukona, Fang, Aicher, Chen, and Magnuson (2011): Adults listened to phrases such as *Toby will arrest the…,* while viewing an array of characters (e.g., a robber, a policeman, and a fireman). On hearing *arrest,* adults gazed to characters associated with the action (e.g., robber, policeman), suggesting that they predicted on the basis of semantic associations. However, they were more likely to gaze toward characters that were not only associated, but also were about to be mentioned (i.e., the robber), especially when given more time. This suggests that expectations are not only generated based on semantic associations, but also based on verbal argument structure.

Moreover, when adults make predictions, they make them at a fine level of detail, from semantics to syntax and all the way down to the phonological form of upcoming words. For instance, Dikker, Rabagliati, and Pylkkanen (2009) used magnetoencephalography to demonstrate that adults show enhanced "mismatch" responses in sensory cortices when they read words whose written form does

not match their expectations (see also Kim & Lai, 2012). DeLong, Urbach, and Kutas (2005) used electroencephalography to show that adults generate expectations that include the form of upcoming articles, in particular showing a mismatch response to *an* when reading the phrase *the day was windy so the boy went out to fly an….* The context in this phrase leads them to expect the word *kite*, which should be preceded by *a* not *an*. Since the two articles have the same semantic content, the mismatch response appears to require a phonological prediction.

Finally, adults also appear to make predictions in parallel. For instance, when participants read unexpected words, their time gazing on that word is directly related to its predictability (Smith & Levy, 2013): Highly-expected words are read faster than less-expected words, which are read faster than even-less-expected words, and so on. The same appears to be true of the N400, a component of the event-related-potential response to reading a word: The size of the N400 indexes the degree to which a word is expected or unexpected in its context (e.g., DeLong et al., 2005; Kutas & Federmeier, 2000) and it also reflects the number of words that are similar in form or lexical meaning to the target word (i.e., how difficult it is to select the target word amongst other possible candidates Laszlo & Federmeier, 2011).

The predictions that adults make, therefore, would seem to be ideal for learning from (and indeed there is evidence that adults' grammatical preferences change through error-based learning, Jaeger & Snider, 2013). This raises the possibility that language-learning children may be generating similar predictions, allowing them to use online processing mechanisms to acquire language.

Prediction and processing in young children.

While it is clear that children do eventually learn to make predictions of all types (after all, children eventually become adults), it is not clear whether children can make predictions that are useful for learning *before* they have fully mastered their language, and it is even less clear whether children can *use* these predictions for the purpose of mastering their language. Here, we survey the evidence that language-learning children make predictions that are incremental, that are based on linguistic structure, that include phonological information, and that are made in parallel.

A prerequisite of any prediction-based theory is that children must process language in an online fashion. In particular, if children are to make incremental predictions, then they must interpret language incrementally: They should interpret words as they hear them, and not wait until a sentence ends in order to

determine what it means. Consistent with this, young children's language processing does appear, in many respects, to be almost adult-like. For instance, children in their second year can interpret words quickly and incrementally: Eye tracking work suggests that 18 month-olds can reliably identify the referents of familiar words within 300ms of their onset (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Swingley, Pinto, & Fernald, 1999). Similarly, preschool children are able to process complex syntactic structures on-line (Thothathiri & Snedeker, 2008; Trueswell, Sekerina, Hill, & Logrip, 1999) and can use multiple sources of information (e.g., prosody, lexical co-occurrence statistics) to determine the most likely meaning of a syntactically ambiguous sentence (Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008).

Beyond this, there are now a number of demonstrations that children can make guesses about what will be referred to next in a sentence. The importance of this for learning, however, is unclear. Early work by Nation, Marshall, and Altmann (2003) used a visual world paradigm to demonstrate that older children can generate semantic predictions: 10- and 11-year-olds viewed scenes while listening to sentences where the final word could be predicted (*Jane watched her mother eat a cake*) or not (*Jane watched her mother choose a cake*). Children's eye movements suggested that, immediately after encountering the verb, they could predict which object in the scene would be mentioned next: They tended to fixate that object before it was mentioned. Of course, these children were 10

years old, and had typically mastered the grammar of their language. Evidence that (much) younger children also make predictions comes from Mani and Huettig (2012). They used the same basic paradigm (but with much simpler scenes) to show that even 24-month-olds can use verb meanings to generate expectations about upcoming meanings (contrasting *The boy eats the big cake* and *The boy sees the big cake*). This effect was only seen in children with larger expressive vocabularies (see also Borovsky, Elman and Fernald, 2012), which suggests that, in some way, children's predictions are tied to linguistic ability, which we discuss further in the next section. But these experiments do not show that children are making predictions that can be, or are, used for learning: It is unclear if the predictions are incremental, if they are based on linguistic structure, if they include phonological information, and if they are made in parallel.

Potential evidence for incremental predictions comes from Borovsky, Elman, and Fernald (2012). They demonstrated that children were able to combine semantic information from the subject and verb of a sentence to anticipate an upcoming referent. For example, children were shown an array containing a ship, a cat, a bone, and some treasure, while listening to phrases such as *the pirate chased the…* or *the dog chased the….* On hearing the former phrase, they gazed toward the ship before it was even mentioned, whereas on hearing the latter phrase they anticipatorily gazed toward the cat. Further work has shown that children can generate these predictions even for newly learned-about events (e.g., after

learning that monkeys tend to ride buses, and tigers tend to ride trains, they will look at a bus on hearing *the monkey rides…,* Borovsky, Sweeney, Elman & Fernald, 2014) This suggests that children's predictions are generated based on the incremental comprehension of multiple words (e.g., *pirate* and *chase* activate *ship*; *dog* and *chase* activate *cat*).

However, this work leaves unclear the exact process by which the predictions are generated. In particular, the method cannot distinguish between predictions generated by the composition of multiple words (e.g., where children generate predictions by entering *pirate* as an argument for *chase*) and predictions generated by independent lexical access to the meaning of each of these words (e.g., *pirate* primes *ship, chase* primes *ship*, and so when both are accessed *ship* is highly primed). To our knowledge, there is no current data that distinguishes these possibilities. That is to say, children's predictions do appear to be incremental (in that they are made as words are heard), but it is not clear that they are based on incremental composition of linguistic units, which is the type of incrementality that would be necessary for learning.

Better evidence that children are generating predictions based on the incremental composition of linguistic units comes from work on syntactic priming. Thothathiri and Snedeker (2008) show that priming a verb's argument structure affects children's predictions for upcoming syntax. They had 3- and 4-year-olds children

listen to instructions containing verbs that can occur in either prepositional object constructions such as *Give the birdhouse to the sheep* or double object constructions such as *Give the bird the bone*. Children then acted out these instructions with displays that contained appropriate objects, in this case a birdhouse, a bird, a sheep, and a bone. Importantly, these instructions are identical (and thus ambiguous) up to the end of the word *bird.* However, Thothathiri and Snedeker found that children's interpretations of these ambiguous phrases were biased by previously occurring phrases: On hearing *bird,* they were more likely to look at the birdhouse when they had just heard another sentence with a prepositional object structure (e.g., *bring the ball to the lion*) rather than a double object structure (*bring the lion the ball*). This suggests that children had some sort of expectation about the syntactic structure of the sentence that they were about to hear. However, this work can also be interpreted as an effect of *integration*, where children do not pre-specify the appropriate argument structure, but rather begin building it on hearing appropriate auditory input.

This problem – the difficulty of disentangling effects of prediction and effects of integration – affects any study in which the dependent measure is a response to a predicted or unpredicted word (for a discussion with specific reference to the N400 ERP component, see Hagoort, Baggio, & Willems, 2009; Kutas & Federmeier, 2011), rather than a direct estimate of the prediction itself (for

examples of this, see for instance DeLong et al., 2005; Stokes, Thompson, Nobre, & Duncan, 2009). The difficulty of distinguishing between predictive and integrative accounts can be seen in an ERP study by Bernal, Dehaene-Lambertz, Millotte and Christophe (2010). They examined 2-year-olds' EEG responses to syntactically grammatical (i.e., expected) or ungrammatical (unexpected) words, and found a consistent response to words that were ungrammatical in their syntactic context. They suggest that this response might reflect a violation of a syntactic prediction (e.g., children expected a noun, but heard a verb), but are also clear that it could be due to the difficulty of integrating an ungrammatical word into a syntactic structure (e.g., children search for, but cannot find, a phrase structure rule that will allow them to add a noun phrase to their syntactic structure). Difficulties disentangling prediction from integration can also be seen in work on children's N400 responses (e.g., Friedrich & Friederici, 2004; Friedrich & Friederici, 2006), and eye tracking work looking at the effects of context on word recognition (Lew-Williams & Fernald, 2007).

As should be clear, we currently lack direct evidence that children make the types of predictions that would aid learning. What about indirect evidence? This turns out to be mixed. Some indirect evidence is consistent with the idea that children might have the wherewithal to make phonological predictions. For instance, work on "implicit naming" suggests that children may be able to actively hold the phonological form of unspoken words in mind before their second birthday (this

ability is presumably a prerequisite for making a phonological prediction). Mani and Plunkett (2010) found that, after showing 18-month-olds a picture (e.g., of a dog), the infants were faster to recognise phonologically related words (e.g., *door* vs. *boat*), even though that initial picture was never explicitly named. This suggests that viewing the initial picture primed the phonological form of its name, which then primed the recognition of related target words.

Recent work by Khan (2013) provides further evidence for implicit phonological priming. She found that, after passively viewing a picture, 24-month-olds tended to gaze toward "phonosemantically" related targets. For instance, after viewing a cup, infants gazed toward a picture of a dog over a picture of a box. This may seem surprising, but it makes sense when you realise that *cup* is phonologically related to *cat*, which semantically primes *dog* (this phonosemantic priming effect has also been found in five-year-olds, Huang and Snedeker, 2011). These results are not only important because they suggest that children can generate phonological and semantic representations of unspoken words (providing a potential basis for predictions at both levels of representation), but also because they indicate that even 2-year-olds possess an extremely interactive cognitive architecture, which permits extensive pre-activation of representations associated with the current input.

However, other indirect evidence raises questions about the true interactivity of children's early cognitive architectures. A body of recent work on children's sentence processing has shown that children as old as 5 years have difficulty using high-level linguistic information, such as discourse structure, to constrain processes such as syntactic ambiguity resolution or semantic interpretation (Arnold, Brown-Schmidt, & Trueswell, 2007; Snedeker & Trueswell, 2004; Trueswell et al., 1999). As one example, eye-tracking work has shown that it takes 5-year-olds about one-and-a-half seconds to resolve an ambiguous pronoun based on discourse cues (e.g., referring to the first-mentioned referent in a context such as *Sally ran in front of Mary. She…*) while adults resolve this type of pronoun quickly and proactively (Arnold et al., 2007; Hartshorne et al., 2014). Results like this raise questions about whether children have the processing capacity to quickly generate predictions from which they can learn.

The degree to which children consistently generate predictions from which they can learn is therefore unclear. We know that children do appear to have mechanisms for generating expectations, but we still need to know what information children use to generate those expectations, how detailed those expectations are, and whether expectations are generated rapidly enough to allow prediction-based learning.

Predicting to learn or learning to predict?

PREDICTING TO LEARN?

While the evidence so far is suggestive that children may indeed "predict to learn", it is also consistent with an alternative possibility: Once children begin to master language, then they also start to master the mechanisms necessary to produce linguistic predictions. That is to say, children "learn to predict". In this case, prediction would be a *consequence* of language learning, and not a cause.

Unfortunately, distinguishing between these two accounts – predicting to learn and learning to predict – is not simple. To do so, we would not only need to conduct experiments that test if children make the types of prediction that are necessary for learning, but we would also need to determine whether prediction is a cause or effect of language learning. This would require detailed longitudinal assessments of children's ability to predict and process upcoming language, and an understanding of how this relates to their linguistic knowledge more broadly. The relationship, of course, is unlikely to be simple, as accurate predictions necessarily require a detailed knowledge of the language. For instance, it is hard to distinguish between a child who does not make predictions at all, and a child who makes very inaccurate predictions due to their incomplete linguistic knowledge. In order to discover the degree to which children predict to learn, and the degree to which children learn to predict, we would need to determine whether learners make *inaccurate* predictions early in development. Then, we would need to test if these initially inaccurate predictions help children to (eventually) learn an accurate grammar. For instance, we could determine

whether children who are better predictors (e.g., children who consistently make predictions about upcoming words, even if they are inaccurate) end up learning faster than language-matched peers. A somewhat-analogous approach has already been successfully pursued in work by Fernald, Marchman and colleagues (Fernald & Marchman, 2012; Fernald, Perfors, & Marchman, 2006; Marchman & Fernald, 2008), whose studies have assessed how variations in children's linguistic processing speed (at a word recognition task) affects their later language development.[1]

The successful completion of this enterprise will be greatly helped by the explicit computational models that have been developed in this area (e.g., Chang et al., 2006; Ramscar et al., 2013). These can generate accurate hypotheses about what children predict, and also about how children should behave when they receive evidence that mismatches their predictions. The theoretical clarity and computational precision of the models described in Section 2 is impressive, and they will clearly play an important part in solving the puzzles posed by children's predictions. However, we would like to emphasize that these are not the only

---

[1] Note that Fernald's work raises a potential alternative route by which prediction could influence language development. So far, we have contrasted views in which prediction is specifically used as a learning mechanism, with views in which prediction is a characteristic of more expert systems and is not used for learning. But Fernald's work suggests that, under the expert system account, prediction could still facilitate language acquisition, by increasing processing speed and thereby acting as a crutch for acquiring knowledge of words and grammar. Prediction can therefore facilitate learning even if children do not use predictive, error-driven learning.

possible accounts by which prediction-based learning could help children acquire language, and that evidence for prediction-based learning should not necessarily be taken as evidence that children learn language in a manner that is either connectionist (Chang et al., 2006; Elman, 1990) or neo-behaviourist (Ramscar et al., 2013).

For example, prediction-based learning could also play an important role for models in which children learn explicit, structured representations. Informally, this can be seen in Chomsky's (1981) suggestion that children could use prediction error to overcome the no negative evidence problem. More formally, predictive learning could play an important role in mechanistically implementing models of language learning that are typically couched at a "Computational" (i.e., amechanistic) level of analysis, such as Perfors et al.'s (2011) Bayesian model of syntactic development.

Support for this idea, that predictive-learning can aide acquisition of structured representations, comes from outside the domain of language learning. In particular, recent research into reinforcement learning has provided evidence for so-called "model-based" reinforcement learning approaches, in which learners acquire explicit and structured representations of the environment that they are learning about (such as an explicit map of a spatial environment, Daw, 2012; Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Doll, Simon, & Daw, 2012;

Otto, Gershman, Markman, & Daw, 2013). These models are acquired via prediction about future states of the world followed by adjustment from an error signal, just as in the "model-free" approaches of e.g., Ramscar et al. (2013). Model-based approaches are particularly promising because, when some of the basic parameters of the model can be specified in advance, then it is easy to learn an accurate representation of the world from relatively little data (Dayan & Daw, 2008; Doya, Samejima, Katagiri, & Kawato, 2002). Model-based learning has another advantage, in that possessing a model of the world allows the learner to rapidly change how they want to interact with it, based on only limited changes in information about the environment. By contrast, it is hard to immediately change the habits learned by a model-free approach (Dayan & Daw, 2008). There is a clear analogy here between model-based learning and language development and processing: Children could use predictive learning to explore a space of possible grammars, and eventually determine the best-fitting model of the sentences that they have already heard. Innate restrictions on possible grammars would facilitate this learning. Possessing an explicit representation of the grammar (and of how other speakers tend to behave) would allow the learner to change its behaviour with limited incoming evidence (for evidence that adults behave in a way that is as flexible as this, see Nieuwland & Van Berkum, 2006).

The role of dialogue in children's predictions

We began this paper by talking about conversation, and about the important roles that prediction may play in explaining both how adults converse and how non-linguistic children enter conversation for the first time. We have argued that, despite the importance of prediction for a number of different theories of language learning, and despite the strong evidence that prediction plays an important role in adult language processing, we currently lack grounds to conclude that prediction plays a role in acquisition – the evidence is suggestive, but not yet conclusive. However, this does not mean that prediction plays no role, and we have described potential empirical strategies for confirming the role of prediction in learning.

In our discussion of prediction-based learning, we have therefore strayed far from our opening discussion of conversation. At least, we have focused on much smaller levels of analysis, such as words and sentences. Words and sentences are important: The major questions of language acquisition are focused on how children master meanings and syntactic structures. But dialogues between speakers are probably the most natural form of language, and so are likely to play a vital role in how children learn about words and syntax in the first place. Moreover, dialogues are also the situations in which young children would most likely benefit from using prediction to produce rapid responses, as well as usually providing the conditions for such predictions to be more accurate (as dialogues are typically situated in a rich extra-linguistic context and tend to involve a lot of

repetition). It is therefore important to return to the question of how children master dialogue.

Current theories of dialogue do not provide a natural explanation of how children learn to interrelate turn-taking and linguistic structure to create conversation. While there is a strong role for "expert" predictions in almost all theories, there has been little discussion of how that expert behaviour develops. Levinson's (2006) Interaction Engine Hypothesis assumes that human children are born with the basic building blocks of conversation, so there is no need to explain their development. Pickering and Garrod's (2004) Interactive Alignment model simply does not touch on the development of dialogue. One way to interpret the Interactive Alignment model is that dialogue is a consequence of mastering other types of linguistic representation, and not an independently learnt skill. In the model, levels of representation automatically prime one another across partners (one speaker's syntactic representations align with her partner's syntactic representations, her prosodic representations align with his prosodic representations, etc.), and also prime one another across levels of representation within a partner. That is to say, children's ability to align at one level of representation will automatically allow them to begin aligning at other levels of representation. If these sorts of automatic priming mechanisms were present in young children, then it would help them to engage in adult-like dialogue while minimizing the need to learn additional mechanisms for conversation. There is

indeed evidence that children as young as 3 are sensitive to syntactic priming (Branigan, McLean, & Jones, 2005; Huttenlocher, Vasilyeva, & Shimpi, 2004; Rowland, Chang, Ambridge, Pine, & Lieven, 2012; Thothathiri & Snedeker, 2008), although we do not know how these priming effects relate to children's conversational abilities.

The Interactive Alignment model suggests a very "bottom-up" model of how children learn language in its conversational context. In particular, the theory stands in contrast to other models (e.g., Frank, Goodman, & Tenenbaum, 2009; Levinson, 2006; Tomasello, 1999) in which children are assumed to use skills for reading high-level intentions to master the lower-level linguistic structures that define their language. This difference reflects a fundamental property of the Interactive Alignment model: Its mechanistic explanation of how interlocutors use low-level, perceivable cues to "read one another's minds". Developmentally, this suggests that children may initially learn to align at lower levels of representation, such as form (e.g., prosodic form or lexical form, Morgan & Demuth, 1995) before they learn to align at higher-level representations such as semantic or discourse structure. This fits with the evidence that even children as old as five years have difficulty using discourse and pragmatic information to interpret spoken sentences online (Huang & Snedeker, 2009; Huang & Snedeker, 2013; Snedeker & Trueswell, 2004).

These key properties of the Interaction Engine and Interactive Alignment models mean that while prediction can play a role in both, that role is subtly yet importantly different. Under the Interaction Engine hypothesis, children can automatically use (possibly innate) knowledge of social interaction, such as mental states and principles of conversational interaction, to inform their prediction-based learning of their language's words and grammar. By contrast, under the Interactive Alignment model, children use prediction to not only discover the words and structures of their language, but also (presumably at the same time) to discover how those words and structures relate to factors such as mental states and principles of conversational interaction.

These considerations suggest some changes to our initial story about the role of prediction. At the start of the paper, we noted that mechanisms of prediction were now being used to explain two distinct things: adults' ability to rapidly converse, and children's ability to learn language. The considerations above suggest that, in fact, these may not be so distinct after all. In particular, prediction, conversation, and learning may all interact, such that prediction and learning help to explain how, as adults, we engage in conversation, while conversation, in concert with prediction, helps to explain how children eventually learn.

Acknowledgments

References

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes, 22*(4), 527-565.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn Sci, 11*(7), 280-289.

Bernal, S., Dehaene‐Lambertz, G., Millotte, S., & Christophe, A. (2010). Two‐year‐olds compute syntactic structure on‐line. *Developmental Science, 13*(1), 69-76.

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

Bowerman, M. (1987). Commentary: Mechanisms of Language Acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 443–466). Hillsdale, NJ: LEA.

Braine, M. D. (1963). The ontogeny of English phrase structure: The first phase. *Language*, 1-13.

Branigan, H. P., McLean, J. F., & Jones, M. W. (2005). *A blue cat or a cat that is blue? Evidence for abstract syntax in young childrenís noun phrases.*

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science, 26*(5), 609-651.

Chang, F., Baumann, M., Pappert, S., & Fitz, H. (2014). Do Lemmas Speak German? A Verb Position Effect in German Structural Priming. *Cognitive Science*.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological review, 113*(2), 234.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht, Netherlands: Foris.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181-204. doi: doi:10.1017/S0140525X12000477

Daw, N. D. (2012). Model-based reinforcement learning as cognitive search: neurocomputational theories. *Cognitive Search: Evolution, Algorithms and the Brain*, 195-208.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204-1215.

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience, 8*(4), 429-453.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat Neurosci, 8*(8), 1117-1121.

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current opinion in neurobiology, 22*(6), 1075-1081.

Doya, K., Samejima, K., Katagiri, K.-i., & Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural computation, 14*(6), 1347-1369.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179-211.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*(1), 71-99.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late‐talking toddlers. *Child development, 83*(1), 203-222.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology, 42*(1), 98.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 228-231.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*(5), 578-585.

Friedrich, M., & Friederici, A. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *Journal of cognitive neuroscience, 16*(8), 1465-1477.

Friedrich, M., & Friederici, A. D. (2006). Early N400 development and later language acquisition. *Psychophysiology, 43*(1), 1-12.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General, 117*(3), 227.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzanigga (Ed.), *The cognitive neurosciences, 4th ed.* (pp. 819-836). Cambridge, MA: MIT Press.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model *Proceedings of NAACL* (Vol. 2, pp. 159-166).

Hartshorne, J. K., Nappa, R., & Snedeker, J. (2014). Development of the first-mention bias. *Journal of Child Language*, 1-24.

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental psychology, 45*(6), 1723-1739.

Huang, Y. T., & Snedeker, J. (2011). Cascading activation across levels of representation in children's lexical processing. *Journal of Child Language, 38*(03), 644-661.

Huang, Y. T., & Snedeker, J. (2013). The use of referential context in children's online interpretation of adjectives. *Developmental psychology, 49*, 1090-1102.

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language, 50*(2), 182-195.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*(1), 57-83.

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

Khan, M. (2013). *Thinking in words: Implicit verbal activation in children and adults.* (Ph.D thesis), Harvard University, Cambridge, MA.

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

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

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. *Predictions in the brain: Using our past to generate a future*, 190-207.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci, 4*(12), 463-470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology, 62*, 621-647.

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology, 48*(2), 176-186.

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

Levinson, S. C. (2006). On the human "interaction engine". In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39-69): Berg.

Levy, R. (2007). *Expectation-based syntactic comprehension*

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18*(3), 193-198.

MacWhinney, B. (1987). The competition model. In B. Macwhinney (Ed.), *Mechanisms of language acquisition* (pp. 249-308). Hillsdale, NJ: LEA.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language, 31*(04), 883-914.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 843-847.

Mani, N., & Plunkett, K. (2010). In the Infant's Mind's Ear Evidence for Implicit Naming in 18-Month-Olds. *Psychological Science, 21*(7), 908-913.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*(3), F9-F16.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT press.

Medin, D. L., Shanks, D. R., & Holyoak, K. J. (1996). *Causal Learning: Advances in Research and Theory* (Vol. 34): Academic Press.

Morgan, J. L., & Demuth, K. (1995). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*: Psychology Press.

Nation, K., Marshall, C. M., & Altmann, G. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86*(4), 314-329.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience, 18*(7), 1098-1111.

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science*, 0956797612463080.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118*(3), 306-338.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*(4), 633-651.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-190.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(04), 329-347. doi: doi:10.1017/S0140525X12001495

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: HUP.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.

Pinker, S. (1995). Language acquisition. In L. Gleitman & M. Lieberman (Eds.), *An Invitation to Cognitive Science: Language* (pp. 135-182): MIT Press Cambridge, MA.

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.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory, 2*, 64-99.

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition, 125*(1), 49-63.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*(4), 569-588.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*(3), 302-319.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive psychology, 49*(3), 238-299. doi: DOI 10.1016/j.cogpsych.2004.03.001

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58*(2), 574-608.

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences, 106*(46), 19569-19574.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*(2), 73-108.

Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three-and four-year-old children. *Journal of Memory and Language, 58*(2), 188-213.

Tomasello, M. (1992). *First Verbs: A Case Study of Early Grammatical Development*. Cambridge, UK: CUP.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: HUP.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73*(2), 89-134.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition, 99*(2), B63-B72.

Wexler, K., & Culicover, P. (1980). *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.