



Contents lists available at SciVerse ScienceDirect

## Cognition

journal homepage: [www.elsevier.com/locate/COGNIT](http://www.elsevier.com/locate/COGNIT)

# Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience

T. Florian Jaeger<sup>a,b,\*</sup>, Neal E. Snider<sup>c</sup><sup>a</sup> Department of Brain and Cognitive Sciences, University of Rochester, Meliora Hall, Box 270268, Rochester, NY 14627-0268, United States<sup>b</sup> Department of Computer Science, University of Rochester, Meliora Hall, Box 270268, Rochester, NY 14627-0268, United States<sup>c</sup> Nuance Communications, Inc., 1 Wayside Road, Burlington, MA 01803, United States

## ARTICLE INFO

## Article history:

Received 7 July 2010

Revised 5 October 2012

Accepted 22 October 2012

Available online xxxx

## Keywords:

Structural persistence

Syntactic priming

Alignment

Adaptation

Implicit learning

Prediction error

Surprisal

## ABSTRACT

Speakers show a remarkable tendency to align their productions with their interlocutors'. Focusing on sentence production, we investigate the cognitive systems underlying such alignment (syntactic priming). Our guiding hypothesis is that syntactic priming is a consequence of a language processing system that is organized to achieve efficient communication in an ever-changing (subjectively non-stationary) environment. We build on recent work suggesting that comprehenders adapt to the statistics of the current environment. If such adaptation is rational or near-rational, the extent to which speakers adapt their expectations for a syntactic structure after processing a prime sentence should be sensitive to the prediction error experienced while processing the prime. This prediction is shared by certain error-based implicit learning accounts, but not by most other accounts of syntactic priming. In three studies, we test this prediction against data from conversational speech, speech during picture description, and written production during sentence completion. All three studies find stronger syntactic priming for primes associated with a larger prediction error (primes with higher syntactic surprisal). We find that the relevant prediction error is sensitive to both prior and recent experience within the experiment. Together with other findings, this supports accounts that attribute syntactic priming to expectation adaptation.

© 2012 Elsevier B.V. All rights reserved.

## 1. Introduction

When we talk, we align with our conversation partners along various levels of linguistic representation. This includes decisions about speech rate and how we articulate sounds, as well as lexical and structural decisions. Here we focus on alignment of syntactic structure, also known as syntactic priming or structural persistence (Bock, 1986; Pickering & Branigan, 1998). Syntactic priming has received an enormous amount of attention in the psycholinguistic literature (for a recent overview, see Pickering

& Ferreira, 2008). With respect to language production, which we will be concerned with here, syntactic priming refers to the increased probability of re-using recently processed syntactic structures. For example, comprehending a passive sentence (e.g., *The church was struck by lightning*) increases the probability of encoding the next transitive event with a passive rather than an active structure.

A large body of work has investigated under what conditions syntactic priming is observed. Thanks to this work, it is known that syntactic priming is observed in both spoken and written production and that it is observed independent of whether the prime was produced or comprehended (to name just two findings). What has emerged from this work is that priming effects are small but robust. Others have investigated what factors modulate the *strength of syntactic priming* – that is, the

\* Corresponding author at: Department of Brain and Cognitive Sciences, University of Rochester, Meliora Hall, Box 270268, Rochester, NY 14627-0268, United States. Tel.: +1 585 276 3611; fax: +1 419 821 8374.

E-mail address: [fjaeger@bcs.rochester.edu](mailto:fjaeger@bcs.rochester.edu) (T.F. Jaeger).

magnitude of the increase in the probability of re-using the syntactic structure of the prime. For example, a stronger priming effect is observed for target sentences with the same verb as the prime, compared to targets that do not overlap lexically with the prime (the ‘lexical boost’ effect, e.g., Hartsuiker, Bernolet, Schoonbaert, Speybroeck, & Vanderelst, 2008; Pickering & Branigan, 1998; Snider, 2008).

Considerably less is known about what *causes* syntactic alignment. Despite broad agreement on the significance of this question, relatively few studies have addressed it (e.g., Bock & Griffin, 2000; Chang, Dell, & Bock, 2006; Kaschak, 2007; Malhotra, 2009; Pickering, Branigan, Cleland, & Stewart, 2000; Reitter, Keller, & Moore, 2011). We explore the hypothesis that syntactic priming is a consequence of adaptation with the goal to minimize the expected prediction error experienced while processing subsequent sentences, thereby facilitating efficient information transfer (cf. Jaeger, 2010). This view owes intellectual debt to, and builds on, previous accounts of syntactic priming in terms of implicit learning (in particular, Chang et al., 2006; but also Bock & Griffin, 2000; Chang, Dell, Bock, & Griffin, 2000; Kaschak, 2007). We use the term adaptation or expectation adaptation as a mechanism-neutral term to refer to changes in the expectations or beliefs held by producers and comprehenders. By prediction error, we refer to the deviation between what is observed and expectations prior to the observation. In particular, we focus on the syntactic prediction error, the degree to which expectations for syntactic structures are violated during incremental language understanding.<sup>1</sup> The minimization of future prediction errors – or, more cautiously, the maximization of utility, which usually entails the ability to reduce the prediction error – is broadly accepted to be one of the central functions of the brain (for a summary of relevant work, see Clark, *in press*).

In order to situate our approach to syntactic priming within a broader theoretical context, we begin by reviewing the role of prediction errors in language processing. This leads us to recent work on syntactic priming and adaptation in comprehension, and the question as to how comprehenders determine *how much to adapt their expectations* for future sentences whenever a prediction error is experienced. Once we have established this broader context, we discuss the consequences for syntactic priming during language production.

### 1.1. Prediction errors in language comprehension

The prediction error experienced while processing a word or sentence affects the processing difficulty associated with it. For example, the processing difficulty experienced when temporarily ambiguous sequences of words are disambiguated towards a specific interpretation

(so-called ‘garden path’ effects) depends on how expected that interpretation was given the preceding context (e.g., Garnsey, Pearlmutter, Meyers, & Lotocky, 1997; Hare, McRae, & Elman, 2003; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Kello, 1993). Similarly, word-by-word processing difficulty during reading is a function of how expected the word is given preceding context (among other factors, e.g., DeLong, Urbach, & Kutas, 2005; Demberg & Keller, 2008; Levy, 2008; McDonald & Shillcock, 2003; Rayner & Duffy, 1988; Staub & Clifton, 2006).

Sensitivity to prediction errors is a natural consequence of a processing system that has developed to process language efficiently: expectations based on previous experience help to overcome the noisiness of the perceptual input and to deal efficiently with uncertainty about the incremental parse (see also Levy, 2008; Norris & McQueen, 2008; Smith & Levy, 2008). This assumes that comprehenders’ expectations closely match the *actual* statistics of the linguistic environment, thereby minimizing the expected prediction error. This assumption might be seen as in conflict with another well-known property of language: speakers differ with regard to their production preferences, including syntactic preferences (e.g., Tagliamonte & Smith, 2005; Weiner & Labov, 1983). Even within a speaker, syntactic preference can vary dependent on, for example, register (Finegan & Biber, 2001; Sigley, 1997). As a consequence, the actual linguistic distributions frequently change. From the comprehender’s perspective, linguistic distributions are thus subjectively non-stationary. Provided that differences in environment-specific statistics are sufficiently large, this implies that language understanding will be more efficient if comprehenders continuously adapt their syntactic expectations to match the statistics of the current environment (e.g., speaker-specific production preferences).

Indeed, there is evidence for such behavior, which we have dubbed *expectation adaptation* elsewhere (Fine, Jaeger, Farmer, & Qian, *submitted for publication*). One piece of evidence comes from the burgeoning literature on syntactic priming in comprehension. For example, recent exposure to a syntactic structure results in faster processing if the same structure is encountered again (e.g., Arai, van Gompel, & Scheepers, 2007; Traxler, 2008). That these effects are due to changes in expectations is confirmed by evidence from anticipatory eye-movements during language comprehension. In a visual world eye-tracking paradigm, Thothathiri and Snedeker (2008) find that listeners were biased to expect the most recently experienced structure to be used again. These studies provide evidence that the most recently experienced prime affects expectations for upcoming syntactic structure. Other experiments have found that comprehenders integrate, not only the most recent prime, but rather the cumulative recent experience, into environment-specific syntactic expectations (e.g., Fine, Qian, Jaeger, & Jacobs, 2010; Hanulíková, van Alphen, van Goch, & Weber, 2012; Kamide, 2012; Kaschak & Glenberg, 2004b). For example, consider the case of garden path sentences, which are associated with processing difficulty at the disambiguation point. As mentioned above, this processing difficulty is a function of how unexpected the

<sup>1</sup> We do not imply any specific architecture of the language processing system. For example, it is possible that the syntactic prediction error reduces to prediction errors associated with expectations for sequences of words (e.g., Chang et al., 2006; Frank & Bod, 2011). For a discussion of the relation between syntactic and word-by-word expectations, see also Levy (2008).

disambiguated parse is. In self-paced reading experiments, Fine et al. (submitted for publication) found that this processing difficulty at the disambiguation point dissipates rapidly if the a priori unexpected parse is experienced frequently in the current environment. These studies also provide preliminary evidence that comprehenders' adapted expectations actually *converge* against the statistics of the environment.

Findings like these suggest that comprehenders continuously adapt their expectations to match – or at least approximate – environment-specific statistics (in this case, the distribution of syntactic structures in an experiment). But how do comprehenders know *how much to adapt* their expectations at each moment in time? One important source of information is the prediction error. Generally, the larger the prediction error, the more there is a need to adapt one's prior expectations (see also Courville, Daw, & Touretzky, 2006). This insight is, of course, incorporated in many models of learning, regardless of whether the prediction error is explicitly evoked in the learning algorithm (as in error-based learning, e.g., Elman, 1990; Rescorla & Wagner, 1972; Rumelhart, Hinton, & Williams, 1986) or not (as in other forms of supervised learning, such as Bayesian belief update, or unsupervised learning). If our perspective on expectation adaptation during language comprehension is correct, the degree to which comprehenders adapt their syntactic expectations after exposure to a prime sentence should be a function of the prediction error they experienced while processing that prime.<sup>2</sup>

We are not the first to make this prediction. In their seminal paper, Chang et al. (2006) present a connectionist model of language acquisition that shares our prediction. The primary purpose of their model is to account for the acquisition of syntax from sequences of words. In their model, the language comprehension and production systems are structured networks of simple recurrent networks (Elman, 1990). Syntactic acquisition is achieved by error-based implicit learning: during comprehension learners predict the next word, and the deviation between the comprehenders' expectations and the actually observed word (i.e., the prediction error) serves as an error signal that is used to adjust the weights in the network (Chang et al., 2006, p. 270; via backpropagation, Rumelhart et al., 1986). Henceforth, we will refer to this as the error-based model.

Syntactic priming follows from the error-based model with just the additional assumption that the same error-based learning processes operating during acquisition continue to operate throughout adult life (see also Botvinick & Plaut, 2004; Plaut, McClelland, Seidenberg, & Patterson,

1996). As we detail below, this also predicts that the strength of syntactic priming should be a function of the prediction error experienced while processing the prime (Chang et al., 2006, p. 255). The error-based model thus shares this prediction with the perspective advanced here, that syntactic priming is a consequence of adaptation with the goal to minimize the expected prediction error. In the discussion, we review important conceptual differences between the approaches as well as the predictions they make for future work. For now, we focus on their shared prediction.

## 1.2. Prediction errors and syntactic priming

Preliminary support for this prediction comes from syntactic priming in *comprehension*: a re-analysis of Thothathiri and Snedeker (2008) discussed above found that syntactic primes with larger prediction errors result in bigger changes in expectations following the prime (Fine & Jaeger, *in press*). Here, we turn to production.

Prima facie, speakers could contribute to mutual expectation alignment by aligning their production preferences with (their beliefs about) their interlocutors' expectations. Such alignment can contribute to the minimization of the joint future prediction error, thereby facilitating both faster information transfer (Levy, 2008; Levy & Jaeger, 2007) and more efficient information transfer (Aylett & Turk, 2004; Jaeger, 2006; Jaeger, 2010). Specifically, if interlocutors' production preferences are at least in part reflective of their expectations in comprehension, speakers should adapt their production preferences based on the prediction error they experienced while comprehending their interlocutors' utterances. Although the error-based model is not typically described in these terms, it describes an architecture that essentially achieves this type of mutual expectation alignment.

In the error-based model, comprehension and production share the same sequencing system (Chang et al., 2006, pp. 238–239). This system generates predictions about word sequences. In comprehension, these predictions contribute to the expectations speakers have during word-by-word sentence understanding. In production, they affect which word is produced next.<sup>3</sup> An elegant consequence of this – which to the best of our knowledge has not been discussed before – is that syntactic priming in production contributes to the reduction of interlocutors' prediction error during comprehension (and hence to the processing difficulty experienced by interlocutors): not only

<sup>2</sup> Note that the prediction error is only part of the answer to the problem of learning in a non-stationary environment. The fundamental problem is one of uncertainty about the *cause* for the prediction error. Any deviation from the expected could be informative in that it provides information about the statistics of the current environment. On the other hand, deviations from the expected are also bound to occur by chance. How likely a prediction error is to signal a change in the environment depends – among other things – on prior expectations for such a change. How the brain manages to learn under non-stationarity is a complex question that is just beginning to be explored (for an overview, see Qian, Jaeger, & Aslin, 2012).

<sup>3</sup> In the error-based model, the sequencing system is just one of several systems that determine language production (cf. Chang et al., 2006, Fig. 2, p. 237). Hence, the model does not predict that adaptation in comprehension needs to exactly mirror production (which would be wrong, e.g., Kraljic, Brennan, & Samuel, 2008). Here, we adopt these architectural assumptions. More generally, the assumption that speakers make predictions is shared by a variety of accounts (contrary to Pickering & Garrod, *in press*; see, e.g., Chang et al., 2006; Dell, Oppenheim, & Kittredge, 2008; Jaeger, 2010; Levy & Jaeger, 2007; Maurits, Perfors, & Navarro, 2010). Evidence for this claim comes, for example, from the observation that the realization of linguistic units (e.g., words or syntactic structures) is sensitive to their surprisal in context, mirroring what is observed in comprehension (e.g. Aylett & Turk, 2004; Frank & Jaeger, 2008; Jaeger, 2010; Resnik, 1996; Wasow, Jaeger, & Orr, 2011; van Son & van Santen, 2005).

would comprehenders adapt their expectations to match the interlocutor's production preferences, but they would also adjust their own production preferences so as to more closely resemble the expectations of their interlocutor.

This means that both producers and comprehenders would be contributing to mutual expectation alignment (see also Pickering & Garrod, 2004 although the current account is considerably more specific), thereby contributing to the minimization of the joint effort experienced by interlocutors (cf. Clark, 1996). Among other things, the perspective outlined here also offers an explanation as to why syntactic alignment between interlocutors seems to facilitate both faster production (Ferreira, Kleinman, Kraljic, & Siu, 2011) and better communication (Reitter & Moore, 2007) – because it reduces the prediction error experienced during both production and comprehension.

This brings us to the prediction we aim to test: the relative increase in the strength of syntactic priming should increase as a function of the prediction error experienced while processing the prime structure. This prediction is in principle compatible with other implicit learning accounts that commit neither to specific forms of learning nor to specific aims of learning (e.g., Bock & Griffin, 2000; Kaschak, 2007). Competing accounts that take syntactic priming to be solely a consequence of short-term activation boosts of recently processed representations do not share this prediction (e.g., Dubey, Keller, & Sturt, 2008; Kaschak & Glenberg, 2004a; Pickering & Branigan, 1998). A recently proposed third type of account, hybrid accounts, that evoke both short-term activation boosts and unsupervised implicit learning mechanisms, require additional assumptions to accommodate our prediction (Reitter et al., 2011). In the interest of brevity, we postpone the discussion of these accounts to the general discussion.

Preliminary evidence in line with the prediction that syntactic priming in production is sensitive to the prediction error comes from the observation that less frequent structures tend to prime more strongly (the 'inverse frequency' or 'inverse preference' effect, e.g., Bock, 1986; Ferreira, 2003; Hartsuiker & Kolk, 1998; Kaschak, Kutta, & Jones, 2011; Scheepers, 2003). For example, English passives, which are hugely less frequent than active structures, elicit reliable syntactic priming whereas active structures exhibit almost no detectable priming effect (Bock, 1986). Since frequency tends to be associated with processing times (e.g., Rayner & Duffy, 1988), it is natural to assume that frequency is inversely correlated with the prediction error experienced while processing a structure. Under this assumption, the inverse frequency effect provides evidence that syntactic priming is sensitive to the prediction error.

The inverse frequency effect is, however, only a weak test of the prediction that we are interested in. If language processing is indeed inescapably tied to learning, the relevant prediction error should be *context-dependent*: there is broad agreement that the processing difficulty associated with a word or structure is affected by its *contextual* predictability (e.g., Altmann & Kamide, 1999; DeLong et al., 2005; Garnsey et al., 1997; Kamide, Altmann, & Haywood, 2003; MacDonald et al., 1994; McDonald & Shillcock, 2003;

Staub & Clifton, 2006; Trueswell et al., 1993). Some even hold that contextual expectations are the primary source of processing difficulty compared to frequency effects (Smith & Levy, 2008).

Hence, the prediction made here and by the error-based model implies that the strength of syntactic priming in language production is a function of the prediction error given context-dependent expectations based on both prior and recent experience. We test these predictions against data from conversational speech and experiments on written as well as spoken language production. The first of these two predictions (the effect of prediction errors based on prior experience) is also investigated in a recent study on the Dutch ditransitive alternation by Bernolet and Hartsuiker (2010). We return to their results below.

### 1.3. Overview of studies

The studies presented below investigate syntactic priming in the dative alternation (also called ditransitive alternation, e.g., Bresnan, Cueni, Nikitina, & Baayen, 2007). In the dative alternation, speakers choose between two near-meaning equivalent syntactic variants, as exemplified in (1a) and (1b). In the double object (DO) variant the verb (*give* below) is followed by two noun phrase arguments, with the *recipient* argument (*a country*) preceding the *theme* argument (*money*). In the prepositional object (PO) variant, the dative verb is followed by the theme and then a prepositional phrase, the argument of which is the recipient noun phrase.

- 
- (1) a. **Double object (DO):** *We give a country money.*  
 b. **Prepositional object (PO):** *We give money to a country.*
- 

Study 1 investigates the effect of the prime's prediction error on the strength of syntactic priming in conversational speech (based on data from Bresnan et al. (2007) and Recchia (2007)). Study 2 investigates syntactic priming due to written primes in written sentence completion (based on data from Kaschak (2007) and Kaschak & Borreggine (2008)). Study 3 investigates syntactic priming due to spoken primes in spoken picture description. Studies 1 and 2 employ estimates of the prediction error based on the prior 'average' language experience that speakers have before entering the context in which we assess syntactic priming. These studies hence follow the majority of previous work in employing estimates of prior language experience that are neither individualized nor sensitive to the statistics of the current environment (for some exceptions, see Fine & Jaeger, 2011; Fine et al., submitted for publication; MacDonald & Christiansen, 2002; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). In this sense, these studies assess the effect of the prediction error based on *prior* experience.

However, if there is indeed life-long continuous adaptation, we would expect the prediction error associated with



a syntactic prime to depend on all experience prior to a sentence, including, in particular, recent experience with the statistics of the current environment. For example, in the error-based model the same structure might be associated with different prediction errors, depending on what has recently been processed, since everything that is processed leads to changes in the model's weights. Given that expectation adaptation in comprehension seems to be rather rapid (Fine et al., submitted for publication; Kaschak & Glenberg, 2004b), changes to the prediction error might be detectable within a single experiment. If so, it should be possible to detect changes to the strength of syntactic priming based on recent experience within an experiment. Study 3 tests this prediction by manipulating the sequential distribution of syntactic structures participants are exposed to during the experiment, thereby manipulating the prediction error based on recent experience. To the best of our knowledge, this is the first study that investigates effects on syntactic priming as a function of the prediction error based on recent experience.

#### 1.4. Operationalizing the prediction error

Before we describe the experiments, one last piece of background information is necessary. In order to assess the effect of prediction errors on syntactic priming, we need to operationalize the prediction error. Here, we estimate the prediction error associated with processing a prime as the surprisal of the prime structure (see also Jaeger & Snider, 2008; Malhotra, 2009, p. 185). The surprisal of a linguistic unit (e.g., a word or syntactic structure) in context is equivalent to the amount of Shannon information it adds to the preceding context,  $\log \frac{1}{p(\text{unit}|\text{context})} = -\log p(\text{unit}|\text{context})$  (Shannon, 1948). As should be the case for a measure of prediction error, surprisal is larger, the less expected the encountered word or structure is given previous experience. Specifically, a prime's surprisal is 0 when its structure is perfectly expected ( $p(\text{unit}|\text{context}) = 1$ ), and increases the less expected the structure. We chose surprisal rather than other measures of 'unexpectedness' because surprisal has been found to be a good predictor of word-by-word processing times in self-paced reading and eye-tracking reading experiments (Smith & Levy, 2008; see also Alexandre, 2010; Boston, Hale, Kliegl, Patil, & Vasisht, 2008; Demberg & Keller, 2008; Frank & Bod, 2011).

Interestingly, surprisal as an estimate of the prediction error also follows naturally from the error-based model. In the words of Chang and his colleagues, the prediction error associated with comprehending a word is calculated as follows:

The word ... units used a soft-max activation function, which passes the activation of the unit through an exponential function and then divides the activation of each unit by the sum of these exponential activations for the layer. ... Error on the word units was measured in terms of divergence –  $\sum t_i \log \frac{t_i}{o_i}$  – where  $o_i$  is the activation for the  $i$  output unit on the current word and  $t_i$  is its target activation because of the soft-max activation function. (Chang et al., 2006, p. 270)

Although Chang and colleagues describe the prediction error with reference to activation, the distribution of relative exponentiated activation over word units at each point in time also characterizes a probability distribution (each value falls in  $[0, 1]$  and the values sum to 1). With this in mind, it becomes apparent that the error signal employed in the error-based model is the Kullback–Leibler divergence or relative entropy between the desired – i.e., actually observed – output distribution over words in the mental lexicon (which is 1 for the actual word and 0 for all other words) and the distribution of word probabilities given previous experience. This is, of course, no coincidence since the relative entropy is a measure of the divergence between two probability distributions and hence constitutes a rational choice for the prediction error that is sought to be minimized via backpropagation (see also MacKay, 2003, chap. 5).

Since  $t_i$  is 0 for all but the actual word, the equation for the error signal reduces to  $1 * \log \frac{1}{o_i} = -\log o_i$ , where  $o_i$  is the network's estimate of  $p(\text{word}|\text{previous experience})$ . Hence, the error used to adjust the network's weights after processing a word is identical to the word's surprisal given the network's weights before encountering the word. In the general discussion, we will return to the relation between our operationalization of the prediction error and the error signal used in the error-based model.

## 2. Study 1: Prediction error and syntactic priming in conversational speech

Study 1 assesses the effect of a prime's prediction error on the strength of syntactic priming against data from the dative alternation in spontaneous speech that was generously provided by Bresnan et al. (2007). As an example, take a verb like *owe*, which is highly biased towards DO:

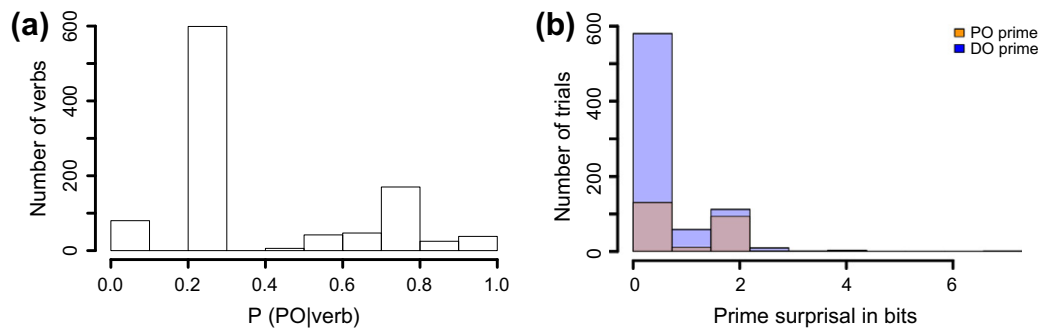
- 
- |     |   |
|-----|---|
| (2) | ... you don't owe your city anything... |
| (3) | ... you owe that back to God ...        |
- 

The prediction we are testing is that processing a prime structure that is unexpected given the prime's verb, as in (3), will make speakers more likely to produce the structure later (in the target) than if the prime's verb is biased towards the structure, as in (2).

### 2.1. Methods

#### 2.1.1. Data

The database provided by Bresnan et al. (2007) contains 2,349 instances of the dative alternation from the full Switchboard corpus (about 2 million words, Godfrey, Holiman, & McDaniel, 1992). Since we were interested in how the prediction error associated with a prime affects the strength of syntactic priming, we excluded all cases without preceding primes (i.e., the first dative in each conversation). This reduced the database to 1249 tokens. We further removed all cases with verbs for which it was unclear whether they participate in the dative alternation (i.e., for which the verb occurred fewer than 5 times in



**Fig. 1.** Histograms of prime verbs' PO biases (a) and prior prime surprisal (b) in Study 1. Note that verb biases are independent of the actual prime structure (PO or DO), whereas prime surprisal is the surprisal of the prime's structure given the prime's verb bias.

either the PO or DO). This left a total of 1007 target productions (and 25 target verb types, of the original 38 in the database), comprising 234 PO and 773 DO structures.

### 2.1.2. Estimating the prediction error (prime surprisal)

Prime surprisal was estimated as the negative log of the conditional probability of the prime structure given the verb: for DO primes, the surprisal is  $-\log_2 p(\text{DO}|\text{verb})$ ; for PO primes, the surprisal is  $-\log_2 p(\text{PO}|\text{verb})$ . Hence, the larger the prime surprisal is, the greater the difference between the expected structure (given the verb) and the observed structure in the prime. A large number of studies have shown that comprehenders are sensitive to verb subcategorization biases and these biases affect what structure comprehenders expect to follow the verb (e.g., Garnsey et al., 1997; MacDonald, Bock, & Kelly, 1993; Trueswell et al., 1993; specifically, for the dative alternation, see Brown, Savova, & Gibson, 2012).

Our estimates of subcategorization frequencies of the prime verbs were based on the full Switchboard corpus. Hence, prime surprisal in Study 1 is based on an estimate of the average prior language experience of the speakers recorded in the Switchboard corpus. The database contains a range of verb biases, as attested in Fig. 1a. Fig. 1b shows the distribution of prior prime surprisal based on the actual prime structures occurring in the database. Note that most primes had rather low surprisal values. This unsurprising since the primes for Study 1 came from conversational speech, rather than a balanced psycholinguistic experiment: in real life, verbs most often occur with the structures that they are biased towards.

### 2.1.3. Analysis

In a mixed logit regression (for an introduction, see Jaeger, 2008), we analyzed the occurrence of PO vs. DO structures based on the prior prime surprisal associated with the most recent prime, while controlling for (a) effects on speakers' preferences in the dative alternation found in previous work (Bresnan et al., 2007; Gries, 2005), (b) additional predictors specific to syntactic priming (added to this database by Recchia (2007)).

We included twelve properties of the target sentence as control predictors based on previous work by Bresnan et al. (2007) and Gries (2005). These are listed in the first block of Table 1. Bresnan and colleagues found that the accessi-

bility and complexity of both the theme and the recipient expression affect speakers' preferences, with more accessible and less complex phrases usually being ordered before less accessible and more complex phrases (see also Arnold, Wasow, Losongco, & Ginstrom, 2000; Bock & Warren, 1985). Beyond properties of the theme and recipient, both the semantic class of the target verb and the target verb's subcategorization bias (here operationalized as the log-odds of a PO structure given the verb in the Switchboard corpus) were found to account for unique proportions of the variance in predicting speakers' preferences (Bresnan et al., 2007; Gries, 2005). Here we modeled semantic class as a binary distinction, indicating whether the dative verb in the target sentence was a 'prevention of possession' verb (as in "cost/deny the team a win") or not.<sup>4</sup>

Seven additional control predictors were included to account for effects of syntactic priming identified in previous work. These controls are shown in the second block of Table 1. This includes the basic effect of syntactic priming (i.e., whether the most recent prime was a PO or DO structure) and the cumulative effect of all preceding primes (Kaschak, 2007; Kaschak, Loney, & Borreggine, 2006). Since previous work has found stronger priming effects if the prime and target share a verb lemma (the 'lexical boost', e.g., Hartsuiker et al., 2008; Pickering & Branigan, 1998; Szmrecsányi, 2005), both the main effect of verb repetition and its interaction with the prime structure were included. Similarly, we included the distance to the most recent prime (in words) and its interaction with the prime structure to control for potential decays in the strength of the priming effect (Reitter, Moore, & Keller, 2006). Finally, we included the 3-way interaction since recent work suggests that the lexical boost on syntactic priming, but not

<sup>4</sup> Bresnan et al. (2007) distinguish between five semantic classes of the verb: abstract, as in "give it some thought"; transfer of possession, as in "give/send someone an armband"; future transfer of possession, as in "owe/promise someone some money"; prevention of possession, as in "cost/deny the team a win"; and communication as in "tell/give me your name". We model semantic class as a binary factor because, in the smaller dative priming database, not all the semantic classes were significant predictors of the dative alternation, and we want to keep the number of degrees of freedom in the model as low as possible given the smaller data set and large number of control factors in the model. This coding was chosen because further analyses revealed that none of the other contrasts was significant on the reduced data set employed in Study 1.

**Table 1**

Summary of Study 1 results ( $c$  = centered,  $r$  = residualized; see text for details). For each effect, we report the coefficient estimate (in log-odds and odds), its standard errors and two tests of significance: Wald's  $Z$  statistic, which tests whether the coefficients are significantly different from zero (given the estimated standard error), as well as the  $\chi^2$  over the change in data likelihood,  $\Delta(-2A)$ , associated with the removal of the unresidualized predictor from the final model. Finally, we report the partial Nagelkerke  $R^2$  as a measure of effect size (cf. Jaeger, 2010, footnote 2).

Predictor (independent variable)	Parameter estimates			Wald's test		$\Delta(-2A)$ -test		Partial pseudo- $R^2$
	Log-odds	S.E.	Odds	Z	$p_z$	$\chi^2$	p	
<i>Properties of target sentence</i>								
Theme pronominal (r)	2.25	0.39	9.5	5.7	$\ll 0.001$	18.2	$\ll 0.001^*$	
Theme given (r)	1.32	0.45	3.7	2.9	$< 0.005$	13.8	$\ll 0.001^*$	0.047
Theme indefinite	-2.41	0.41	0.1	-5.9	$\ll 0.001$	35.6	$\ll 0.001^*$	0.112
Theme singular	-0.83	0.38	0.4	-2.2	$< 0.05$	5.6	$< 0.05^*$	0.017
Recipient pronominal	-0.54	0.50	0.6	-1.1	$> 0.2$	1.2	$> 0.2$	0.002
Recipient given (r)	-2.53	0.58	0.1	-4.4	$\ll 0.001$	24.9	$\ll 0.001^*$	0.081
Recipient indefinite	0.27	0.53	1.3	0.5	$> 0.6$	0.3	$> 0.6$	0.001
Recipient inanimate	3.29	0.58	26.8	5.7	$\ll 0.001$	33.4	$\ll 0.001^*$	0.104
Recipient third person	0.39	0.41	1.5	0.9	$> 0.3$	1.1	$> 0.3$	0.004
Log argument length difference	-2.55	0.23	0.1	-10.9	$\ll 0.001$	66.7	$\ll 0.001^*$	0.193
Verb class = 'prevention of possession'	-3.75	2.50	0.1	-1.6	$> 0.11$	2.4	$> 0.12$	0.017
Target verb bias (log-odds)	4.45	0.63	85.6	6.9	$\ll 0.001$	117.3	$\ll 0.001^*$	0.302
<i>Basic priming effects</i>								
PO prime (c)	0.56	0.33	1.8	1.7	$< 0.1$	3.2	$< 0.1^*$	0.004
Cumulative PO primes	0.29	0.23	1.3	1.2	$> 0.2$	1.8	$> 0.2$	0.005
Prime-target verb identity (c)	0.01	0.34	1.0	0.1	$> 0.9$	1.4	$> 0.3$	0.004
Log prime-target distance (c,r)	0.02	0.10	1.0	0.2	$> 0.8$	0.2	$> 0.9$	0.001
PO prime $\times$ verb identity (r)	1.04	0.68	2.8	1.6	$> 0.10$	2.6	$> 0.10$	0.009
PO prime $\times$ distance (r)	-0.17	0.21	0.8	-0.4	$> 0.4$	1.4	$> 0.4$	0.002
PO prime $\times$ verb identity $\times$ distance (r)	-0.24	0.43	0.8	-0.8	$> 0.5$	0.4	$> 0.5$	0.001
<i>Effect of prime's prediction error</i>								
Prior prime surprisal given the prime verb (c,r)	-0.07	0.20	0.9	0.4	$> 0.7$	0.2	$> 0.6$	0.001
PO prime $\times$ prior prime surprisal (r)	0.81	0.40	2.2	2.0	$< 0.05$	4.3	$< 0.05^*$	0.012

syntactic priming itself, decay over time (Hartsuiker et al., 2008). The model also contained a by-speaker random intercept, to control for individual differences in speakers' preference for PO over DO structures.

Here and in all studies reported below, all predictors were centered and all interactions were residualized against their main effects in order to reduce collinearity in the models. Additionally, prime-target distance was residualized against verb identity since the two predictors were correlated. Prime surprisal was residualized against the main effect of prime structure. In the analysis reported below, all fixed effect correlation  $r$ s  $< 0.25$ . We report the full model, which is generally recommendable whenever one has sufficient data (for guidelines, see references in Jaeger, 2011) because of the inherent problems of stepwise predictor removal (for a concise overview, see Harrell, 2001). Removing predictors that have  $p > 0.7$  (a frequently proposed threshold, see Harrell, 2001) does not change the results of any of our studies.

## 2.2. Predictions

Under the assumption that the prediction error experienced while processing a prime can be approximated as prime surprisal given the prime's verb, we predict that more surprising primes should lead to larger priming effects.

Specifically, we predict an interaction between the prime structure (PO vs. DO) and prime surprisal given the verb ( $-\log_2 p(\text{PO}|\text{verb})$  for PO primes and  $-\log_2 p(\text{DO}|\text{verb})$  for DO primes). Fig. 2a illustrates this

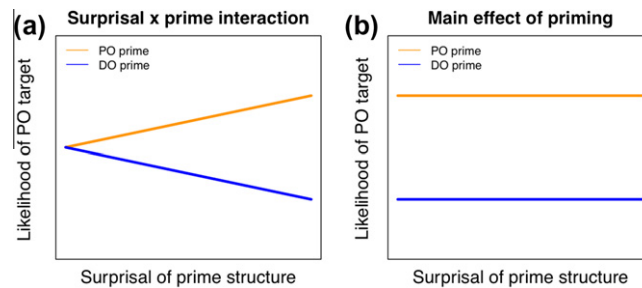
prediction. For PO primes, the PO structure should be more likely to be produced in the target the more surprising (less expected) the PO prime was. This is illustrated by the positive slope for a PO prime in Fig. 2a. Conversely, for DO primes, DO targets should be more likely the more surprising it was to see a DO prime. This is illustrated by the negative slope in Fig. 2a. Activation-boost accounts, on the other hand, do not predict any effect of prime surprisal (Fig. 2b).

## 2.3. Results

Table 1 summarizes the effects of control predictors and the predictors of interest, both of which we discuss below. Coefficients are given in log-odds (the space in which logit models are fitted to the data). Significant positive coefficients indicate increased log-odds (and hence increased probabilities) of a PO structure in the target sentence.

### 2.3.1. Non-priming controls

As shown in the first block of Table 1, all control predictors replicated previous findings, with effects pointing in the same direction as in Bresnan et al. (2007): more accessible and less complex themes favored the PO structure, whereas more accessible and less complex recipients favored the DO structure. Verbs that overall favor the PO structure did so also in our sample. The coefficients as well as significance levels of all control predictors are only minimally different from Bresnan et al. (2007).



**Fig. 2.** Illustration of the predicted interaction of prime structure and prime surprisal (a) as contrasted with the main effect of prime structure (b) expected if syntactic priming is unaffected by the prediction error.

### 2.3.2. Basic priming effects

The second block of rows of Table 1 summarize the priming-related controls. There was a marginally significant main effect of priming ( $p_z < 0.1$ ). There was no main effect of prime-target distance ( $p_z > 0.9$ ), nor a main effect of repeating the verb in the prime ( $p_z > 0.7$ ). There also was no main effect of the cumulative proportion of PO in the dialogue ( $p_z > 0.1$ ). None of the interactions reached significance ( $p_{zs} > 0.1$ ).

### 2.3.3. Prior prime surprisal

As predicted, we found a significant interaction between the prime structure and prior prime surprisal given the verb ( $p_z < .05$ ). There was no main effect of prior prime surprisal ( $p_z > 0.6$ ). The positive interaction indicates that the slope of the surprisal effect in PO primes is significantly greater than the slope of surprisal in DO primes. The interaction is illustrated in Fig. 3, which reflects the predictions illustrated in Fig. 2. Surprisal has a positive trend in PO primes (the more surprising a PO prime, the more likely the target is to be a PO), but a negative trend in DO primes (the more surprising the DO structure, the more likely it is to be repeated, which means a PO is *less* likely in the target). A simple effect analysis revealed that the effect of surprisal was significant for PO primes ( $p_z < .05$ ). The simple effect of surprisal for DO primes did not reach significance, although the numerical effect had the expected negative trend ( $p_z > 0.2$ ).

## 2.4. Discussion

We find that, the more surprising a prime structure is given the subcategorization bias of the prime's verb, the more likely this structure is to be repeated in the target. This sensitivity of syntactic priming to the prediction error is predicted by the hypothesis that syntactic priming is a consequence of adaptation with the goal to minimize the expected future prediction error. It also follows from the error-based model (Chang et al., 2006) and is broadly compatible with less specific implicit learning accounts (e.g., Bock & Griffin, 2000; Chang et al., 2000). Short-term activation boost accounts do not predict this effect (Pickering & Branigan, 1998).

Interestingly, the simple effect of surprisal reached significance only for PO primes, but failed to reach significance for DO primes. It is theoretically possible that this

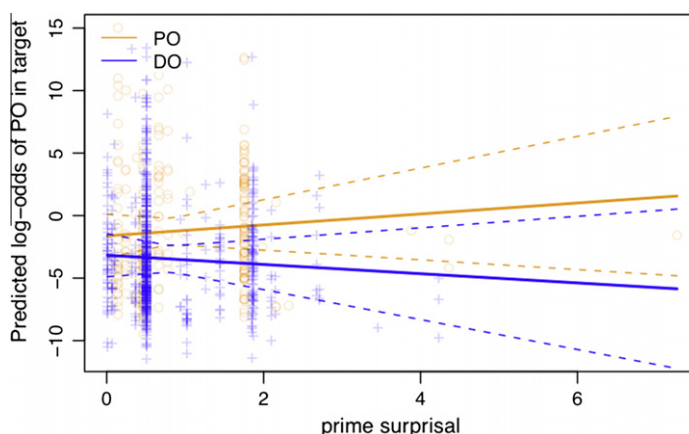
is due to a floor effect for DO primes due to the fact that DO target structures were overall more frequent in the database (773 DOs vs. 234 POs; see also Fig. 3, where the overall the log-odds of a PO structure are smaller than 0). However, Studies 2 and 3 reported below argue against this interpretation.

Another explanation of the observed asymmetry is that priming effects are overall weaker for DO structures, perhaps *because* DO structures are more frequent (cf. the inverse-frequency effect, Hartsuiker & Kolk, 1998; Hartsuiker & Westenberg, 2000; Kaschak, 2007; Scheepers, 2003). Kaschak (2007) found weaker priming effects for DO than PO primes in a series of production experiments. To test whether the same holds for the current data, we examined the main effect of prime structure against the entire data set, including unprimed turns (2349 target tokens). This provides a baseline against which to compare both DO and PO primes. Replicating Kaschak (2007) for the Switchboard corpus, we found that PO priming is stronger ( $p_z > 0.01$ ) than DO priming – in fact, the simple effect of priming is non-significant for DO primes ( $p_z > 0.6$ ). It is hence possible that the failure to find a significant simple effect of prime surprisal for DO primes is due to the overall insignificance of DO priming. We will return to this question below, since – to anticipate the results of our remaining studies – we find the same asymmetry between PO and DO priming in Studies 2 and 3.

We find no evidence that syntactic priming decays (replicating Bock & Griffin, 2000; Bock, Dell, Chang, & Onishi, 2007; Snider, 2008; but unlike Branigan, Pickering, & Cleland, 1999; Reitter et al., 2006). The speakers in the corpus were equally likely to repeat dative structures used previously in the conversation no matter how many words had been subsequently uttered in the conversation. This is remarkable considering the median prime-target distance was 148 words, with a maximum distance of 2000 words.

Contrary to the prediction of some implicit learning accounts, we found no cumulative effect of primes beyond the most recent prime (Bock & Griffin, 2000; Kaschak, 2007). This is most likely due to the fact that there was little variation in the number of primes preceding target structures: the median number of datives per dialogue in our database is 2. Studies 2 and 3, where we find clear cumulative effects of syntactic priming, support this interpretation.





**Fig. 3.** Interaction between prior prime surprisal (in bits) and priming in Study 1. Circle and crosses indicate the log-odds for PO targets according to the model. Orange circles represent PO primes, and blue crosses represent DO primes. Dotted curves represent the 95% confidence intervals of the slopes.

Finally, the interaction between prime structure and verb repetition did not reach significance, though it was near-marginal and in the predicted direction ( $p_z = 0.11$ ). We hypothesize that the distribution of verb lemmas across primes did not provide sufficient power: while verb repetition between prime and target was common in our database (46.7% of tokens), 74.5% of the verb repetitions involve the verb *give*, leaving only 120 cases with 17 types of repeated verbs. This made it difficult to distinguish between effects of the target verb bias and effects of verb repetition.

### 3. Study 2: Prediction error and syntactic priming in written production

Study 1 provides evidence from conversational dialogue data that larger prediction errors while processing the syntactic structure of the prime lead to larger syntactic priming effects. Study 2 extends our investigation to written production. We test whether the effect of prior prime surprisal on prime strength is also observed under more controlled conditions in laboratory production experiments. We present a meta-analysis of data from three priming experiments generously provided by Mike Kaschak and colleagues (Kaschak, 2007; Kaschak & Borreggine, 2008).

#### 3.1. Methods

##### 3.1.1. Data

The data set consisted of 4508 prime and 1703 target trials from three production experiments on syntactic priming (a total of 392 participants and 18 target items). All experiments employed written sentence completion: participants completed a partial sentence presented on a computer by typing into a text box. Prime stimuli contained a subject, verb, and either an animate recipient (to induce a DO) or an inanimate theme (to induce a PO). Targets contained only a subject and verb to allow for

either PO or DO structures to be produced. All non-dative target completions were excluded from the analysis.

The design of the experiments is summarized in Table 2. The three experiments were designed to investigate cumulative priming. During an exposure phase, participants saw 16–20 primes, which consisted of 100%/0%, 75%/25%, 50%/50%, 25%/75%, or 0%/100% PO and DO primes, respectively. The cumulative effect of these primes was assessed in the test phase. The test phase of Experiment 1 from Kaschak and Borreggine (2008) consisted of 10 prime–target pairs, while the test phases of Experiments 1 and 2 from Kaschak (2007) consisted only of target completions (6 and 12 completions, respectively).

In prime trials, participants saw stems that enforced either a PO or a DO completion (e.g., *Meghan gave the toy ...* vs. *Meghan gave her mom ...*). In target trials, participants saw stems that could be continued with either PO or DO structures (e.g., *Meghan gave ...*; Kaschak, 2007, p. 928). We assessed the effect of the prediction error on the strength of priming in the target trials.

##### 3.1.2. Estimating prime surprisal

Although the original experiments did not manipulate prime surprisal, the prime trials in all three experiments employed a sufficiently large variety of verbs, making it possible to examine the effects of prime surprisal in a meta-analysis. Prime surprisal was operationalized following Study 1. Since the data set for the meta-analysis contained some verbs that were not in the Switchboard corpus, we estimated the PO bias of all prime and target verbs based on the subcategorization frequencies from the database in Roland, Dick, and Elman (2007); to obtain reliable estimates, we combined the subcategorization estimates from all corpora.

Hence, like in Study 1, our estimate of prime surprisal in Study 2 is an estimate of prior prime surprisal was based on an estimate of the average previous language experience. Unlike in Study 1, ‘previous language experience’ for Study 2 had to be approximated by averaging over

**Table 2**

Summary of experiments included in the meta-analysis for Study 2.

	Kaschak and Borreggine (2008) Expt. 1	Kaschak (2007) Expt. 2	Kaschak (2007) Expt. 1
Training phase	100%DO/0%PO or 0%DO/100%PO or 50%DO/50%PO or		100%DO/0%PO or 0%DO/100%PO or 50%DO/50%PO or 75%DO/25%PO or 75%DO/25%PO
Testing phase	20 primes 10 prime-target pairs	16 primes 12 targets	20 primes 6 targets

corpora consisting of a more heterogeneous mix of genres, styles, and including both written and spoken data. More importantly, the estimate of previous language experience in Study 2 also does not include data from the participants in the Kaschak and colleagues' experiments and hence might be a less accurate estimate of the actual surprisal experienced by participants in the experiments than that employed in Study 1. This biases against finding the predicted interaction of prime structure and prime surprisal. The distribution of PO bias and prior prime surprisal is summarized in Fig. 4a and b, respectively.

The surprisal values for two verbs (*throw* and *sell*) in the Kaschak (2007) experiments were outliers in that they were much higher than the other other verbs: they were more than 3 standard deviations higher than the mean. These accounted for only 2.2% of target trials, which were excluded from the analysis, leaving 1665 trials for the analysis, thereof 900 PO and 765 DO structures.

### 3.1.3. Analysis

We again employed a mixed logit regression to predict the occurrence of PO over DO structures based on the surprisal associated with the most recent prime. Following Study 1, we included the PO bias of the target verb as a control predictor. Except for prime-target distance, which did not vary in the experiments, we also included all priming-related controls from Study 1. We coded as the most recent prime whatever trial immediately preceded the target trial, regardless of whether this was a target or prime trial. Similarly, the cumulative count of preceding prime structures included both prime and target trials.

Finally, both by-participant and by-item random effects were included in the analysis. Model comparison was conducted to assess which random effects were significant. Since all results presented in this paper hold both for the

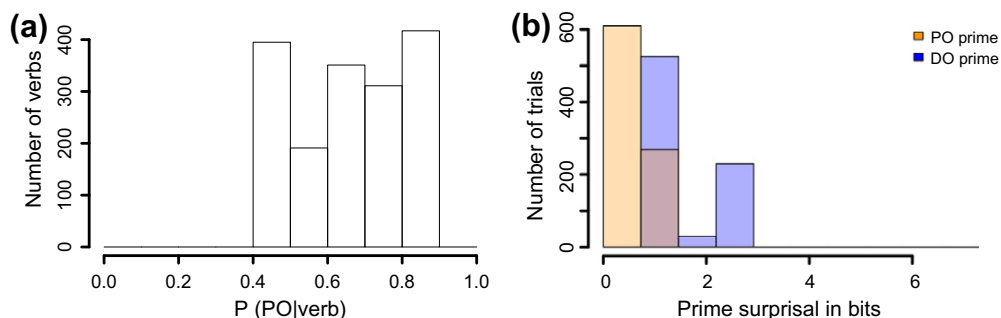
full random effect structure – all intercepts, all slopes, and all covariances – and for the maximal random effect structure justified by model comparison, we simply present models with full random effect structures. That is, we present the results from the most conservative analysis (see Appendix B for a summary of the random effect structure). All fixed effect correlation  $r_s < 0.3$ . The predictions are the same as in Study 1.

### 3.2. Results

The results of Study 2 are summarized in Table 3. Here, this is only the subcategorization bias of the target verb, which has the expected effect. The second block summarizes the basic priming effects and the final block the effects of interest. Next, we summarize these effects.

#### 3.2.1. Basic priming effects

There was a main effect of prime structure such that PO primes were likely to lead to PO targets ( $p_z \ll 0.001$ ). There was no main effect of verb repetition between prime and target ( $p_z > 0.6$ ), but, consistent with previous findings, there was an interaction with prime structure such that the prime structure is more likely to be repeated in the target when the verbs are the same ( $p_z < .005$ ). There also was a main effect of cumulative priming such that the more POs participants had encountered, the more likely they were to produce a PO ( $p_z \ll 0.001$ ). In short, both the cumulative priming effect observed in the original experiments (Kaschak & Borreggine, 2008; Kaschak, 2007) and the verb repetition effect observed in previous work (Hartsuiker et al., 2008; Pickering & Branigan, 1998; Szmrecsányi, 2005) replicated.

**Fig. 4.** Histograms of prime verbs' PO biases (a) and prior prime surprisal (b) in Study 2.

**Table 3**

Summary of Study 2 results (c = centered, r = residualized; for further information see the caption of Table 1).

Predictor (independent variable)	Parameter estimates			Wald's test		$\Delta(-2.1)$ -test		Partial pseudo- $R^2$
	Log-odds	S.E.	Odds	Z	$p_z$	$\chi^2$	p	
<i>Properties of target sentence</i>								
Target verb bias (log-odds)	0.90	0.20	2.5	4.5	$\ll 0.001$	17.0	$\ll 0.001^*$	0.015
<i>Basic priming effects</i>								
PO prime (c)	1.08	0.15	2.9	6.9	$\ll 0.001$	33.0	$\ll 0.001^*$	0.033
Cumulative PO primes	0.27	0.06	1.3	4.2	$\ll 0.001$	13.4	$< 0.005^*$	0.014
Prime-target verb identity (c)	0.09	0.25	1.1	0.4	$> 0.7$	0.2	$> 0.9$	0.001
PO prime $\times$ verb identity (r)	1.20	0.37	3.3	3.2	$< 0.005$	9.6	$< 0.001^*$	0.009
<i>Effect of prime's prediction error</i>								
Prior prime surprisal given the prime verb (c,r)	0.07	0.13	1.1	0.5	$> 0.5$	0.5	$> 0.4$	0.001
PO prime $\times$ prior prime surprisal (r)	0.82	0.32	2.3	2.5	$< 0.05$	6.0	$< 0.05^*$	0.005

### 3.2.2. Prior prime surprisal

Like in Study 1, we found the predicted significant interaction between prime structure and prior prime surprisal given the verb ( $p_z < .05$ ; if the tokens of verbs *throw* and *sell* are not excluded, the results are the same except the surprisal interaction would be  $p_z = .05$ ). There was no main effect of prior prime surprisal ( $p_z > 0.5$ ). The interaction is illustrated in Fig. 5. As in Study 1, simple effect analyses revealed that the effect of prime surprisal was significant for PO primes ( $p_z < 0.05$ ), whereas the effect did not reach significance for DO primes ( $p_z > 0.4$ , although it is in the predicted negative direction).

### 3.3. Discussion

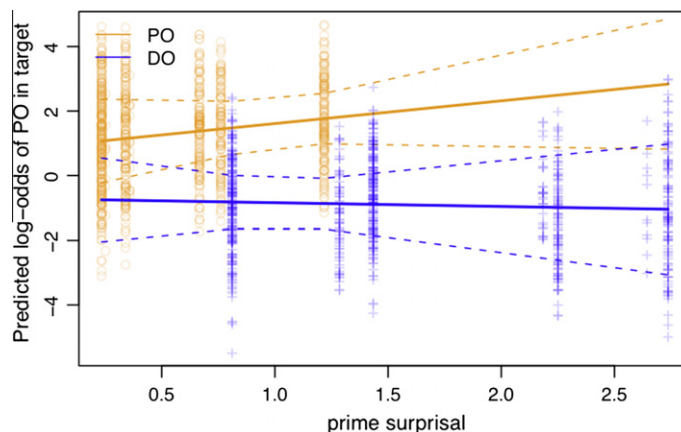
Study 2, which employed written production data, replicates the effect observed for conversational data in Study 1. In both studies, we find that priming is sensitive to the magnitude of the prediction error. The more surprising a prime structure is given the prime verb, the more likely this structure is to be repeated in the target.

As in Study 1, we find the prediction confirmed only for PO primes, where the effect of prime surprisal is not significant for DO primes. Unlike in Study 1, PO and DO production were about equally likely in Study 2, so that this lack

of an effect for DO primes is unlikely to be due to a floor effect. The data are, however, compatible with the second explanation advanced in our discussion of Study 1: it is possible that there is overall very little DO priming so that it is hard to detect effects of surprisal. We postpone further discussion of this issue until after Study 3.

Taken together, Studies 1 and 2 provide evidence from conversational speech and from written production that syntactic priming is sensitive to the prediction error associated with the processing of the prime structure. This, as we laid out in the introduction, is expected if syntactic priming is the consequence of mechanisms that aim to align interlocutors' expectations, thereby reducing the prediction errors on upcoming sentences. If this is indeed the correct interpretation of the effects observed here, then we would expect prediction errors – and hence syntactic priming – to be sensitive to *current* expectations. For example, if speakers aim to align with their interlocutors, they should be sensitive to that specific interlocutor's productions.

Indeed, as outlined in the introduction, there is evidence that *comprehension* is exquisitely sensitive to the statistics of the current linguistic environment. Most of this evidence comes from lower-level perceptual and phonetic processes. For example, the processing difficulty initially experienced with non-native or non-standard



**Fig. 5.** Interaction between prior prime surprisal (in bits) and priming in Study 2. Orange circles represent PO primes, and blue crosses represent DO primes. Dotted curves represent the 95% confidence intervals of the slopes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

pronunciations can dissipate after only a few minutes of exposure (Bradlow & Bent, 2008; Maye, Aslin, & Tanenhaus, 2008). Perceptual adaptation of phonetic categories has been observed after exposure to only a handful of auditorily ambiguous but visually or lexically disambiguated instances of a sound (e.g., Vroomen, van Linden, de Gelder, & Bertelson, 2007; van Linden & Vroomen, 2007). Evidence that this type of adaptation can be understood as expectation adaptation comes from more recent work that has found perceptual recalibration to be sensitive to the relative *distribution* of phonetic cues in recent experience (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2012; Sonderegger & Yu, 2010). More recent work has also revealed evidence for similarly rapid expectation adaptation during syntactic comprehension (Farmer, Monaghan, Misyak, & Christiansen, 2011; Fine et al., submitted for publication; Kaschak & Glenberg, 2004b).

Very little is known about such processes in production. Here we begin to address this question for sentence production (for phonological production, see Warker & Dell, 2006). We ask whether the prediction error that affects syntactic priming in *production*, as observed in Studies 1 and 2, is based on recent experience with the current linguistic environment. In short, is the strength of syntactic priming in production sensitive to how unexpected the prime structure was given the distribution of syntactic structures preceding the prime? Studies 1 and 2 do not speak to this question since we estimated the prediction error based on verb's subcategorization frequencies as observed in corpora. The prime surprisal estimate in Studies 1 and 2 therefore was an estimate of the surprisal based on *prior* experience (here: experience that is not environment- or speaker-specific). Studies 1 and 2 therefore leave open the question as to whether syntactic priming in production is sensitive to the prediction error as it is experience during the processing of a prime sentence. The primary purpose of Study 3 is to address this question.<sup>5</sup> A secondary motivation was to test whether the effect of the prediction error based on prior experience that was observed in Studies 1 and 2 would replicate in an experiment explicitly designed to test this hypothesis.

#### 4. Study 3: Prediction error depending on prior and recent experience

Study 3 investigates the effect of both prior and recent experience on syntactic priming in a picture description paradigm under the guise of a memory task. Specifically, the design of Study 3 crossed prime structure (PO or DO) with the prediction error based on prior experience and the prediction error based on recent experience within the experiment. To increase the power to detect effects of prior surprisal, we include 48 different verbs in the experiment spanning a wide range of subcategorization biases

(cf. Study 1, which contained 25 verbs, of which only 9 occurred more than 20 times, and Study 2, which contained only 7 verbs). The prediction error based on recent experience was manipulated by either alternating prime structures throughout the experiment (PO–DO–PO–DO–... or vice versa) or blocking prime structures, so that all prime trials within the first half of the experiment used one prime structure and the prime trials in the remainder of the experiment the other prime structure (PO–PO–...–PO–DO–DO–...–DO or vice versa). Under the assumption that participants continuously adapt their syntactic expectations after each prime, participants' expectations for a PO or DO prime structure should depend, not only on prior experience, but also on how many PO and DO structures they have processed within the experiment prior to a prime trial. Hence, we predict that syntactic priming in our experiment is sensitive to both prior surprisal and adapted surprisal (i.e., surprisal based on recent experience).

##### 4.1. Estimating and manipulating prior prime surprisal

The prior surprisal of the prime stimuli employed for Study 3 (see below) was estimated in a norming experiment using magnitude estimation (Bard, Robertson, & Sorace, 1996). Forty-one Stanford University undergraduate students participated for course credit. Participants were shown one prime sentence at a time in both of the possible word orders (PO or DO). The order of presentation of the two structures was counterbalanced within items and participants. The first sentence was assigned a score of 100 and participants were asked to rate on a multiplicative scale how much better or worse the second word order variant was compared to the first.

The rating task was preceded by instructions that illustrated the task, using examples not used during the norming experiment. This included examples of how fractions may be necessary if the second was more than two orders of magnitude less acceptable than the first.

Following Bard et al. (1996), ratings were log-transformed and then standardized by participant, providing z-scores. We then calculated the average PO z-score for each sentence. These z-scores were normalized into probabilities, based on which prime surprisal was calculated. The distributions of PO bias and prior prime surprisal are summarized in Fig. 6a and b, respectively.

##### 4.2. Estimating and manipulating adapted prime surprisal

Adapted prime surprisal was manipulated within the priming experiment by varying how many prime structures of the same type (PO or DO) were presented to participants in a row, while keeping the overall count of both structures constant (50% PO and 50% DO). Specifically, half of participants saw primes in block order, whereas the other half saw primes in alternating order. In the block order, the first half of the experiment contained only instances of one prime structure (e.g., 12 PO primes, with fillers interspersed) and the second half of the experiment contained only instances of the other prime structure (e.g., 12 DO primes, with fillers interspersed). In the alternating

<sup>5</sup> Unfortunately, the database for Study 1 contained too few primes per conversations to test this prediction. Study 2, although seemingly perfectly suited for investigations of adapted surprisal, almost perfectly confounds adapted prime surprisal with both the proportion of previously experienced PO and DO primes and the structure of the most recent prime (see design details in Kaschak & Borreggine (2008) and Kaschak (2007)).



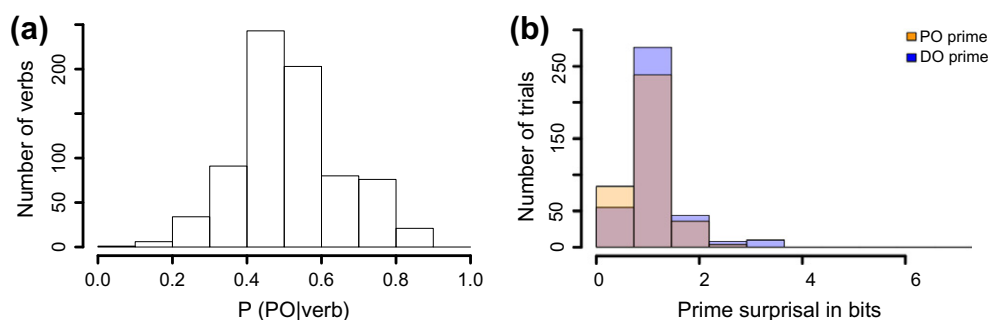


Fig. 6. Histograms of prime verbs' PO biases (a) and prime surprisal (b) in Study 3.

condition, PO and DO prime structures alternated (with fillers interspersed). Both for the block and the alternating orders, the order of PO and DO primes was counterbalanced across participants. Fig. 7 visualizes the block and alternating condition.

Adapted prime surprisal was calculated based on the cumulative proportion of PO and DO primes encountered by the participant.<sup>6</sup> For PO primes, adapted surprisal was based on the proportion of PO primes in previous trials; for DO primes, adapted surprisal was based on the proportion of DO primes in previous trials. To smooth these estimates and to avoid undefined values for the first prime trial (for which  $p(\text{prime structure} \mid \text{previous trials}) = 0$ ), counts were first converted to empirical logits ( $\frac{\text{Count}_{PO} + 0.5}{\text{Count}_{PO} + 0.5}$ ), before being converted into proportions (by taking the inverse logit of the the empirical logit values). Thus the first trial always has a prime surprisal of 1 bit, which corresponds to a uniform prior expectation of  $p(\text{prime structure}) = 0.5$ . In the block condition, the structure presented in the first block should become less surprising given experience during the experiment as the block proceeds. When the first prime of the second half of the experiment is encountered, adapted prime surprisal should rise starkly and then again decrease through the rest of the experiment. In the alternating condition, the surprisal of both prime structures should be distributed more homogenously around 1, reflecting the average proportion of the two prime structures closer 50% compared the block condition.

It should be noted that our estimate of adapted prime surprisal does not combine recent experience with the prior bias due to the prime verb's subcategorization bias and other properties of the prime stimuli. In this sense, what we refer to here as adapted prime surprisal is *not* the same as the posterior surprisal as experienced by participants in our experiment. That posterior surprisal is a combination of the prior surprisal, which we estimated under consideration of the prime verb's subcategorization

bias, and the adapted surprisal, which takes into consideration the distribution of PO and DO structures in recent experience but does not take into account the prime verb's subcategorization bias.<sup>7</sup>

### 4.3. Methods

#### 4.3.1. Participants

The analysis includes data from 25 students at the University of Rochester, who were paid for their participation, and 20 students at Stanford University, who were given course credit, for a total of 45 participants. Sixty participants were originally run in the experiment, but 2 had to be excluded due to recording errors, and another 13 because they did not produce sufficient variation in their target structures (<15% PO targets). Exclusion of these participants does not change the significance of effects reported below.

#### 4.3.2. Materials

Target pictures consisted of the 24 dative-eliciting pictures generously provided by Bock and Griffin (2000). Each picture was grouped with a specific dative prime. The 24 prime items each consisted of one sentence with high PO bias and one sentence with low PO bias (i.e., 48 dative sentences).

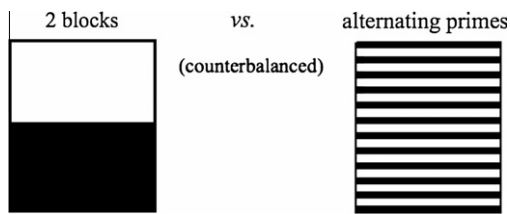
An example item is shown in (4). The complete list of items is given in Appendix A. Prime sentences and target pictures were grouped in such a way that the target picture was unlikely to be described by the prime verb (indeed, the prime and target verbs never overlapped in Study 3).

- (4) a. **PO prime, high surprisal (i.e., low PO bias)**  
The poor painter showed a new work to the art dealer.
- b. **PO prime, low surprisal (i.e., high PO bias)**  
The poor painter sold a new work to the art dealer.
- c. **DO prime, low surprisal (i.e., low PO bias)**  
The poor painter showed the art dealer a new work.
- d. **DO prime, high surprisal (i.e., high PO bias)**  
The poor painter sold the art dealer a new work.

Target trials were mixed with 30 filler sentences and 30 filler pictures and 24 sentence and pictures for an unrelated experiment (all them using transitive structures).

<sup>6</sup> In the general discussion, we also present analyses in which adapted prime surprisal is based on the distribution of preceding PO or DO structures in only the target trials or in both prime and target trials.

<sup>7</sup> To be precise, our estimate of the adapted prime surprisal is identical to the mean of the posterior estimate of  $p(\text{prime structure} \mid \text{previous trials})$  as estimated by a beta-binomial incremental belief update model without consideration of subcategorization and other lexical cues for  $\alpha = 0.5$  and  $\beta = 0.5$  (where  $\alpha$  and  $\beta$  are the 'pseudo count' parameters of the beta-binomial prior at the start of the experiment).



**Fig. 7.** Manipulation of recent experience in Study 3: prime structures were presented either in two blocks, each consisting of 12 primes with the same structure (e.g., 12 PO followed by 12 DO), or in alternating order (e.g., PO, DO, PO, DO, etc.)

Items and fillers were presented in the same order to each participant, regardless of the experimental condition. The two prime structure conditions and the two prior prime surprisal conditions each occurred equally often within the experiment (Latin square design; the order of prime structures used to manipulate adapted prime surprisal was a between-participant factor, although it results in within-participant variability in adapted prime surprisal).

#### 4.3.3. Procedure

Participants were seated in front of a computer and told that they would be participating in an experiment that investigated how language production affected their memory for sentences and pictures. Stimuli were presented using the program Linger (Rohde, 1999). Prime trials were always immediately followed by target trials. The procedure is illustrated in Fig. 8.

On prime trials, participants listened to a prime sentence read by a female voice played over a set of speakers. After the prime sentence had been played, participants pressed the space bar (to begin recording) and repeated the sentence exactly as it was presented. After participants finished speaking, they pressed the space bar again, and then they were asked whether they had heard the sentence before. Participants pressed the 'F' key for 'yes', and the 'J' key for 'no'.

On target trials, participants were presented a picture that they had to describe in one sentence after pressing the space bar (to start recording). Once they were done speaking, they pressed the space bar again and a question appeared asking if they had seen the picture before, which they answered just like in the prime trials.

Items were always separated by at least five fillers (sentence or picture, chosen randomly). On filler trials, participants either repeated a filler sentence or described a filler picture, and also indicated whether they had seen the stimulus before, using the same procedure as for item trials

described above. In order to implement the memory cover task, 18 of the 30 filler sentences and 18 of the 30 filler pictures were repeated a second time in the experiment. To convince participants early on in the experiment that their memory would indeed be tested, the first repeated filler occurred on the third trial, repeating the first trial. Including the 24 experimental trials, the 24 unrelated trials and the 96 filler trials, there were a total of 144 trials in the experiment.

#### 4.3.4. Scoring

Participants responses were scored and annotated by two independent raters. One participant was scored by both raters in order to estimate inter-annotator agreement, which was high (Cohen's  $\kappa = 0.91$ ). For each prime or target response, the annotators transcribed the entire utterance and scored the prime structure, the prime verb, the subject, recipient, and theme NPs.

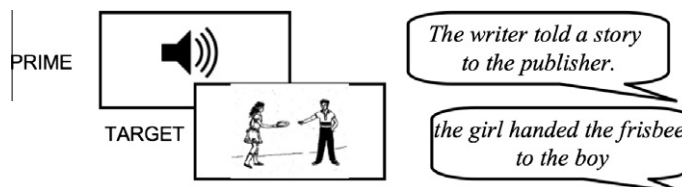
On 4.4% of the prime trials, participants did not repeat the prime correctly. These prime trials were excluded from the analysis and from the count of prime structures based on which adapted surprisal was calculated. Fig. 9 shows the distribution of estimated adapted prime surprisal by item order.

Two of the original items were excluded because they produced low variation in the target productions (<15% PO). This left a total of 755 trials for the analysis. Inclusion of these trials does not change the results reported below.

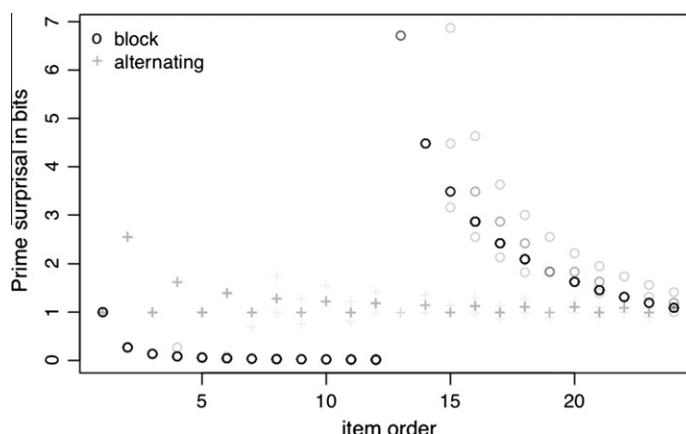
#### 4.3.5. Analysis

As in Studies 1 and 2, we employed a mixed logit regression to predict the occurrence of PO over DO structures. Three random effects were included in the analysis. For by-participant and by-item random effects, the analysis contained the full random effect structure – all intercepts, all slopes, and all covariances. Some of these random effects were not justified by model comparison, but since their removal did not change results we simply report the full model here (but see Appendix B for more details). Since participants varied in what verbs they produced for a picture (there were 24 items in Study 3 but across participants 34 different verbs were produced), we also included a random intercept by target verb, which was justified by model comparison ( $\chi^2(1) = 12.0, p < 0.001$ ).

Finally, we included all priming-related controls from Studies 1 and 2 in the analyses, except for prime-target distance and prime-target verb identity, which were held constant in Study 3. As in Studies 1 and 2, the cumulative count of preceding prime structures included both prime and target trials. All fixed effect correlation  $r_s < 0.15$ .



**Fig. 8.** Procedure for priming trials in Study 3.



**Fig. 9.** Estimates of adapted prime surprisal plotted versus item order. Circles represent the block condition. Crosses represent the alternating condition.

#### 4.4. Predictions

We predicted that the interaction between prime structure and prior prime surprisal observed in Studies 1 and 2 should replicate. If participants indeed rapidly adapt their syntactic expectations based on recent experience including the distribution of prime structures they are exposed to during the experiment (as proposed by, e.g., Chang et al., 2006 and supported by recent work, Farmer et al., 2011; Fine et al., submitted for publication; Kaschak & Glenberg, 2004b), we should also observe the same type of interaction between adapted prime surprisal and prime structure.

#### 4.5. Results

Table 4 summarizes the results. As in Studies 1 and 2, we first summarize the basic priming effects and then turn to the effects of interest.

##### 4.5.1. Basic priming effects

Replicating previous work, we found a significant main effect of prime structure such that PO primes made PO structures more likely to be produced in the target ( $p_z < 0.001$ ). Replicating work by Kaschak and colleagues, we also found a significant effect of cumulative priming ( $p_z < .001$ ): the higher the proportion of PO structures in the trials preceding the target, the more likely speakers were to produce a PO structure.

##### 4.5.2. Prior prime surprisal

We found the predicted interaction of prior prime surprisal and PO prime ( $p_z < 0.05$ ). This effect is plotted in Fig. 10. Replicating Studies 1 and 2, simple effect analyses revealed that the effect of surprisal was significant for PO primes ( $p_z < 0.05$ ) but not DO primes ( $p_z > 0.2$ ), although in the predicted direction. As in Studies 1 and 2, there was no main effect of prior prime surprisal ( $p_z > 0.7$ ).

##### 4.5.3. Adapted prime surprisal

As with prior surprisal, we found the predicted interaction between adapted prime surprisal and the prime structure ( $p_z < .05$ ). This effect is plotted in Fig. 11. Again, the

simple effect of adapted surprisal is significant for PO primes ( $p_z < 0.05$ ) and non-significant for DO primes ( $p_z > 0.3$ , in the expected direction). There was no main effect of adapted prime surprisal ( $p_z > 0.2$ ).

#### 4.6. Discussion

The large effects of prime structure and cumulativeness suggest that Study 3 was successful in eliciting strong syntactic priming effects, which make it a suitable paradigm for the current purpose. More crucially, the experiment confirms and extends the results of Studies 1 and 2. We find that priming is stronger the more surprising the prime structure based both on prior experience and recent experience within the experiment. Before we discuss the consequences of these findings in more detail, we address a potential caveat to our results.

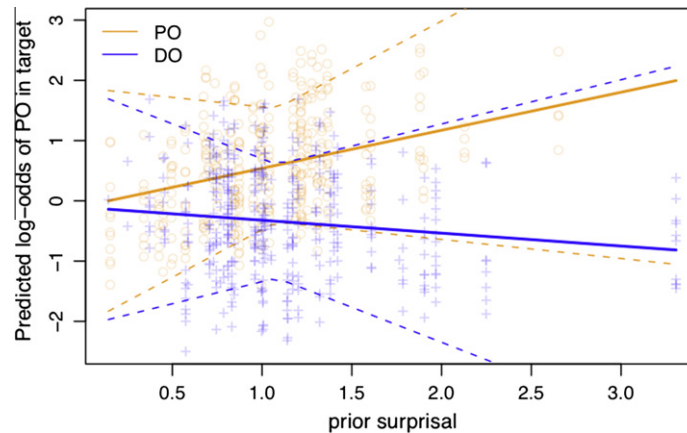
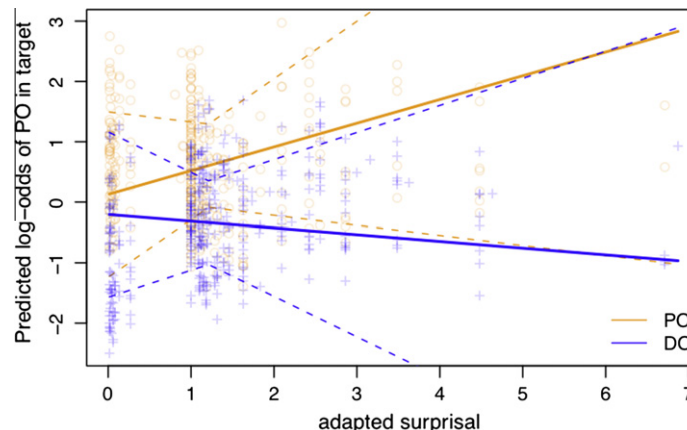
As in Studies 1 and 2, the simple effect of prime surprisal went in the predicted direction for both PO and DO primes. However, in all three studies, the effect only reached significance for PO primes. In Studies 2 and 3, this effect is unlikely to be due to a floor effect since PO and DO structures were about equally likely in the target sentences (cf. mean between orange and blue lines, Figs. 5 and 11). A more likely explanation is that DO primes exhibit only very weak priming effect to begin with, as observed in Study 1 and previous work (Kaschak, 2007; Kaschak et al., 2011). Since this hypothesis cannot be tested here because Study 3 did not contain a baseline condition against which we could compare the effect of PO and DO priming. There is, however, supporting evidence from a number of previous studies that have found weaker or no priming effects for the more frequent structure in an alternation compared to the less frequent structure (e.g., Bock, 1986; Ferreira, 2003; Hartsuiker & Kolk, 1998; Hartsuiker & Westenberg, 2000; Jaeger, 2010; Scheepers, 2003). For example, Kaschak et al. (2011) do not find cumulative priming effects for the more frequent DO structure, but find cumulative priming for the less frequent PO structure.

Most relevant to the current context are the experiments reported in Bernolet and Hartsuiker (2010). Building on earlier presentations of the current work, Bernolet and Hartsuiker investigate the effect of prime surprisal in the

**Table 4**

Summary of Study 3 results (c = centered, r = residualized; for further information see the caption of Table 1).

Predictor (independent variable)	Parameter estimates			Wald's test		$\Delta(-2\mathcal{A})$ -test		Partial pseudo- $R^2$
	Log-odds	S.E.	Odds	Z	$p_z$	$\chi^2$	p	
<i>Basic priming effects</i>								
PO prime (c)	0.94	0.17	2.5	5.3	$\ll 0.001$	21.0	$\ll 0.001^*$	0.054
Cumulative PO primes	1.72	0.46	5.6	3.6	$< 0.001$	14.8	$< 0.001^*$	0.027
<i>Effects of prime's prediction error</i>								
Prior prime surprisal (c,r)	0.11	0.25	1.1	0.3	$> 0.7$	0.2	$> 0.6$	0.001
PO prime $\times$ prior prime surprisal (r)	1.23	0.61	3.4	2.5	$< 0.02$	4.4	$< 0.05^*$	0.007
Adapted prime surprisal (c,r)	0.11	0.09	1.1	1.3	$> 0.18$	2.6	$> 0.11$	0.005
PO prime $\times$ adapted prime surprisal (r)	0.50	0.24	1.6	2.0	$< 0.05$	5.6	$< 0.02^*$	0.011

**Fig. 10.** Interaction between prior prime surprisal (in bits) and priming in Study 3. Orange circles represent PO primes, and blue crosses represent DO primes. Dotted curves represent the 95% confidence intervals of the slopes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)**Fig. 11.** Illustration of the interaction between adapted prime surprisal (in bits) and priming in Study 3. Orange circles represent PO primes, and blue crosses represent DO primes. Dotted curves represent the 95% confidence intervals of the slopes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Dutch dative alternation. They find stronger priming for more surprising DO primes, but no such effect for PO primes. As a matter of fact, Berrolet and Hartsuiker do not observe *any* priming for PO primes. That is, they seem to find the opposite of our result. As Berrolet and Hartsuiker point out though, DOs are actually the *less* frequent structure in Dutch (whereas POs are the less frequent

structure in English). That is, once we take into consideration that English and Dutch differ in terms of the relative frequency of PO and DO structures, the pattern of results found in Studies 1 to 3 is identical to those reported in Berrolet and Hartsuiker. We conclude that the repeated failure to find a significance simple effect of prime surprisal for DO primes is likely due to an overall weak priming



effect for the more frequent prime structure (DOs for English, POs for Dutch).

Still, this is a question that clearly deserves further attention. For example, in an ongoing study, we have found DO priming for English for a different set of verbs (Weath-erholz, Campbell-Kibler, & Jaeger, 2012). One question that needs to be addressed in future work is what determines when priming by the more frequent structure is observed. Another question is whether the strength of DO priming – when it is observed – is dependent on the prediction error. Without further assumptions, this is the prediction made by both the error-based model and the perspective advanced here.

## 5. General discussion

At the outset of this article, we hypothesized that syntactic priming is a consequence of adaptation with the goal to minimize the expected future prediction error. Specifically, we hypothesized that comprehenders continuously adapt their expectations to the statistics of the current environment (see also Botvinick & Plaut, 2004; Chang et al., 2006; Fine et al., submitted for publication; Plaut et al., 1996) and that speakers contribute to the minimization of the mutual prediction error. From this, we derived the prediction that the strength of syntactic priming in production should be correlated with the prediction error speakers' experienced while processing the most recent prime. This is the prediction that we tested and found confirmed in our studies.

In all of our studies, the strength of syntactic priming increased with the surprisal associated with the prime's syntactic structure. Study 1 found this effect in conversational speech, Study 2 found the effect for written production, and Study 3 found the effect for spoken picture descriptions. The three studies also differed in terms of how the prediction error was assessed. For Studies 1 and 2 the prediction error was based on the prime verb's subcategorization bias (see also Bernolet & Hartsuiker, 2010). For Study 3, we estimated the prediction error based on the relative preference for a PO or DO structure given the entire lexical content of the sentence. Finally, we found that syntactic priming during language production is affected by the prime's prediction error given both prior experience (Studies 1–3) and recent experience (Study 3).

The results obtained here have implications for accounts of syntactic priming. We first briefly (re)introduce those accounts. We discuss the extent to which previous evidence distinguishes between these accounts and how that evidence relates to the perspective advanced here. Once this background has been established, we discuss the implications of the current findings for previous accounts and for future work on syntactic priming.

### 5.1. Accounts of syntactic priming

Accounts of syntactic priming can be grouped into two broad classes. Both types of explanations consider syntactic priming to be a mostly automatic process that is not under conscious control (e.g., Bock & Griffin, 2000; Pickering

& Garrod, 2004). They differ, however, in the mechanisms that they assume to underlie syntactic priming. One proposal holds that syntactic priming is due to short-term boosts in the activation of recently processed representations (Pickering & Branigan, 1998; see also Kaschak & Glenberg, 2004a). The competing proposal holds that syntactic priming is a consequence of implicit learning (Bock & Griffin, 2000; Kaschak, 2007). These accounts have been instrumental in advancing our theoretical understanding of syntactic priming and are, in many ways, the inspiration for the current work. In their original specifications they are, however, not sufficiently precise to be distinguishable by empirical means (see also comments in Brown-Schmidt & Tanenhaus (2004), Kaschak & Glenberg (2004a) and Pickering & Ferreira (2008)). For those reasons, we focus on more recent work that has committed to more specific architectural assumptions (Chang et al., 2000; Chang et al., 2006; Dubey et al., 2008; Malhotra, 2009; Reitter et al., 2011). Specifically, we focus on two prominent models and their relation to the approach adopted here (for additional discussion, see also Malhotra, 2009). The first model is the error-based implicit learning model (Chang et al., 2006), which we have already described. The second is the hybrid model by Reitter et al. (2011).

Like the error-based model, the hybrid model derives syntactic priming from rather general assumptions about learning and memory retrieval. Building on work by Dubey et al. (2008), Reitter et al. (2011) present an account of syntactic priming within the ACT-R cognitive architecture (Anderson et al., 2004). In their model, syntactic priming is understood as a consequence of both spreading activation due to recent memory retrieval (cf. activation-boost accounts, Pickering & Branigan, 1998; Kaschak & Glenberg, 2004a) and unsupervised implicit learning.<sup>8</sup> The probability that a certain syntactic structure will be produced is determined by its activation in memory relative to the activation of all other structures that are compatible with the retrieval cues (including the semantic and syntactic constraints). Two mechanisms that are general to ACT-R and have been shown to account for a variety of other behaviors (Anderson et al., 2004) are assumed to contribute to the activation of a structure in memory. First, every time a syntactic structure is retrieved from memory it receives a boost in activation, which is assumed to decay as a power-law over time. Short-term syntactic priming effects arise as a consequence of the activation boost, which is also assumed in other work on sentence processing within the ACT-R cognitive architecture (e.g., Lewis & Vasishth, 2005). Second, activation is assumed to spread between related nodes, including from lexical to syntactic nodes and vice versa. Among other things, this property of the model accounts for the lexical boost effect described in the introduction (Pickering & Branigan, 1998).

The activation a node receives through priming is assumed to decrease as a power-law, thereby predicting that

<sup>8</sup> Chang and colleagues also propose an additional, memory-based, mechanisms in addition to error-based implicit learning (Chang et al., 2006, p. 256). In their model, this mechanism is necessary to explain the lexical boost effect (for discussion, see Chang, Janciuskas, & Fitz, 2012; Rowland, Chang, Ambridge, Pine, & Lieven, 2012; Reitter et al., 2011, p. 593).

the strength of syntactic priming decays over time. Additionally, the assumption of power-law decay also means the activation boost experienced while processing a prime is never *completely* undone. As a consequence, repeated retrieval increases the base activation of a structure, so that more frequent structures have higher base activation. This constitutes a form of unsupervised learning that is commonly referred to as *base-level learning*. Base-level learning explains why structures that are more frequently observed are produced more frequently. As we'll detail below, base-level learning also provides an explanation for the inverse frequency effect (the less frequent structures prime more strongly).

Our own approach is closely related to the error-based model, although the two models are derived from different assumptions and with different goals in mind. For example, in the error-based model life-long implicit learning is an assumption, whereas for us rationality or near-rationality is the assumption from which life-long learning follows given the subjective non-stationarity of the linguistic signal, as perceived by the comprehender. Another difference is that, unlike Chang and colleagues, we do not assume error-based learning. In the error-based model, the prediction error is the quantity that is used to adjust the weights of the network. This is, however, not a necessary assumption to derive the prediction that syntactic priming is a function of the prediction error associated with the prime. Any mechanisms that leads to rational or near-rational learning should be sensitive to the prediction error, without necessarily directly referring to it (see also Courville et al. (2006) and Qian et al. (2012) for recent discussions).<sup>9</sup>

Next, we discuss four arguments that have been brought forward in order to distinguish between implicit learning and activation boosts accounts. In doing so, we also point to further differences between the error-based model and the perspective adopted here.

## 5.2. Previous evidence

The first argument is based on the firmly established observation that syntactic priming is generally stronger when the prime and target overlap lexically (the 'lexical boost' effect, e.g., Gries, 2005; Pickering & Branigan, 1998; Snider, 2008). Reitter et al. (2011, p. 601) argue that implicit learning alone is insufficient to account for this property. However, while this argument might hold for both the base-level learning component in Reitter and colleagues' hybrid model and the error-based model (as also acknowledged by Chang et al. (2006, p. 256)), it is not apparent why this argument holds for all types of implicit learning. Specifically, the lexical boost effect is perfectly

compatible with the hypothesis that syntactic priming is a consequence of expectation adaptation with the goal to minimize the expected future prediction error: a prime with the same verb as the target sentence arguably contains more information about what structure is to be expected in the target than a prime with a different verb. The same holds for other types of lexical overlap, which have also been observed to result in stronger priming (Snider, 2008). That is, stronger priming in the presence of lexical overlap might be a rational property of expectation adaptation (for further discussion, see Fine et al., *submitted for publication*). This would predict that the magnitude of the lexical boost effect depends on the informativity of the lexical cue with regard to the target structure.

The second argument holds that only implicit learning accounts predict that syntactic priming effects should be long-lived (Bock & Griffin, 2000). Some studies have found syntactic priming to persist over many sentences or even throughout conversations (e.g., Bock & Griffin, 2000; Bock et al., 2007; Branigan, Pickering, Stewart, & McLean, 2000; Ferreira, Bock, Wilson, & Cohen, 2008; Snider, 2008). This is in line with the results of Study 1 (Studies 2 and 3 did not speak to this issue). Some recent studies even report robust syntactic priming effects as long as a whole week after exposure (Kaschak, Kutta, & Schatschneider, 2011).

However, others have found that the effects of syntactic priming decay (e.g., Branigan et al., 1999; Gries, 2005; Reitter et al., 2011; Szmrecsányi, 2005). A potential resolution of the apparent conflict between these findings is offered by Hartsuiker et al. (2008). Hartsuiker and colleagues note that previous studies on the decay of syntactic priming differed in whether prime and target overlapped lexically, which is known to affect syntactic priming effect (Pickering & Branigan, 1998). In a series of experiments, Hartsuiker and colleagues find that the lexical boost effect decays rapidly, while syntactic priming effects that are independent of lexical overlap persist over a longer time. Some have concluded from this that both short-term activation boosts and longer-lived implicit learning are involved in syntactic priming (e.g., Ferreira & Bock, 2006; Hartsuiker et al., 2008).

Under the assumption of rational or near-rational expectation adaptation, however, these differences in decay rates might reduce to properties of the relevant linguistic distributions. Consider that repetitions of the same lexical material tend to cluster (i.e., they are overdispersed, Altmann, Pierrehumbert, & Motter, 2009): in natural discourses, speakers do not switch topics every sentence – instead, we tend to talk about one thing for some time until we switch topics. If comprehenders are sensitive to this property of lexical recurrence distributions (for preliminary evidence, see Heller & Pierrehumbert, 2011), faster decay rates for the lexical boost effect would be a rational response to the statistics of the environment (akin to the effects reported in Anderson & Schooler, 1991). More generally, if comprehenders weigh cues relative to their informativity about the statistics of the current environment, decay of syntactic priming might well be expected (cf. Anderson & Milson, 1989). In the perspective adopted here, the question for future work will be to what extent hearing a speaker produce a certain syntactic struc-

<sup>9</sup> For example, Bayesian belief updating models do not explicitly refer to the prediction error in their learning rule. Still, even a simple Beta-binomial belief updating model predicts larger changes in beliefs after more surprising events. Thus, the position taken here is a weaker hypothesis with regard to the type of implicit learning that is assumed to underlie syntactic priming. We prefer this position because, to the best of our knowledge, there are currently no data – including those presented here – that clearly show that error-based, rather than other types of supervised or unsupervised, learning is required to explain syntactic priming.

ture provides information about subsequent productions of that speaker and how fast this information decays (for example, for evidence that the informativity of lexical cues about upcoming lexical content decays relatively rapidly, see Clarkson & Robinson, 1997; Qian & Jaeger, 2012). In short, evidence from the differential decay of syntactic priming and the lexical boost effect does not necessarily rule out implicit learning accounts (contrary to Reitter et al., 2011, pp. 601, 621–623).

The third argument that has been put forward to distinguish between implicit learning and activation boost accounts derives from the second: if syntactic priming is due to implicit learning, it should be affected by the cumulative recent experience, rather than just the most recent prime (Kaschak, 2007). Indeed, work by Kaschak and colleagues provides evidence that syntactic priming is cumulative: the more often a structure has been mentioned in recent history, the more likely speakers are to repeat it (e.g., Kaschak, 2007; Kaschak & Borreggine, 2008). Such cumulativeness follows naturally from the prediction that syntactic priming is long-lived. However, cumulativeness can also be accommodated within activation-boost accounts under certain assumptions about the rate with which activation decays (e.g., Reitter et al., 2011; for discussion, see also Kaschak & Glenberg, 2004b; Malhotra, 2009). Both under the perspective advanced here and under the error-based model, cumulativeness follows without further assumptions: as the proportion of a structure in recent experience increases, speakers should be more likely to produce it.

Finally, the fourth argument refers to 'inverse frequency' effect (e.g., Bock, 1986; Hartsuiker & Kolk, 1998; Kaschak et al., 2011; Scheepers, 2003). There is as of yet no activation-boost account that explains the inverse frequency effect. In contrast, both unsupervised (Reitter et al., 2011) and supervised implicit learning accounts (Chang et al., 2006) can account for the inverse frequency effect. In the error-based model, the inverse frequency effect follows from the assumption of error-based learning, which leads to larger changes in the connection weights after less expected primes. In the hybrid model, base-level learning leads to increased resting activation for more frequent structures. As a consequence, the activation boost associated with processing a prime leads to a smaller *relative* increase in the activation of more frequent structures. Since the probability of producing a structure depends on its activation compared to the activation of other structures, this correctly predicts that more frequent structures should lead to weaker priming.

In sum, the picture that is emerging from previous work is that implicit learning is at least one of several mechanisms that together bring about syntactic priming (Ferreira & Bock, 2006; Chang et al., 2012; Hartsuiker et al., 2008; Reitter et al., 2011). What is less clear is what *type* of implicit learning underlies syntactic priming. Both unsupervised (Reitter et al., 2011) and supervised implicit learning mechanisms (Chang et al., 2006) have been proposed.

### 5.3. Implications of the current results

Our finding that syntactic priming in production is affected by the prediction error provides further support

for implicit learning accounts. In principle, both supervised and unsupervised implicit learning accounts are compatible with this finding, as long as they employed learning mechanism predicts that on average more expectation adaptation takes place after primes with larger prediction errors. The results obtained here do, however, distinguish between existing *models* of implicit learning and syntactic priming. They also constrain the future development of these models. We first discuss the effect of prior prime surprisal and then turn to the effect of adapted prime surprisal. Some implicit learning accounts of syntactic priming do not further specify the type of learning they assume (Bock & Griffin, 2000; Bock et al., 2007). These accounts are broadly compatible with the results observed here, although they do not predict them. We therefore do not further discuss them.

#### 5.3.1. Prior prime surprisal

Studies 1 to 3 provide evidence that the strength of syntactic priming increases with the prediction error based on the prime verb's subcategorization bias (see also Bernolet & Hartsuiker, 2010). Sensitivity to the prediction error is predicted by our guiding hypothesis since rational and near-rational learning under uncertainty implies sensitivity to the prediction error (cf. belief-updating accounts of sequential learning under uncertainty, Dayan & Yu, 2001; Yu & Cohen, 2008). Simply put, showing stronger adaptive responses following surprising events is a rational response to learning in non-stationary environments in which the underlying distributions are expected to undergo rapid changes. Surprising events can be evidence for rapid changes of the distributions that underlie the observable events and hence are evidence that beliefs about these distributions should be adapted (Courville et al., 2006; Qian et al., 2012). These Bayesian accounts make an interesting prediction that distinguishes them from error-based accounts (Elman, 1990; Chang et al., 2006): the extent to which producers and comprehenders should adjust their expectations based on recent surprising events should depend on their prior estimate of how likely the distribution of relevant syntactic structures is to change rapidly (see Qian et al., 2012).

Qualitatively, our results are also predicted by the error-based model. There are, however, several open questions that need to be addressed in future work. First, our analyses tested whether there was a linear relation between surprisal and the log-odds of producing a PO (rather than DO) structure. The error-based model uses the prediction error to adjust the weights of a simple recurrent network. This results in a highly non-linear relation between the prediction error and the strength of syntactic priming.<sup>10</sup> While it seems safe to assume that the relation between the prediction error and the strength of syntactic priming predicted by the error-based model is

<sup>10</sup> The error-based model employs backpropagation to adjust the weights, likely leading to changes in the weights that are non-linear in the prediction error. Additionally, the probability of producing a PO structure is a non-linear function of these weight changes. The transformation of these probabilities into log-odds, which we used here to assess the strength of syntactic priming, also is non-linear.

positive and monotonic (hence correctly predicting the qualitative effect), it is an open question to what extent the error-based model provides a good quantitative fit for the type of data observed here. Here, we tested (and found) only a linear relation between the prediction error and the increase in the log-odds of producing a certain structure.

Second, our estimate of the prediction error was based on the probability of the *syntactic structure*. The sequencing system in the error-based model predicts words, not structures, and the error signal used to change weights in the model is a word-based prediction error. Syntactic expectations are only implicitly encoded in the hidden nodes of the model in a distributed fashion. Hence, while our estimate of the prediction error will tend to be correlated with the word-by-word prediction errors employed in the error-based model, the two measures are not the same. In this context, it is interesting to consider recent work that has compared simple recurrent models (SRNs) against models with explicit hierarchical syntactic structures (probabilistic phrase structure grammars, PCFGs) as predictors of processing difficulty in reading. This work has found that surprisal estimates based on recurrent networks predict at least some measures of word-by-word processing difficulty better than probabilistic phrase structure grammars do (Alexandre, 2010; Frank & Bod, 2011). The results obtained here suggest that evidence from syntactic priming can be used to similarly test what types of representations are accessed during sentence production. For example, future work could compare the extent to which SRN- vs. PCFG-based prediction errors predict the strength of syntactic priming.

Next, we turn to the hybrid model (Reitter et al., 2011). Reitter and colleagues found that unsupervised base-level learning can account for the inverse frequency effect, which – we argued above – is potentially related to the effect of prime surprisal observed here. It is theoretically possible that base-level learning operates not only over syntactic structures, but also over associations between syntactic structures and their lexical heads (or even their dependents). Reitter and colleagues describe a second learning mechanism, *associative learning* (Anderson, 1993), which they assume to operate in addition to the standard ACT-R assumptions. In the hybrid model, this learning mechanisms seems to be used only to acquire the associative links between lexical material and syntactic structures that account for the lexical boost effect (Reitter et al., 2011, p. 635). If such a mechanism was, however, to be combined with power-law activation decay, as assumed for base-level learning, it might capture base-level learning of verb-structure associations. Such a modified hybrid model would predict inverse effect of the *joint* frequency of, for instance, prime verbs and prime structures on the strength of syntactic priming (paralleling the inverse frequency effect observed for standard base-level learning over prime structures).

While a implementation of this hypothesis into the model of Reitter and colleagues is beyond the scope of this paper, we can utilize the data from Studies 1 to 3 to test whether the joint frequency of prime verb and structure exhibits a significant interaction with the prime structure. We calculated the joint frequency of the prime verb and

structure for all primes in our studies. For the re-analysis of Study 1, we estimated joint frequencies based on the full Switchboard corpus (Godfrey et al., 1992). For the re-analysis of Studies 2 and 3, we estimated joint frequencies based on the Roland et al. (2007) database. That is, for re-analysis of Studies 1 and 2, we used the same data sources to estimate the joint frequencies that we had used to estimate prior prime surprisal (the norming experiment employed for Study 3 above cannot be used to estimate joint frequencies). We then entered the joint frequency estimate and its interaction with the prime structure into the analyses from Studies 1 to 3, replacing the predictors for prime surprisal and its interaction with prime structure.

For all three data sets, we found that the joint probability of the prime structure and the prime verb had no effect on participants' likelihood of repeating the prime (Study 1:  $p > .1$ ; Study 2:  $p > .1$ ; Study 3:  $p > .5$ ; all  $p$ s based on model comparison). More crucially, neither did the joint probability interact with the prime structure (Study 1:  $p > .1$ ; Study 2:  $p > .9$ ; Study 3:  $p > .5$ ). The same was observed if the joint probability was log-transformed, akin to prime surprisal (main effect, Study 1:  $p > .1$ ; Study 2:  $p > .4$ ; Study 3:  $p > .5$ ; interaction, Study 1:  $p > .3$ ; Study 2:  $p > .5$ ; Study 3:  $p > .1$ ).<sup>11</sup>

In recent work on syntactic priming in *comprehension*, we have found the same pattern as observed here: a significant interaction between prime surprisal (given the verb) and prime structure, but no significant interaction between the joint frequency of the prime verb and prime structure (Fine & Jaeger, *in press*). Together with the current results, we take this as tentative evidence against an account of our results in terms of base-level learning.

### 5.3.2. Adapted prime surprisal

If language users indeed continuously adapt their expectations based on recent input, this would mean that the prediction error itself should change as a function of recent experience. Study 3 lends support for this hypothesis.

The effect of adapted surprisal is generally compatible with the error-based model, which ties language processing to learning: every time a word is processed, the prediction error experienced while processing it is used to adjust the weights of the simple recurrent network(s). With changes to the weights, the predictions of the model change after every word and hence also after every prime. This means that the model should in principle be able to account for the type of effect observed in Study 3.

A question to be addressed in future work is *how well* the error-based model captures the cumulative effect of recent exposure. In the error-based model, the speed of adaptation is determined by the learning rate, which governs how much the connection weights change as a

<sup>11</sup> For the re-analysis of the data from Study 3, the failure to find an effect might be due to the fact that the 48 verbs employed in Study 3 included many infrequent verbs, for which joint frequency estimates are unreliable (recall that Study 3 – for this reason – employed a norming experiment to estimate prior prime surprisal). However, for Studies 1 and 2, we found a significant interaction between prime structure and prior prime surprisal, whereas the same analysis conducted with joint frequencies rather than surprisal does not return significant interactions.



function of the prediction error. This learning rate is not itself subject to learning or other constraints. It is set by the researcher. Chang et al. (2006, p. 270) assume that the learning rate decreases throughout life, an assumption they seem to make in order to successfully model both language acquisition and syntactic priming.

This would, however, suggest relatively *slow* expectation adaptation in adults, which seems to be in conflict with recent work on comprehension that has found rapid expectation adaptation within a few tens of sentences (Farmer et al., 2011; Fine et al., submitted for publication; Kamide, 2012; Kaschak & Glenberg, 2004b). A similar question applies to the hybrid model. It is unclear whether the model can be parameterized in such a way that it predicts the inverse frequency effect, while at the same time allowing for rapid expectation adaptation. It is possible that the assumption of power-law decay yields the right balance between rapid activation changes and long-lasting changes in the base-line activation. Ultimately, these questions will only be answered by applying the error-based and hybrid model to the type of data obtained here.

In this context, it is interesting to note that there is also evidence that syntactic adaptation effects can persist over at least several days (Fine & Jaeger, 2011; Wells et al., 2009; see also Kaschak et al., 2011). In these experiments, participants were exposed to unusual syntactic distributions of syntactically well-formed sentences (e.g., an unusual number of garden path sentences, Fine & Jaeger, 2011). Such exposure led to rapid expectation adaptation, as reflected in reduced processing difficulty for syntactic structures that were frequently encountered in recent experience (cf. Fine et al., submitted for publication). Crucially, the effects persisted when participants were tested again a few days later in the same environment. Since participants presumably experienced a substantial number of instances of the relevant syntactic structures in the days that passed between the exposure and test phase, this raises the question as to why the adaptation effects persisted. If confirmed, these findings challenge models in which learning always proceeds at the same rate and in which knowledge about previously encountered linguistic environments (e.g., speaker-specific properties) is not maintained (e.g., Chang et al., 2006; Reitter et al., 2011).

Of course, it is possible that entirely different processes underlie rapid expectation adaptation and the longer-lasting effects. Here, we are less interested in what additional mechanisms need to be assumed to reconcile these models with phenomena they were never intended to account for. We note, however, that the perspective we are exploring here offers a potential unifying explanation for both rapid and long-lasting effects: if the goal of implicit learning during language processing is to minimize the expected future prediction error, then it is expected that comprehenders do not just adapt to the statistics of novel environments (such as a novel speaker or being in an experiment), but that they also *remember* these statistics (in particular, when there are reasons to believe that the environment will be encountered again, Anderson & Schooler, 1991; see also Qian et al., 2012).

If speakers indeed adapt their syntactic expectations to novel environments and remember those these statistics

later, our hypothesis that syntactic priming is a consequence of these adaptation processes would predict that the relevant prediction error is based on these environment-specific statistics – even days later (or, however long the adaptation effects last). More generally, we would expect that the prediction error that affects syntactic priming reflects whatever expectations comprehenders hold for the current environment. For example, if exposed to a novel environment, comprehenders might generalize prior expectations to this novel environment based on the perceived similarity of the current environment to previous environments. If comprehenders have reasons to believe that a novel speaker has similar production preferences as a previously experienced speaker, this should affect the prediction error and thus the strength of syntactic priming experienced while listening to the novel speaker.

We close by briefly raising two further questions we consider particularly important for future work on syntactic priming, for one of which we present preliminary data from Study 3.

#### 5.4. Are prediction errors only experienced in comprehension?

So far, we have assumed that prediction errors that cause syntactic priming are experienced during language comprehension (see also Chang et al., 2006). However, it is possible that speakers also monitor their *own* speech for prediction errors. While most existing models of sentence production do not characterize production as a predictive process, this perspective has been productively employed in work on articulation (e.g., Guenther, 1995) and, more generally, motor planning (e.g., Wei & Körding, 2009; Wolpert, 1997; Wolpert & Kawato, 1998; see also Clark, *in press* and references therein).

Indeed, there is evidence that language production involves prediction in ways similar to comprehension (e.g. Aylett & Turk, 2004; Frank & Jaeger, 2008; Jaeger, 2010; Resnik, 1996; Wasow et al., 2011; van Son & van Santen, 2005; but see Ferreira, 2008). Additionally, there is evidence that speakers can learn based on their perception of own productions (e.g., Dell et al., 2008; Houde & Jordan, 1998; Jones & Munhall, 2005; Warker & Dell, 2006). Particularly powerful evidence for this type of adaptation comes from perturbation studies: when speakers' perception of their own articulations are perturbed so that they seem to deviate from the intended target pronunciation (e.g., by real-time manipulation of the first formant during the articulation of a vowel), speakers adjust subsequent articulations to compensate for the perceived deviation (e.g., Houde & Jordan, 1998; Jones & Munhall, 2005).<sup>12</sup>

Relatively little is known about comparable processes during syntactic production (but see Butler, Jaeger, Bohnemeyer, & Furth, 2011; Roche, Dale, & Kreuz, *in preparation* for preliminary evidence). It is possible that prediction errors experienced during both production and comprehension affect syntactic priming.

<sup>12</sup> Interestingly, these adjustments seem to be relative to a target in acoustic, rather than articulatory, space (Perkell et al., 2004; see also Frank, 2010).

This brings us to the final question we hope to address in more detail in future work. If syntactic priming is indeed a consequence of expectation adaptation with the goal to facilitate efficient information transfer, we would expect speakers to be primarily sensitive to recent *comprehension* experience. That is, regardless of whether prediction errors are experienced during production or comprehension or both, the prediction error should primarily be based on expectations that stem from recent comprehension experience with the interlocutors' productions.

Although Study 3 was not designed to address this question, it is suited for a preliminary test. In Study 3, primes were presented auditorily and then repeated by participants. Target sentences were elicited through pictures. That is, while both primes and targets were produced, only primes constitute comprehension experience in the sense relevant here (i.e., only the primes convey information about the statistics of the current linguistic environment). For the original analysis of Study 3, we estimated adapted prime surprisal based the distribution of only primes. As was shown in Table 4, we found the predicted interaction between adapted prime surprisal and prime structure ( $\chi^2(1) = 5.6, p < .05$ ). We repeated this analysis, using adapted surprisal estimates that were based on only the previous targets or both on previous primes and targets. Adapted surprisal given only previous targets does not result in a significant interaction with prime structure ( $\chi^2(1) = 0.6, p > .4$ ) and neither does adapted surprisal given both previous prime and target structures ( $\chi^2(1) = 2.0, p > .15$ ).

These findings suggest that the recent experience that determines the prediction error is primarily due to recently comprehended, rather than produced, materials – in line with our hypothesis. If confirmed by future studies, the current findings thus offer a rather concrete explanation as to why syntactic priming in production facilitates efficient communication (as evidenced in Reitter & Moore, 2007): syntactic priming in production reduces the average prediction error experienced during comprehension, which – as we discussed in the introduction – is well-known to be positively correlated with comprehension difficulty (DeLong et al., 2005; Demberg & Keller, 2008; Frank & Bod, 2011; Levy, 2008; McDonald & Shillcock, 2003). This way, the perspective advanced here holds the potential to bring us one step closer towards understanding the alignment mechanisms that underlie successful communication (Brennan & Clark, 1996; Clark, 1996; Pickering & Garrod, 2004). If confirmed, the current results also offer further support for the hypothesis that the cognitive systems underlying language production are organized so as to facilitate efficient information transfer (Jaeger, 2006; Jaeger, 2010).

There are, however, alternative interpretations of the result obtained here. Perhaps it is the fact that primes in Study 3 were both comprehended *and* produced that resulted in a stronger effect of the prediction error. It is also possible that prediction errors in production and comprehension are affecting speakers equally strongly, but that prediction errors in production tend to be smaller. We leave these questions to future work.

## 6. Conclusion

We find that syntactic priming is sensitive to the prediction error experienced while processing the prime structure (see also Bernolet & Hartsuiker, 2010). As would be expected based on research on syntactic comprehension (e.g., Garnsey et al., 1997; MacDonald et al., 1994; Kamide et al., 2003; Trueswell et al., 1993), this prediction error is context-dependent, including expectations based both on prior and recent experience within the current linguistic environment. As far as previous accounts of syntactic priming are concerned, this result provides further support for error-based implicit learning accounts (Chang et al., 2006). More broadly, our results are compatible with, though not predicted by, accounts that attribute syntactic priming at least partially to implicit learning (Bock & Griffin, 2000). Existing activation-boost accounts that aim to explain syntactic priming without reference to implicit learning are incompatible with our results (Dubey et al., 2008; Pickering & Branigan, 1998).

Beyond syntactic priming, the results obtained here adds to previous findings that the systems underlying sentence processing are remarkably sensitive to recent experience (e.g., Fine et al., submitted for publication; Kamide, 2012; Kaschak & Glenberg, 2004b; see also work on adaptation in speech processing, Bradlow & Bent, 2008; Eisner & McQueen, 2005; Kraljic & Samuel, 2007; Kraljic et al., 2008; Vroomen et al., 2007). Evidence from speech perception, syntactic, prosodic, and pragmatic processing suggests that comprehenders continuously adapt their expectations so as to match or approximate the statistics of the current linguistic environment. Our results suggest that language production is either subject to similar adaptation processes or that adaptation processes operating during comprehension are reflected during production (e.g., due to sequencing systems that are shared between comprehension and production).

## Acknowledgments

We would like to thank Joan Bresnan for generously providing the data for Study 1, Mike Kaschak for the data for Study 2, and Kathryn Bock for the picture stimuli for Study 3. We thank Katrina Furth, Sarah Brown, and Amanda Perlman for help with the stimuli and annotation for Study 3, and Dana Subik for help in running Study 3. We are also grateful to Thomas Farmer, Alex Fine, Mike Kaschak, Franklin Chang, Maureen Gillespie, and Nicole Craycraft for comments on earlier versions of this paper. Parts of the results presented here were first presented at the 2007 CUNY Sentence Processing Conference. This work was partially funded by an Alfred P. Sloan Research Fellowship, NSF award BCS-0845059, and NSF CAREER Award (NSF IIS-1150028) to TFJ and a post-doctoral fellowship at the Center of Language Sciences at the University of Rochester to NES (NIH training Grant #T32 DC000035).

## Appendix A

1. LO PO: The new employee lent a CD to her co-worker.  
LO DO: The new employee lent her co-worker a CD.  
HI PO: The new employee swapped a CD to her co-worker.  
HI DO: The new employee swapped her co-worker a CD.
2. LO PO: The scheming gambler mailed an ace to the casino owner.  
LO DO: The scheming gambler mailed the casino owner an ace.  
HI PO: The scheming gambler flipped an ace to the casino owner.  
HI DO: The scheming gambler flipped the casino owner an ace.
3. LO PO: The clumsy student owed a new racket to his roommate.  
LO DO: The clumsy student owed his roommate a new racket.  
HI PO: The clumsy student tossed a new racket to his roommate.  
HI DO: The clumsy student tossed his roommate a new racket.
4. LO PO: The wealthy socialite awarded a scholarship to the needy child.  
LO DO: The wealthy socialite awarded the needy child a scholarship.  
HI PO: The wealthy socialite presented a scholarship to the needy child.  
HI DO: The wealthy socialite presented the needy child a scholarship.
5. LO PO: The writer told a story to the publisher.  
LO DO: The writer told the publisher a story.  
HI PO: The writer passed a story to the publisher.  
HI DO: The writer passed the publisher a story.
6. LO PO: The pastry chef loaned a pie recipe to the contest judge.  
LO DO: The pastry chef loaned the contest judge a pie recipe.  
HI PO: The pastry chef submitted a pie recipe to the contest judge.  
HI DO: The pastry chef submitted the contest judge a pie recipe.
7. LO PO: The master painter handed a portrait to the museum curator.  
LO DO: The master painter handed the museum curator a portrait.  
HI PO: The master painter left a portrait to the museum curator.  
HI DO: The master painter left the museum curator a portrait.
8. LO PO: The English teacher gave a take-home test to

- the class.  
LO DO: The English teacher gave the class a take-home test.  
HI PO: The English teacher assigned a take-home test to the class.  
HI DO: The English teacher assigned the class a take-home test.
9. LO PO: The oafish teenager cost fifty dollars to the store owner.  
LO DO: The oafish teenager cost the store owner fifty dollars.  
HI PO: The oafish teenager took fifty dollars to the store owner.  
HI DO: The oafish teenager took the store owner fifty dollars.
  10. LO PO: The governor allotted a thousand dollars to the town mayor.  
LO DO: The governor allotted the town mayor a thousand dollars.  
HI PO: The governor charged a thousand dollars to the town mayor.  
HI DO: The governor charged the town mayor a thousand dollars.
  11. LO PO: The poor painter showed a new work to the art dealer.  
LO DO: The poor painter showed the art dealer a new work.  
HI PO: The poor painter sold a new work to the art dealer.  
HI DO: The poor painter sold the art dealer a new work.
  12. LO PO: The losing gambler bet a poker chip to the cocktail waitress.  
LO DO: The losing gambler bet the cocktail waitress a poker chip.  
HI PO: The losing gambler bought a poker chip for the cocktail waitress.  
HI DO: The losing gambler bought the cocktail waitress a poker chip.
  13. LO PO: The proud principal sent a certificate to the winning athlete.  
LO DO: The proud principal sent the winning athlete a certificate.  
HI PO: The proud principal issued a certificate to the winning athlete.  
HI DO: The proud principal issued the winning athlete a certificate.
  14. LO PO: The restaurateur baked an apple turnover for the newspaper critic.  
LO DO: The restaurateur baked the newspaper critic an apple turnover.  
HI PO: The restaurateur packed an apple turnover for the newspaper critic.  
HI DO: The restaurateur packed the newspaper critic an apple turnover.
  15. LO PO: The bank officer paid a large sum to the

(continued on next page)

young couple.  
 LO DO: The bank officer paid the young couple a large sum.  
 HI PO: The bank officer quoted a large sum to the young couple.  
 HI DO: The bank officer quoted the young couple a large sum.

16. LO PO: The Sunday school teacher saved a cupcake for the hungry child.  
 LO DO: The Sunday school teacher saved the hungry child a cupcake.  
 HI PO: The Sunday school teacher made a cupcake for the hungry child.  
 HI DO: The Sunday school teacher made the hungry child a cupcake.

17. LO PO: The comic collector offered a first edition to the begging kid.  
 LO DO: The comic collector offered the begging kid a first edition.  
 HI PO: The comic collector allowed a first edition to the begging kid.  
 HI DO: The comic collector allowed the begging kid a first edition.

18. LO PO: The teenage vandal promised a new lawnmower to the nextdoor neighbor.  
 LO DO: The teenage vandal promised the nextdoor neighbor a new lawnmower.  
 HI PO: The teenage vandal brought a new lawnmower to the nextdoor neighbor.  
 HI DO: The teenage vandal brought the nextdoor neighbor a new lawnmower.

19. LO PO: The underdog wished bad luck to the champion.  
 LO DO: The underdog wished the champion bad luck.  
 HI PO: The underdog caused bad luck to the champion.  
 HI DO: The underdog caused the champion bad luck.

20. LO PO: The famous journalist did an article for the magazine editor.  
 LO DO: The famous journalist did the magazine editor an article.  
 HI PO: The famous journalist wrote an article for the magazine editor.  
 HI DO: The famous journalist wrote the magazine editor an article.

21. LO PO: The babysitter fed a chocolate to the screaming toddler.  
 LO DO: The babysitter fed the screaming toddler a chocolate.  
 HI PO: The babysitter threw a chocolate to the screaming toddler.  
 HI DO: The babysitter threw the screaming toddler a chocolate.

22. LO PO: The stingy sculptor afforded a statue to the town museum.  
 LO DO: The stingy sculptor afforded the town museum a statue.  
 HI PO: The stingy sculptor rented a statue to the town

museum.  
 HI DO: The stingy sculptor rented the town museum a statue.

23. LO PO: The prison guard denied a hot meal to the rowdy prisoners.  
 LO DO: The prison guard denied the rowdy prisoners a hot meal.  
 HI PO: The prison guard served a hot meal to the rowdy prisoners.  
 HI DO: The prison guard served the rowdy prisoners a hot meal.

24. LO PO: The consultant taught a lesson to the new trainee.  
 LO DO: The consultant taught the new trainee a lesson.  
 HI PO: The consultant read a lesson to the new trainee.  
 HI DO: The consultant read the new trainee a lesson.

## Appendix B

### Study 2 random effects:

Predictor	p-Value
Participant	$\ll 0.001$
Item	$\ll 0.001$
Item $\times$ prime	$> 0.5$
Item $\times$ cumulativity condition	$< 0.005$
Item $\times$ prime $\times$ cumulativity condition	$> 0.05$

### Study 3 random effects:

Predictor	p-Value
Participant	$< 0.001$
Item	$> 0.1$
Item $\times$ prime	$> 0.2$
Item $\times$ bias	$< 0.05$
Item $\times$ cumulativity	$> 0.3$
Item $\times$ prime $\times$ bias	$> 0.3$
Item $\times$ prime $\times$ cumulativity	$> 0.2$
Item $\times$ bias $\times$ cumulativity	$> 0.4$
Item $\times$ prime $\times$ bias $\times$ cumulativity	$> 0.5$

## References

- Alexandre, J. D. (2010). Modeling implicit and explicit processes in recursive sequence structure learning. In S. Ohlsson & R. Catrambone (Eds.), *The 32nd annual meeting of the Cognitive Science Society (CogSci10)*. Austin, TX: Cognitive Science Society.
- Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264.
- Altmann, E., Pierrehumbert, J., & Motter, A. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLoS One*, 4(11), e7678.



- Anderson, J. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Anderson, J., Bothell, D., Byrne, M., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 10–36.
- Anderson, J., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96(4), 703–719.
- Anderson, J., & Schooler, L. (1991). Reflections of the environment in memory. *Psychological Science*, 2(6), 396.
- Arai, M., van Gompel, R., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, 54(3), 218–250.
- Arnold, J. E., Wasow, T., Losongco, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55.
- Aylett, M. P., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31–56.
- Bard, E., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72, 32–68.
- Bernolet, S., & Hartsuiker, R. (2010). Does verb bias modulate syntactic priming? *Cognition*, 114(3), 455–461.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Bock, J. K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, 104(3), 437–458.
- Bock, J. K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology*, 129(2), 177–192.
- Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21(1), 47–67.
- Boston, M., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *Journal of Eye Movement Research*, 2(1), 1–12.
- Botvinick, M., & Plaut, D. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, 111(2), 395–429.
- Bradlow, A., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729.
- Branigan, H., Pickering, M., & Cleland, A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin and Review*, 6(4), 635–640.
- Branigan, H., Pickering, M., Stewart, A., & McLean, J. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition*, 28(8), 1297–1302.
- Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482.
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Kršmer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen.
- Brown, M., Savova, V., & Gibson, E. (2012). Syntax encodes information structure: Evidence from online reading comprehension. *Journal of Memory and Language*, 66, 194–209.
- Brown-Schmidt, S., & Tanenhaus, M. (2004). Priming and alignment: Mechanism or consequence? *Behavioral and Brain Sciences*, 27(2), 193–194.
- Butler, L., Jaeger, T. F., Bohnemeyer, J., & Furth, K. (2011). Learning to express visual contrasts in the production of referring expressions in yucatec maya. In K. van Deemter, A. Gatt, R. van Gompel, & E. Krahmer (Eds.), *Proceedings of the CogSci workshop on the Production of Referring Expressions: Bridging the gap between computational, empirical and theoretical approaches to reference (PRE-CogSci 2011)*. Austin, TX: Cognitive Science Society.
- Chang, F., Dell, G. S., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272.
- Chang, F., Dell, G. S., Bock, J. K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29(2), 217–230.
- Chang, F., Janciuskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass*, 6(5), 259–278.
- Clark, H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Clark, A. (in press). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*.
- Clarkson, P., & Robinson, A. (1997). Language model adaptation using mixtures and an exponentially decaying cache. In *Proceedings of the 1997 international conference of acoustics, speech, and signal processing* (Vol. 1, pp. 799–802). Munich, Germany.
- Clayards, M., Tanenhaus, M., Aslin, R., & Jacobs, R. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804–809.
- Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, 10(7), 294–300. <<http://www.ncbi.nlm.nih.gov/pubmed/16773323>>.
- Dayan, P., & Yu, P. (2001). ACh, uncertainty, and cortical inference. *Advances in neural information processing systems (NIPS)* (Vol. 14, pp. 189–196). Cambridge, MA: MIT Press.
- Dell, G., Oppenheim, G., & Kittredge, A. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, 23(4), 583–608.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193–210.
- Dubey, A., Keller, F., & Sturt, P. (2008). A probabilistic corpus-based model of syntactic parallelism. *Cognition*, 109(3), 326–344.
- Eisner, F., & McQueen, J. (2005). The specificity of perceptual learning in speech processing. *Attention, Perception, & Psychophysics*, 67(2), 224–238.
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.
- Farmer, T., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (2011). Phonological typicality influences sentence processing in predictive contexts: A reply to Staub et al. (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1318–1325.
- Ferreira, V. S. (2003). The persistence of optional complementizer mention: Why saying a “that” is not saying “that” at all. *Journal of Memory and Language*, 48, 379–398.
- Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. *Learning and Motivation*, 49, 209–246.
- Ferreira, V., & Bock, K. (2006). The functions of structural priming. *Language and Cognitive Processes*, 21(7–8), 1011–1029.
- Ferreira, V., Bock, K., Wilson, M., & Cohen, N. (2008). Memory for syntax despite amnesia. *Psychological Science*, 19(9), 940.
- Ferreira, V. S., Kleinman, D., Kraljic, T., & Siu, Y. (2011). Do priming effects in dialogue reflect partner- or task-based expectations? *Psychonomic Bulletin & Review*, 1–8.
- Fine, A. B., & Jaeger, T. F. (in press). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*.
- Fine, A. B., Qian, T., Jaeger, T. F., & Jacobs, R. (2010). Is there syntactic adaptation in language comprehension? In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics (ACL): Workshop on cognitive modeling and computational linguistics*. Uppsala, Sweden.
- Fine, A. B., Jaeger, T. F., Farmer, T., & Qian, T. (submitted for publication). Rapid expectation adaptation during syntactic comprehension.
- Finegan, E., & Biber, D. (2001). Register variation and social dialect variation: The register axiom. In P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 235–267). Cambridge, UK: Cambridge University Press.
- Fine, A. B., & Jaeger, T. F. (2011). Language comprehension is sensitive to changes in the reliability of lexical cues. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *The 33rd annual meeting of the Cognitive Science Society (CogSci11)* (pp. 925–930). Austin, TX: Cognitive Science Society.
- Frank, A. (2010). *Integrating linguistic, motor, and perceptual information in language production*. Unpublished doctoral dissertation. University of Rochester.
- Frank, S., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, 22(6), 829–834.
- Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *The 30th annual meeting of the Cognitive Science Society (CogSci08)* (pp. 939–944). Austin, TX: Cognitive Science Society.
- Garnsey, S. M., Pearlmutter, N. J., Meyers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Godfrey, J., Holliman, E., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Proceedings of the*

- 1992 international conference on acoustics, speech, and signal processing (ICASSP-92) (Vol. 1, pp. 517–520).
- Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4), 365–399.
- Guenther, F. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3), 594.
- Hanulíková, A., van Alphen, P., van Goch, M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience* (24), 878–887.
- Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, 48(2), 281–303.
- Harrell, F. E. J. (2001). *Regression modeling strategies*. Oxford: Springer-Verlag.
- Hartsuiker, R., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, 58(2), 214–238.
- Hartsuiker, R., & Kolk, H. (1998). Syntactic persistence in Dutch. *Language and Speech*, 41(2), 143–184.
- Hartsuiker, R., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, 75(2), B27–B39.
- Heller, J., & Pierrehumbert, J. B. (2011). Word burstiness improves models of word reduction in spontaneous speech. In *Architectures and mechanisms for language processing (AMLAP 2011)*. Paris, France.
- Houde, J., & Jordan, M. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216.
- Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. Phd thesis. Stanford, CA: Stanford University.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61, 23–62.
- Jaeger, T. F. (2011). Corpus-based research on language production: Information density and reducible subject relatives. In E. M. Benders & J. E. Arnold (Eds.), *Language from a cognitive perspective: Grammar, usage, and processing. studies in honor of tom wasow* (pp. 161–197). Stanford: CSLI Publications.
- Jaeger, T. F., & Snider, N. E. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In *Proceedings of the 29th annual Cognitive Science Society (CogSci09)* (pp. 1061–1066). Austin, TX: Cognitive Science Society.
- Jones, J., & Munhall, K. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, 15(19), 1768–1772.
- Kamide, Y., Altmann, G., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156.
- Kamide, Y. (2012). Learning individual talkers structural preferences. *Cognition*, 124(1), 66–71.
- Kaschak, M. (2007). Long-term structural priming affects subsequent patterns of language production. *Memory and Cognition*, 35(5), 925.
- Kaschak, M., & Borreggine, K. (2008). Is long-term structural priming affected by patterns of experience with individual verbs? *Journal of Memory and Language*, 58, 862–878.
- Kaschak, M., & Glenberg, A. (2004a). Interactive alignment: Priming or memory retrieval? *Behavioral and Brain Sciences*, 27(2), 201–202.
- Kaschak, M., & Glenberg, A. (2004b). This construction needs learned. *Journal of Experimental Psychology: General*, 133(3), 450.
- Kaschak, M., Kutta, T., & Jones, J. (2011). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin & Review*, 1–7.
- Kaschak, M., Kutta, T., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, 39, 381–388.
- Kaschak, M., Loney, R., & Borreggine, K. (2006). Recent experience affects the strength of structural priming. *Cognition*, 99(3), 73–82.
- Kleinschmidt, D., & Jaeger, T. F. (2012). A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. In *Proceedings of the 34th annual meeting of the Cognitive Science Society (CogSci12)* (pp. 107–115). Austin, TX: Cognitive Science Society.
- Kraljic, T., Brennan, S., & Samuel, A. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107(1), 54–81.
- Kraljic, T., & Samuel, A. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS)* (Vol. 19, pp. 849–856). Cambridge, MA: MIT Press.
- Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- MacDonald, J. L., Bock, J. K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25(2), 188–230.
- MacDonald, M., & Christiansen, M. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and waters and caplan (1996). *Psychological Review*, 109, 35–54.
- MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.
- MacKay, D. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.
- Malhotra, G. (2009). *Dynamics of structural priming*. Unpublished doctoral dissertation. Edinburgh, UK: University of Edinburgh.
- Maurits, L., Perfors, A., & Navarro, D. (2010). Why are some word orders more common than others? A uniform information density account. *Advances in neural information processing systems (NIPS)* (Vol. 23, pp. 1585–1593). Cambridge, MA: MIT Press.
- Maye, J., Aslin, R., & Tanenhaus, M. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543–562.
- McDonald, S., & Shillcock, R. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6), 648–652.
- Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357.
- Perkell, J., Guenther, F., Lane, H., Matthies, M., Stockmann, E., Tiede, M., et al. (2004). The distinctness of speakers productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, 116, 2338–2344.
- Pickering, M., & Garrod, S. (in press). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*.
- Pickering, M., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39, 633–651.
- Pickering, M., Branigan, H., Cleland, A., & Stewart, A. (2000). Activation of syntactic information during language production. *Journal of Psycholinguistic Research*, 29(2), 205–216.
- Pickering, M., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134(3), 427.
- Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), 169–190.
- Plaut, D., McClelland, J., Seidenberg, M., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115.
- Qian, T., & Jaeger, T. F. (2012). Cue effectiveness in communicatively efficient discourse production. *Cognitive Science*, 36(7), 1312–1336.
- Qian, T., Jaeger, T. F., & Aslin, R. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology*, 3, 1–9.
- Rayner, K., & Duffy, S. (1988). On-line comprehension processes and eye movements in reading. *Reading Research: Advances in Theory and Practice*, 13–66.
- Recchia, G. (2007). *Strata: Search tools for richly annotated and time-aligned linguistic data*. Undergraduate honors thesis.
- Reitter, D., Keller, F., & Moore, J. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, 35, 587–637.
- Reitter, D., & Moore, J. (2007). Predicting success in dialogue. In *Proceedings of the 45th annual meeting of the Association of Computational Linguistics (ACL)* (pp. 808–815). Prague, Czech Republic: Association for Computational Linguistics.
- Reitter, D., Moore, J., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of the 28th annual conference of the Cognitive Science Society (CogSci06)*. Austin, TX: Cognitive Science Society.
- Rescorla, R., & Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black & Prokasy (Eds.), *Classical conditioning ii: Current research and theory*. New York, NY: Appleton-Century-Crofts.

- Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61, 127–159.
- Roche, J., Dale, R., & Kreuz, R. (in preparation). Dont rush the navigator: Disambiguating strategies require cognitive flexibility.
- Rohde, D. (1999). Tgrep2 manual [computer software manual]. <<http://tedlab.mit.edu/~dr/Tgrep2/tgrep2.pdf>>.
- Roland, D., Dick, F., & Elman, J. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3), 348–379.
- Rowland, C., Chang, F., Ambridge, B., Pine, J., & Lieven, E. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition*, 125(1), 49–63.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
- Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, 89, 179–205.
- Shannon, C. (1948). A mathematical theory of communications. *Bell Systems Technical Journal*, 27(4), 623–656.
- Sigley, R. (1997). The influence of formality and channel on relative pronoun choice in New Zealand English. *English Language and Linguistics*, 1(2), 207–232.
- Smith, N., & Levy, R. (2008). Optimal processing times in reading: A formal model and empirical investigation. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *The 30th annual meeting of the Cognitive Science Society (CogSci08)* (pp. 595–600). Austin, TX: Cognitive Science Society.
- Snider, N. (2008). *An exemplar model of syntactic priming*. Unpublished doctoral dissertation. Stanford, CA: Stanford University.
- Sonderegger, M., & Yu, A. (2010). A rational account of perceptual compensation for coarticulation. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society (CogSci10)* (pp. 375–380). Austin, TX: Cognitive Science Society.
- Staub, A., & Clifton, C. J. (2006). Syntactic prediction in language comprehension: Evidence from either ... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 425–436.
- Szmrecsányi, B. M. (2005). Language users as creators of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, 1, 113–149.
- Tagliamonte, S., & Smith, J. (2005). No momentary fancy! The zero in English dialects. *English Language and Linguistics*, 9(2), 289–309.
- 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.
- Traxler, M. (2008). Lexically independent priming in online sentence comprehension. *Psychonomic Bulletin and Review*, 15(1), 149–155.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528–553.
- van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1483.
- van Son, R. J. J. H., & van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47(1), 100–123.
- Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45(3), 572–577.
- Warker, J., & Dell, G. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 387.
- Wasow, T., Jaeger, T. F., & Orr, D. (2011). Lexical variation in relativizer frequency. In H. Wiese & H. Simon (Eds.), *Proceedings of the workshop on expecting the unexpected: Exceptions in grammar at the 27th annual meeting of the German Linguistic Association* (pp. 175–196). Berlin/New York: Mouton de Gruyter.
- Weatherholz, K., Campbell-Kibler, K., & Jaeger, T.F. (2012). Syntactic alignment is mediated by social perception and conflict management. In *Architectures and mechanisms for language processing (AMLap 2012)*. Riva del Garda, Italy.
- Wei, K., & Körding, K. (2009). Relevance of error: What drives motor adaptation? *Journal of Neurophysiology*, 101(2), 655–664.
- Weiner, E. J., & Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics*, 19, 29–58.
- Wells, J., Christiansen, M., Race, D., Acheson, D., & MacDonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58(2), 250–271.
- Wolpert, D. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209–216.
- Wolpert, D., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7–8), 1317–1329.
- Yu, A., & Cohen, J. (2008). Sequential effects: Superstition or rational behavior? *Advances in neural information processing systems (NIPS)* (Vol. 21, pp. 1873–1880). Cambridge, MA: MIT Press.