



Comprehension priming as rational expectation for repetition: Evidence from syntactic processing



Mark Myslín*, Roger Levy

Department of Linguistics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0108, United States

ARTICLE INFO

Article history:

Received 23 June 2013

Revised 7 October 2015

Accepted 31 October 2015

Keywords:

Psycholinguistics

Priming

Expectation

Rational analysis

Syntax

Probabilistic models of cognition

ABSTRACT

Why do comprehenders process repeated stimuli more rapidly than novel stimuli? We consider an *adaptive* explanation for why such facilitation may be beneficial: priming is a consequence of *expectation for repetition* due to rational adaptation to the environment. If occurrences of a stimulus cluster in time, given one occurrence it is rational to expect a second occurrence closely following. Leveraging such knowledge may be particularly useful in online processing of language, where pervasive clustering may help comprehenders negotiate the considerable challenge of continual expectation update at multiple levels of linguistic structure and environmental variability. We test this account in the domain of structural priming in syntax, making use of the sentential complement–direct object (SC–DO) ambiguity. We first show that sentences containing SC continuations cluster in natural language, motivating an expectation for repetition of this structure. Second, we show that comprehenders are indeed sensitive to the syntactic clustering properties of their current environment. In a series of between-groups self-paced reading studies, we find that participants who are exposed to clusters of SC sentences subsequently process repetitions of SC structure more rapidly than participants who are exposed to the same number of SCs spaced in time, and attribute the difference to the learned degree of expectation for repetition. We model this behavior through Bayesian belief update, showing that (the optimal degree of) sensitivity to clustering properties of syntactic structures is indeed learnable through experience. Comprehension priming effects are thus consistent with rational expectation for repetition based on adaptation to the linguistic environment.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Repetition facilitates processing. Human comprehenders process words, pictures, faces, and everyday environmental sounds more rapidly when these stimuli are immediate repetitions than when they are novel (for review, see [Bigand, Tillmann, Poulin-Charronnat, & Mandlerier, 2005](#)). Why do comprehenders remain prepared to process a stimulus after its first presentation? We consider an *adaptive*, computational-level account of why such facilitation may be beneficial: priming is a consequence of *expectation for repetition* due to rational adaptation to the environment ([Anderson, 1990; Marr, 1982](#)). Clustering of repeated events in time, rather than uniform spacing, is pervasive in human dynamics, from economic transactions to instant messages to the occurrence of words in newspaper headlines over time ([Anderson & Schooler, 1991; Vazquez et al., 2006](#)). Given such clustering, it is rational for comprehenders to increase their expectations for

another instance of an event closely following a first occurrence. Where possible, then, it would be adaptive for comprehenders to learn and deploy knowledge of the clustering properties of the current environment, coming to more strongly expect repetitions in environments where stimuli cluster than in those where they do not.

Leveraging such knowledge may be particularly useful in the domain of language comprehension. Online linguistic processing is an incredibly complex cognitive feat, requiring comprehenders to continually update expectations at multiple levels of structure while negotiating considerable environmental variability. It is also the case that language is the naturalistic clustering environment par excellence, wherein tokens of the same type often occur in closer succession than predicted by chance. Such structure in language, if comprehenders are sensitive to it, may provide invaluable cues in forming accurate expectations of upcoming input, allowing for efficient language comprehension.

Here we present the first test of this account in online language processing, in the domain of sentence processing. It is known that processing of a sentence is faster if its syntactic structure is

* Corresponding author.

E-mail addresses: mmyslin@ucsd.edu (M. Myslín), rlevy@ucsd.edu (R. Levy).

repeated from a preceding sentence. Consider the well-studied sentential complement–direct object (SC–DO) ambiguity (Garnsey, Pearlmutter, Myers, & Lotocky, 1997)¹:

-
- (1) Her friend whispered the solution
 (a) was to dispose *Sentential complement (SC)*
 of the evidence.
 (b) very quietly in her ear. *Direct object (DO)*
-

In this context, verbs such as *whispered* in (1) may subcategorize for one of two syntactic structures, sentential complements as in (a) or direct objects as in (b). Even controlling for factors such as verb repetition and subcategorization bias, comprehenders who have recently encountered the SC structure process subsequent SCs more rapidly (e.g. Fine, Qian, Jaeger, & Jacobs, 2010). If this facilitation is due to an adaptive, rational expectation for repetition of SC structures across sentences, it should be most robust in environments in which SC sentences are very likely to follow SC sentences (regardless of the total number of SCs in the environment). In this paper, we show for the first time that manipulating the clustering properties of the environment indeed affects processing of syntactic structures, such that comprehenders in an environment in which SC sentences cluster process repeated SCs more rapidly than comprehenders in anti-clustering environments. We also present a Bayesian belief-updating model that shows that the relative importance of clustering properties in the environment, as well as the particular shape of the current clustering properties, are indeed learnable through experience. These results support a rational expectation adaptation account in which facilitation of repeated structure is due to adaptation to general environmental experience.

The paper is structured as follows. Section 2 reviews evidence for structural priming in syntactic processing and surveys existing accounts of the phenomenon. Section 3 proposes a rational expectation-based account of these effects. Section 4 presents a corpus study showing that SC sentences indeed cluster in natural language, motivating such an expectation for repetition. Sections 5 and 6 show that comprehenders are indeed sensitive to the clustering properties of the environment through a series of self-paced reading experiments in which clustered and anti-clustered experience had differential effects on the processing of repeated SC structure. Section 7 presents a Bayesian belief-updating model of this adaptation that shows that sensitivity to clustering properties is learnable through experience. Section 8 discusses theoretical implications of an expectation-based account of priming, and Section 9 concludes.

2. Structural priming in syntax comprehension

2.1. Experimental evidence

Repeated syntactic structure facilitates comprehension. Sentences that repeat structure from previously-comprehended sentences are read faster, elicit smaller changes in brain activity, and are rated as more grammatical than sentences that do not repeat structure. For example, reduced relatives are read faster following a reduced relative prime than when following a main verb prime (Pickering & Traxler, 2004; Traxler & Tooley, 2008). These kinds of effects have been experimentally shown to persist for up to sev-

eral weeks (e.g. Wells, Christiansen, Race, Acheson, & Macdonald, 2009), may be elicited with as few as just one prime sentence (e.g. Fine et al., 2010), and are usually not dependent on repetition of particular verbs (Thothathiri & Snedeker, 2008, although this does usually result in a *lexical boost* of priming). Further, interpretations of ambiguous structures are influenced by recently comprehended structures: for example, comprehenders are more likely to choose high-attached interpretations of prepositional phrases after reading a prime expression with a high-attached interpretation, or even after seeing mathematical expressions with analogous parenthetical groupings, suggesting domain generality of structural priming (Branigan, Pickering, & McLean, 2005; Scheepers et al., 2011). The now relatively extensive literature on structural priming in comprehension is reviewed in Pickering and Ferreira (2008) and Tooley and Traxler (2010).

2.2. Theoretical accounts of structural priming

Most accounts of structural priming are cast at Marr's (1982) algorithmic level of analysis, falling into two broad classes: *residual activation* and *implicit learning*. Residual activation accounts of priming, ported from production research to comprehension research, hold that accessing a particular syntactic structure increases that structure's mental activation level for a brief period but rapidly decays, leading to speeded processing of subsequent tokens of the structure (e.g. Pickering & Branigan, 1998). Implicit learning accounts, on the other hand, stipulate that processing a structure leads to unconscious learning of its associated representation, and the amount of exposure determines the strength of learning and ease of subsequent processing (Bock & Griffin, 2000; Chang, Dell, & Bock, 2006). *Dual mechanism* accounts argue that residual activation explains short-term, lexically driven priming, while implicit learning explains longer term, lexically independent priming (Hartsuiker, Beronlet, Schoonbaert, Speybroeck, & Vanderelst, 2008; Reitter, Keller, & Moore, 2011).

Building on implicit learning accounts at the computational level of analysis are *expectation adaptation* accounts, seeking to explain the adaptive benefits of these behaviors. Starting with the premise that context-specific comprehender expectations for upcoming syntactic structures affect processing (the *SURPRISAL* theory; Hale, 2001; Jurafsky, 1996; Levy, 2008), and given the objective of easing processing and allocating resources efficiently, a rational behavior is for these expectations to converge on the statistics of the environment (argued in detail below and in Anderson (1990) and Fine, Jaeger, Farmer, & Qian (2013)). Recent evidence suggests that these expectations can be modulated in the same ways that classic structural priming has been seen to operate. Fine et al. (2013) show that given a verb that may occur as a main verb or as the verb in a relative clause, comprehender expectations initially reflect their prior experience that main verbs are the more frequent continuation, but the more relative clause continuations recently experienced, the more strongly comprehenders come to expect relative clauses. Similar results were obtained by Kaschak and Glenberg (2004), where processing of a novel syntactic construction (*needs done*) sped up with additional exposures over a single experimental session; by Wells et al. (2009), where object relative clauses became easier with more exposure over several weeks; by Fine et al. (2010), where sentential complements (SCs) were processed faster the more they had occurred in recent experience; by Farmer, Fine, and Jaeger (2011), where comprehenders rapidly learned environment-specific verb biases for syntactic continuations; and by Kamide (2012), where comprehenders learned syntactic preferences of individual speakers. Rapid expectation adaptation has also been demonstrated in speech perception (Kleinschmidt & Jaeger, 2011; Kraljic, Samuel, & Brennan, 2008), prosody (Kurumada, Brown, & Tanenhaus, 2012), and pragmatics

¹ Although our case study involves a syntactic ambiguity, our theoretical arguments do not implicate ambiguity resolution as a specialized process. We take a surprisal-based view of syntactic processing, in which structural ambiguity *per se* does not influence processing difficulty, and instead is relevant only insofar as its effect on conditional word probabilities (Levy, 2008, 2013). Thus, in principle, our investigation of syntactic repetition applies to any arbitrary syntactic structure, whether or not it participates in an ambiguity.

(Yildirim, Degen, Tanenhaus, & Jaeger, 2013); this literature is reviewed more comprehensively in Fine et al. (2013).

A further argument for the structural priming as a rational expectation adaptation comes from variability in the strength of priming. Jaeger and Snider (2013) and Fine and Jaeger (2013) observe, in production and comprehension respectively, that this strength is affected by *prediction error*: the more surprising a structure, the more strongly subsequent occurrences are facilitated. Syntactic priming may thus result, they argue, from an effort to minimize expected prediction error for future occurrences of a structure: in order to converge on the statistics of the environment, language users adapt their expectations more dramatically the more unexpected the structure they encounter.

Such a proposal leaves open several possibilities regarding the precise nature of priming as expectation adaptation. On the one hand, expectations themselves are understood to be conditioned finely on preceding context; on the other hand, priming is understood to be the modulation of expectations based on recent experience—that is, based on preceding context. Two rather different—but not mutually incompatible—views of priming as expectation adaptation are thus possible: one in which a prime's effect is an overt shift of the parameters governing a comprehender's distributional expectations, the other in which a prime's effect is to create a conditioning context in which the comprehender's expectations are stronger for repetition of the prime's structure than they would be in different contexts (in other words, comprehenders leverage a form of syntactic dispersion; Jaeger & Snider, 2013:74). In the next section, we build on the expectation-adaptation view of priming by elaborating on these possibilities.

3. Rational expectation for repetition

In this section, we build on the computational-level accounts above to propose that an adaptive motivation for comprehenders to process repeated structures more efficiently is a *rational expectation for repetition* due to the statistics of the ordinary linguistic environment, but that this expectation itself can be modulated given the proper environment. After developing this idea in Section 3.1, we review evidence for its plausibility in Sections 3.2 and 3.3 and preview our empirical and modeling work in Section 3.4.

3.1. Implicit learning, expectation adaptation, and context

As we describe in Section 2.2, Jaeger, Fine, and colleagues refine the notion of priming as implicit learning by pointing out that error-driven adjustment of expectations would be efficient for a rational, adaptive language-processing architecture operating in a changing environment. However, this proposal leaves underspecified the precise nature of priming as inducing shifts in distributional expectations, as conditioning context, or as both. In this section, we consider three possible versions of the expectation-adaptation account with respect to these issues. For clarity, we focus in this discussion on the question of what expectations exist for a structure T the next time there is an opportunity for it to occur, as a function of whether and where T has occurred in the recent history of opportunities for it. For precision, we also briefly employ the notation of probability theory, denoting by $P(T|Context)$ the strength of expectation for T given preceding context.²

In one version of the expectation-adaptation account, the effect of primes—that is, of the occurrence of instances of T in the recent history of opportunities for it—is to change the parameters governing the expectation for T . Critically, in this account, expectations for T are “unigram” with respect to this history of opportunities: if we denote the strength of expectation as $P(T|Context)$, then previous opportunities for T are not part of *Context*. Rather, the effect of the prime is to update the parameters governing this expectation so that it is stronger than it was before the prime—that is, that the distribution of expectations $P(T|Context)$ is rapidly and rationally adjusted. An analogy here would be the repeated rolling of a die that you know is loaded but whose loading you do not know: every time you see it come up six you will expect more strongly that it will come up six in the future because you gain information about the loading, even though the outcomes of consecutive rolls are not causally related in your mental model of how the die falls. This view leads to a natural account of cumulativity effects, in which the strength of priming increases with the number of primes that a speaker is exposed to, regardless of their temporal ordering (Fine et al., 2010, 2013; Fine & Jaeger, 2013; Jaeger, 2008; Jaeger & Snider, 2013; Kaschak, Loney, & Borreggine, 2006; Reitter et al., 2011).

An alternative account is that empirical “priming” effects are simply conditioning on context: language users track co-occurrences of structure across multiple utterances, and their expectations are modulated by what structures occur in earlier utterances, but that the distribution of these context-specific expectations is not rapidly adjusted. That is, language users' syntactic expectations are *higher-order* than unigram with respect to the history of opportunities: if we denote the strength of expectation as $P(T|Context)$, then the recent history of opportunities for T is part of *Context*, but the parameters of $P(T|Context)$ are not rapidly updated. Suppose that this is true, and that comprehenders' expectations are adapted—slowly, over the long term—to the statistics of the linguistic environment. *A priori*, we could just as well see what look like anti-priming effects as what look like priming effects. If structures are positively autocorrelated, or *clustered*, over time, then we would expect to see priming effects, since having T in context makes it more likely that T will appear at the next opportunity. But if structures are negatively autocorrelated, or *anti-clustered*, over time, we would expect to see anti-priming effects, since having T in context makes it *less* likely that T will appear at the next opportunity—language users would not expect repetition. Therefore, if structures are indeed empirically clustered over time in naturalistic language use, it would lend support to this account; we review evidence for such clustering in Section 3.2. Note, however, that this account would to some extent undermine the view of priming as implicit learning: the model generating expectations is not being adjusted rapidly, and structural priming effect simply reflects context sensitivity.

A third version of the expectation-adaptation account would hold that primes have *both* effects: not only do structural expectations for T involve conditioning on the recent history of opportunities for it, but language users also rapidly adapt their expectation distribution $P(T|Context)$ —where the recent history of opportunities for T is part of *Context*—to fit statistics of the local linguistic environment. This proposal would require evidence not only that structures are clustered in naturalistic language use, but that language users adapt to whether this higher-order clustering structure exists in the local linguistic environment—including the ability for comprehenders to *learn to expect non-repetition*. Such evidence (in Section 3.3 we review suggestive cases) would rescue implicit learning accounts of priming, and would have two additional implications. First, such evidence would be particularly difficult for a purely residual-activation theory to accommodate, because it would show that strength of priming can be influenced

² We note, however, that the theory we develop here does *not* depend on expectations actually being probability distributions: in particular, the discussion in this section remains valid even if expectations violate the axioms of properness (i.e., if the same total level of expectation is not available in all contexts) and/or disjoint union.

by higher-order statistics of the linguistic environment that could not be represented by raw activation levels. Second, it would imply that classic, ‘vanilla’ priming effects (where, with no other preceding context, structure T in utterance i facilitates T in utterance $i + 1$) are most likely driven not only by implicit learning but also by contextual conditioning.

3.2. Clustering of repeated events in natural language

Temporal clustering of events in natural language would motivate an expectation for nearby repetition given one occurrence of an event. We now review evidence that if a word or structure occurs once, it indeed has a higher-than-chance probability of occurring again very soon. Anderson and Schooler (1991) provide an early demonstration of this phenomenon in language and beyond, showing that probability of certain naturalistic events, calculated based on intervals between past occurrences of the event, correlates well with the observed probability of that event recurring at a given time (discussed further in Section 3.3.2). These events include occurrence of particular words in newspaper headlines and child speech corpora as well as arrival of messages to one of the authors’ email inboxes. In this section, we survey even broader evidence for linguistic clustering. We begin with the intuitive example of word tokens, and then shift to syntactic structures, the focus of this paper.

3.2.1. Clustering of words

Tokens of a given word type cluster in natural language. This makes sense, since as speakers in a coherent discourse spend time on a given topic, its associated words are highly likely to occur. Church and Gale (1995) and Church (2000) find that within documents, the probability of additional occurrences of an observed word far exceeds chance assuming independence of each occurrence. DeRoeck, Sarkar, and Garthwaite (2007) show that even frequent function words are subject to some degree of clustering. Altmann, Pierrehumbert, and Motter (2009) show that occurrences of tokens of words do not follow a Poisson process, in which time intervals between occurrences are independent, but are better characterized by a stretched exponential distribution implying clustering in time. Further evidence for clustering of words comes from natural language processing, in which caching models, which assign higher probability to a word given a recent occurrence, lead to substantial performance gains (e.g. Goodman, 2001; Kuhn, 1988).

Comprehenders are sensitive to word clustering properties. Using a series of long texts, Heller, Pierrehumbert, and Rapp (2010) find that words whose tokens were strongly clustered in time were read especially rapidly when they were discourse-old, relative to words that were only weakly clustered. They interpret this result in terms of need probabilities: weakly clustering (that is, more uniformly spaced) words are likely to be needed in all contexts, so there should not be a particularly strong advantage when they are discourse-old. Strongly clustering words, however, are very likely to be needed again after an initial occurrence, so they are read especially rapidly when discourse-old. At the word level, then, linguistic stimuli cluster in natural language, and comprehenders are sensitive to this clustering in online processing.

3.2.2. Clustering of syntactic structures

Syntactic structures, too, cluster in naturally occurring speech. Most corpus studies in this domain have investigated specific, approximately meaning-equivalent syntactic alternations.³ Gries

(2005) investigates the dative alternation (*Susan gave toys to the children* versus *Susan gave the children toys*) and particle placement of transitive phrasal verbs (*John picked up the book* versus *John picked the book up*). For both alternations, the realization of any given token was strongly predicted by the realization of the previous token, with influence decaying logarithmically with number of intervening utterances. Szmrecsanyi (2005) investigated, in addition to the particle placement alternation, the choice between synthetic and analytic comparatives (*John is cleverer than Mary* versus *John is more clever than Mary*) and a choice between two future markers (*John will see Mary* versus *John is going to see Mary*). Jaeger (2008) conducted a similar study for active versus passive constructions. Results for all of these alternations showed the same tendency toward repetition of previous choices that was observed by Gries. In other words, given an alternation, observed syntactic structures within that alternation cluster in time.

The above studies share an important conditioning factor: structures participating in an alternation cluster *given the occurrences of the relevant alternation*. Reitter, Moore, and Keller (2006), in contrast, examined the distribution of phrase structure rules over time in each file of several corpora, finding that rules have a high but decaying repetition probability. Dubey, Sturt, and Keller (2005) similarly find evidence of syntactic parallelism in five constructions in written and spoken corpora. These results provide evidence that clustering of syntactic structures in discourses is a general phenomenon, and not limited to the context of particular alternations. In the next section, we review evidence that comprehenders rapidly adapt to whether this higher-order clustering structure exists in the local environment, for both linguistic and non-linguistic phenomena.

3.3. Learning to expect non-repetition

If comprehenders track higher-order contingencies in the temporal patterning of events, there should exist situations in which they do *not* expect repetition of stimuli, if the stimuli do not cluster in prior (or recent) experience. In these situations, existing priming accounts nevertheless predict facilitation for repeated events, whereas a context-sensitive, rational expectation account predicts some degree of surprise, and thus less facilitation. In this section, we identify two possible sources of expectations for structures *not* to repeat: potentially conscious *discourse-level* expectations (which may have sources other than sequential knowledge) and implicitly learned, purely *distributional* expectations.

3.3.1. Discourse-level expectations for non-repetition

One class of discourse-level expectation is for choice of referring expression. In their work on *repeated-name penalties*, Gordon, Grosz, and Gilliom (1993) find that reading times for repeated full-name (rather than pronominal) mentions of a referent depend on the referent’s discourse prominence, as manipulated through sentence position. In other words, they find that degree of expectation for repetition is a function of discourse knowledge, overriding the facilitation for repetitions that would be contributed by strict priming effects.

Traxler, Foss, Seely, Kaup, and Morris (2000) report another discourse-based modulation of expectation for repetition. Intuitively, given two of a certain referent, comprehenders expect the second one to be introduced with a word such as *another* rather than an identical, repeated noun phrase. Eye movements were tracked during reading of sentences such as *The lumberjack greeted the lumberjack early this morning* versus *The young man greeted the lumberjack early this morning*. In early reading time measures, *lumberjack* was read faster when primed by an identical previous occurrence. However, total sentence reading time did not differ between sentences, indicating that sentences with repetition

³ Although many of these studies frame this clustering as a result of structural priming in production, a variety of competing and complementary explanations are plausible, and we make no claims regarding the underlying sources of syntactic clustering. On this matter, see DuBois (2010), Jaeger and Snider (2013), and Pickering and Ferreira (2008).

induced a slowdown elsewhere (since readers spent less time on *lumberjack*). Thus, while a potentially automatic expectation for repetition appears to be at work in early processing, a repetition penalty is observed in later stages, possibly once higher-level discourse knowledge is incorporated in sequential expectation.

3.3.2. Distributional expectations for non-repetition

The discourse-level phenomena above suggest that high-level linguistic knowledge may produce expectations for non-repetition, in contrast to the prediction made by standard repetition priming accounts insensitive to higher-order structure. Another source of sequential expectations is more domain-general and purely distributional: comprehenders may learn and rapidly adapt to the length of intervals between occurrences of events in recent experience. Anderson and Schooler (1991) propose a rational account of memory retrieval along these lines: memory adapts to the structure of the environment, making use of the experienced distributions of occurrences of a stimulus in order to make the most-likely-to-be-used memories most available at the right times. Thus, memories are retrieved according to their *need probability*, the probability that a particular memory trace is needed at a certain time. This is computed by weighing such environmental factors as frequency, recency, and pattern of prior occurrences. If an item is frequent in past experience, it is rational to assume it has a high probability of occurring again, and to make the memory of that item easily retrievable. Similarly, if repetitions of a stimulus cluster in time, it is rational to expect an immediate repetition given an initial occurrence, and therefore to keep the memory of the item highly accessible for a brief period.

Comprehenders indeed appear to be sensitive to clustering patterns in a wide variety of domains beyond linguistic processing. In an offline two-choice task, Anderson and Whalen (1960) found that participants' responses came to approximate the true repetition probability over time. In online processing, Soetens, Boer, and Huetting (1985) and Cho et al. (2002) found that, given a sequence ABAB, participants were faster to respond to a completion A, which is consistent with the pattern, than B, which is a repetition of the previous stimulus and thus expected to be faster on a simple priming account of individual stimuli. More experience led to increasing reliance on the repetition pattern rather than the identity of the just-prior stimulus, suggesting that participants were learning and adapting to higher-order patterns in the environment. Similar evidence comes from a color search task reported in Thomson, D'Ascenzo, and Milliken (2012): when the probability of a repeated color was high given recent experience, the reaction-time cost of processing a switch trial (in which the color was *not* repeated from the previous trial) was high; when repetition probability was low, switch cost was also lower. Jones and Sieck (2003) find that sensitivity to sequential dependencies also affects more complex behaviors such as categorization, where participants accurately learn repetition probabilities of outcomes. Finally, comprehenders are also sensitive to more complex patterns and longer dependencies: Cleeremans and McClelland (1991) and Alexandre (2010) report gradual implicit learning of complex sequences with contingencies spanning up to three intervening elements, with few participants reporting noticing patterning to the sequences.

Perhaps most crucially for a need probability account in which comprehenders calibrate their expectations to the clustering properties of the environment, there is evidence that comprehenders track the length of the interval between repetitions of a stimulus. Peterson, Hillner, and Saltzman (1962) reported a seemingly paradoxical memory effect: when participants memorized arbitrary associations of words to digits, the longer the delay between presentations of associates during training, the better the recall of these associations when a long period of time had passed before testing (relative to testing after a short delay). In fact, these results

are explicable in terms of need probability: if comprehenders implicitly learn during training that a given associate pair occurs at long intervals, they should expect a long interval between its final occurrence during training and its occurrence during test.

Glenberg (1976) investigated this pattern explicitly, manipulating the interval between the first and second training presentations of an associate pair, and the interval between its second presentation and test. Recall was greatest when these quantities were matched: at short test lags, associations that had been trained at short lags were recalled best, and at long test lags, associations that had been trained at long lags were recalled best. Bahrick (1979) found a similar advantage for recall of Spanish–English word lists. Anderson and Schooler (1991) and Anderson, Tweney, Rivardo, and Duncan (1997) model these and analogous results in terms of need probability, showing that accuracy of recall is strongly predicted by degree of expectation for needing the relevant memory trace at test time, based on the intervals between its prior occurrences.

While the above evidence supports an account in which comprehenders track reoccurrence intervals and modulate their expectations accordingly, this may not be possible in all repetition priming phenomena. Soetens et al. (1985) and Cho et al. (2002) found that at extremely short response–stimulus intervals below 100 ms, comprehenders were faster to process immediately repeated stimuli despite experience with a pattern in which stimuli never repeated. Such a behavior may nonetheless be rational: if comprehenders are unable to adapt to reoccurrence intervals at this timescale of processing, it may be most efficient to expect repetition by default, since in general, environmental stimuli cluster in time. We return to the issue of flexibility of environmental adaptation in Section 8.

In sum, comprehenders seem abundantly sensitive to the clustering properties of (at least non-linguistic) repeated stimuli, and appear to expect repetitions at the interval that has been observed in the environment. These effects are not straightforwardly explicable through standard repetition priming accounts without higher-order learning, which do not predict the increased facilitation at *longer* test lags that is observed when training lags are also long. These results support an explanation of facilitation for repeated structure as the result of adaptation to environmental clustering properties.

3.4. Interim summary & overview of the present studies

To summarize the arguments set forth thus far, on current accounts of syntactic priming, the set of computations leading to facilitated processing of repetition may involve learning higher probabilities of a structure in the current environment, or learning simple conditional probabilities of structures given preceding structures (Section 3.1). We propose that structural priming involves learning *both* properties, so that facilitated repetition is a joint effect of the base rate of a structure's occurrence in the current environment *and* the higher-order temporal clustering properties of the structure in the environment. On this kind of *rational expectation for repetition*, memories are made available according to their need probability, which is computed in part on the basis of pattern of prior exposure. For stimuli that cluster in time, expectation for additional occurrences should be higher given a recent occurrence. Comprehenders indeed seem to exhibit this pattern of expectations in simple recall and sequence processing tasks in a variety of non-linguistic domains. Structural priming effects in syntax comprehension, too, may be the result of rational expectation for repetition, to the extent that tokens of syntactic structure types cluster in natural language.

We investigate this claim directly using the sentential complement–direct object (SC–DO) ambiguity. SC continuations are read

more rapidly when the preceding sentence also contains an SC (Fine et al., 2010). If this is due to rational expectation for repetition, (1) tokens of SCs must cluster in natural language, motivating this expectation, and (2) comprehenders must actually be sensitive to clustering properties of syntactic structures in the environment. We present two studies addressing these respective points. First, we present a corpus study showing that SCs indeed cluster in natural language. Second, we present two self-paced reading experiments showing that comprehenders can adapt rapidly to arbitrary clustering patterns of syntactic structures, and in particular that their expectations for repetition of SCs depends on whether SCs clustered or not in recent experience. A Bayesian belief-updating model of these results shows that these properties and the optimal degree of sensitivity to them are indeed learnable through experience. These results support a rational expectation account in which facilitation of repeated structure is due to adaptation to higher-order temporal dependencies in recent experience in the environment.

4. Corpus study: Clustering of sentential complements in natural language

If speeded processing of repeated sentential complement structures is due to rational expectation for repetition, SCs must cluster in natural language in order to motivate this expectation. We investigate this by examining the distribution of distances between SCs in the parsed Brown Corpus. For each corpus file, we extracted the sentence number of each sentence containing an SC structure⁴ and computed the number of intervening sentences between each SC sentence. The distribution of these distances is plotted in black in Fig. 1(a). In total, 1471 of 35,850 sentences, or 4%, contained SCs.

If SCs were distributed randomly throughout each corpus file and did not cluster, distances between SCs would follow a geometric distribution. In particular, if X is the number of intervening sentences before an SC, and the probability that any given sentence is an SC is p , then the probability that an SC is observed after k intervening sentences is:

$$P(X = k) = (1 - p)^k p \quad (3)$$

As mentioned above, the maximum likelihood estimate of the probability p that a sentence is an SC sentence in the Brown Corpus is 0.04. Entering this value and $k = 0$ into Eq. (3), we find that the chance probability of an immediately repeated SC is 0.04. However, the observed probability of an immediate repetition is far greater, at 0.11. According to a binomial test, this observed proportion is significantly greater than the expected proportion ($p < 0.001$). This trend extends beyond immediate repetitions. In Fig. 1, the null-hypothesis geometric distribution is plotted in red for each value of k up to the observed maximum. At short intervals between SC sentences, observed probabilities of SCs (in black) are greater than chance probabilities (in red), and at long intervals, observed probabilities are smaller than chance probabilities.

While this pattern shows that on the whole, SCs in the Brown Corpus indeed cluster to a greater extent than is expected by chance, it is also consistent with the possibility that individual documents within the corpus vary in their base rate of SCs, and that SCs do not cluster within documents. To address this possibility, for each document, we computed the number of immediate SC repetitions expected by chance if all sentences in the document were randomly ordered, and compared this to the observed number of SC repetitions in each document (Fig. 1b; this computation is detailed in Appendix A). For 106 of the 173 documents, the number

of observed repetitions was greater than the number of expected repetitions, a highly significant pattern according to a paired t -test: $t(172) = 3.89$, $p < 0.0001$.⁵ Thus even within documents, SCs cluster to a greater extent than expected by chance. Given this distribution, it is indeed reasonable for comprehenders to expect additional occurrences of SCs in close succession.

5. Experiment 1: Adapting to the clustering properties of the environment

In our corpus study, we showed that SCs cluster in natural language, making it reasonable for comprehenders to expect additional tokens of SCs immediately following an initial occurrence. We next ask whether comprehenders indeed flexibly and rapidly adapt their expectations for repetition to the temporal clustering properties of syntactic structures in their experience in the current environment.

We present two between-groups self-paced reading experiments, in both of which two groups read the same total number of SCs, but experience different clustering properties. For one group, occurrences of SCs cluster strongly in a training phase, and for the other group, they specifically do not cluster. Participants in both training conditions then read *clusters of two* SC sentences in a test phase; our expectation-based account predicts that the second sentence in these clusters is read faster by the group with clustered training experience, whereas standard structural priming accounts, in which the relevant higher-order contingencies between syntactic structures are not tracked, predict no difference between conditions.

5.1. Participants

Participants in Experiment 1 were recruited via Amazon's Mechanical Turk service. Only users with United States IP addresses who certified that they were native speakers of American English were allowed to participate. Participants were paid between \$0.25–\$1.00 depending on the length of the experiment (see Section 5.2). A total of 192 participants were included.

5.2. Materials

In Experiment 1, stimuli were two-sentence vignettes, where each sentence contained either a sentential complement (SC), direct object (DO), or other, unrelated structure. Unrelated structures could be transitive, but never included verbs that could also subcategorize for SCs. The structure of sentences containing SCs or DOs was temporarily ambiguous, except in half of the SC cases, where a disambiguating *that* was included. Sample stimuli are provided in Table 1; the complete set is provided in Appendix A. Order of sentences within vignettes was constant across all participants and conditions.

The experiment consisted of a training phase and a test phase (Fig. 2). For participants in the *clustered* training condition, training vignettes were SC.SC or Other.Other; SCs always clustered. In *anti-clustered* training, SCs never occurred in the same vignette as other SCs, instead occurring with other, unrelated structures. No DO continuations occurred in training vignettes. One third of vignettes were fillers, and these followed an Other.Other structure. Equal numbers of both total SCs and total Other sentences occurred in both training conditions. Accounting for both critical and filler trials, the ratio of SC sentences to Other sentences was 2:3 in both

⁴ In particular, using `tgrep2`, any sentence matching patterns `sc1-sc4` as reported in Roland, Dick, and Elman (2007) was considered an SC sentence.

⁵ Because the data are not normally distributed, we also performed a binomial test as an alternative to the t -test, with the same qualitative result: the proportion of documents with greater-than-expected repetition is significantly greater than 0.5 ($p < 0.01$).

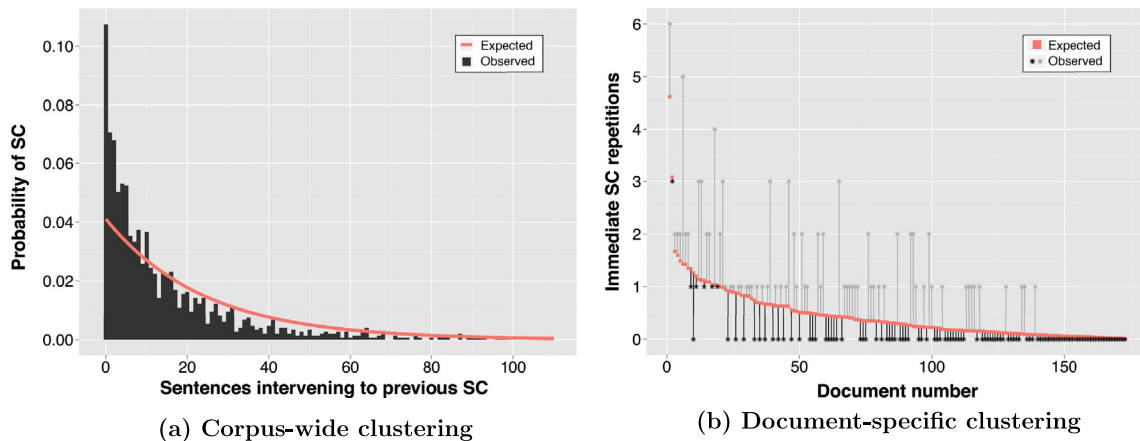


Fig. 1. (a) Distances between sentences containing SCs in the Brown Corpus, computed as number of sentences intervening between SCs. Baseline (geometric distribution) reflecting hypothetical random distribution of SCs throughout the corpus is plotted in red. (b) For each corpus document, observed number of immediate SC repetitions (black or gray stars) versus expected number of repetitions given random ordering of sentences in the document (red squares). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Sample stimuli for Experiment 1.

SC.SC	The chemist observed (that) the mixture was far thicker than expected. She had forgotten (that) the procedure had called for adding water.
SC.DO	The chemist observed (that) the mixture was far thicker than expected. She had forgotten the procedure during her month of vacation.
Other.SC	The muffins Natalie baked turned out extremely dry and hard. She had forgotten (that) the procedure had called for adding water.
SC.Other	The chemist observed (that) the mixture was far thicker than expected. She wondered if she had used the right concentration of chemicals.
Other.	Kara quickly took out her camera and timed her shot perfectly.
Other	The beautiful moment was forever frozen in time on her film roll.

conditions (rounding up to the nearest whole number of filler vignettes as necessary). Length of training was manipulated between groups, and included training lengths of 2, 4, 15, or 30 pairs of critical vignettes (where a pair is defined as ⟨SC.SC, OTHER.OTHER⟩ or ⟨SC.OTHER, OTHER.SC⟩). This resulted in training lengths of 4, 8, 30, or 60 vignettes and the same total number of SCs. The particular SC items in which disambiguating *that* occurred was also counterbalanced between groups. *That*-presence in first versus second sentence position was counterbalanced both within and between groups.

The test phase was equivalent for each group. Test stimuli comprised twelve SC.SC vignettes, twelve SC.DO vignettes, and six Other. Other vignettes. Each SC or DO stimulus throughout the course of the experiment contained a different verb; no critical verb repeated in any participant's experience. Further, the mean SC bias of the verbs (the proportion of tokens of the verb that led to an SC continuation in the Brown Corpus) was balanced throughout the experiment,⁶ and the average sentence length in words was equivalent between sentence types (SC, DO, Other) within each phase.

⁶ Mean SC-bias ranged between 0.34 and 0.41 (sd = 0.14–0.25) for the following categories: training phase sentences in each of the four length conditions, and test phase sentences (i) in first position preceding SCs, (ii) in first position preceding DOs, (iii) in second position resolving as SCs, and (iv) in second position resolving as DOs. The overall proportion of SC outcomes (vs DOs) for tokens of all the verbs we used was 38.9% in the Brown Corpus.

5.3. Procedure

Stimuli were presented to participants using the Ibx web interface for psycholinguistic experiments (Drummond, 2012). Each sentence of a given vignette was presented as a new line of text. The first sentence of the vignette remained visible after the second sentence appeared.

The experiment consisted of a practice stage followed by training and test. Additional filler items were included after practice but before the first critical item. After the initial practice stage, all items were presented in a single block; participants were not alerted to transitions between the various stages. Order of appearance of critical vignettes within training and testing phases was randomized for each participant. In practice and test stages, all stimuli were presented as word-by-word self-paced reading in which participants pressed the spacebar to reveal each successive word, and RTs were defined as the time between a word's presentation and the next spacebar press. In training, however, 1/4 of stimuli were randomly chosen to be presented word-by-word and 3/4 were presented sentence-by-sentence (following the same spacebar-press procedure) in order to keep the duration of the experiment reasonable. In both cases, forward masking with underscores reflected the length of upcoming words. At the conclusion of the experiment, participants were asked to share their impressions and asked what they thought the experiment had investigated.

5.4. Results

5.4.1. Exit survey results

No response indicated any conscious participant awareness of the experimental manipulation, and indeed no response mentioned any properties of the structure or naturalness of the experimental sentences. A majority of participants reported believing that the experiment investigated the general efficiency of reading word-by-word versus sentence-by-sentence.

5.4.2. Reading time results

All RTs that were abnormally low (under 100 ms) or abnormally high (over 5000 ms) were excluded; these represented 1.8% of the data. Outliers more than 3 standard deviations from the overall mean for each region were subsequently removed, which resulted in an additional 1.01% data loss. Remaining RTs were residualized

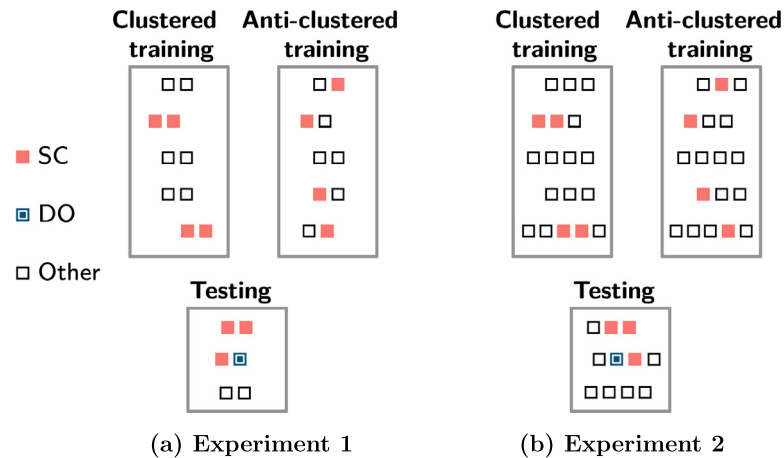


Fig. 2. Structure of the self-paced reading experiments.

to correct for individual differences in word length and participant reading speeds: for each participant, a linear model was fit between word length and reading time, and residual distances were computed between observed RTs and those predicted by the linear model. Critical test sentences were segmented into regions of interest as in Table 2. In all analyses below, we consider reading times for regions in the second sentences in critical test vignettes (vignettes of the form SC.SC and SC.DO). All residual reading times reported and analyzed for multi-word regions are the average of the per-word residual reading times in the region.

In the remainder of this section, we report the results of linear mixed-effects models (Baayen, Davidson, & Bates, 2008) testing the predictions of the expectation-based account. First, we examine reading times post-disambiguation in the second sentence in test vignettes, predicting these SCs (which themselves follow SCs) to be read especially rapidly by participants who experienced the clustered distribution of SCs in training. Second, we examine the role of the disambiguating complementizer *that* in the second SC in test vignettes: participants in clustered training should strongly expect SCs, and so the absence of the additional disambiguating cue provided by *that* should not slow processing as strongly as it should for anti-cluster-trained participants, who are less likely to expect the second SC. Additionally, we predict that these effects should be observable only after some non-trivial length of exposure to a clustering or anti-clustering environment. (As an additional exploration, we investigate whether the size of the effect grows with training length, but do not make a specific prediction about linear or even logarithmic growth, given the nonlinear learning properties reported in Cleeremans & McClelland (1991) and Alexandre (2010) as well as complexities such as the role of fatigue over the course of our experiment.)

We do not make predictions about *where* in the disambiguating part of the sentence expectation-based processing effects should be observed. Our theory is rooted in a surprisal-based view of syntactic processing, in which structural ambiguity *per se* does not influence processing difficulty, and instead is relevant only insofar as its effect on conditional word probabilities (Sections 2.2 and 3; Levy, 2008, 2013). Consequently, the theory does not lead to a specific prediction for the locus of all processing difficulty differential to be, for instance, on the first words in the disambiguation region, instead allowing the possibility of a more diffuse effect. As a further complexity, in word-by-word self-paced reading, much of the processing difficulty imposed by a given word manifests not on reading times for the word itself but 1–3 words

downstream (Mitchell, 1984; Smith & Levy, 2013). For these two reasons, we simply predict that clustering-based learning effects will occur in the part of the sentence where syntactic continuations differ: that is, at some location within the final three regions in Table 2. We present analysis of both (i) a broader region that starts at the disambiguating word and continues to the end of the sentence, and (ii) a narrower, two-word region including the disambiguating word and the immediately following word.

Linear mixed-effects models were fitted that included the maximal random effects structure justified by the design of the experiment (Barr, Levy, Scheepers, & Tily, 2013); these are listed individually for each analysis. We use sum coding for binary fixed effect predictors, and center the one continuous predictor, training length, to have a mean of 0 and standard deviation of 0.5. For each fixed effect predictor, coefficient estimates β and corresponding t -values are reported, where $|t| > 2$ indicates a significant effect at approximately $p < 0.05$, following Baayen et al. (2008).

5.4.3. Test sentence processing by training condition and continuation

Region-by-region residual reading time results for the second sentences in critical vignettes during the test phase are shown in Fig. 3a. Fig. 3b and c are graphs summarizing the 2×2 pattern of residual reading times as a function of (i) $+/-$ CLUSTERING experience during training and (ii) SC versus DO CONTINUATION, first averaged over all training lengths and then broken out separately for each training length. Qualitatively, we see an interactive pattern such that SC continuations are more difficult for anti-clustered training experience, and easier for clustered training experience, with this pattern becoming more visually apparent as training length increases. A single linear mixed-effects model was fit to reading times on the final three regions: Disambiguation, Spillover, and Conclusion (Fig. 3b).⁷ Predictors that were included as fixed effects included centered log-transformed training LENGTH (number of critical vignettes in the training phase), training CLUSTERING experience (clustered or anti-clustered), and CONTINUATION (SC or DO). The log transformation did not change any qualitative patterns of significance in any of the models reported here, but is discussed further in the overall review of training length effects in Section 5.4.5. Random effects included subject- and item-specific intercepts,

⁷ Separate models for each of the three regions were also fit. For each of the Disambiguation and Conclusion regions, significance patterns were identical to those of the linear model reported in Table 3; for the Spillover region, results were numerically similar but the crucial CLUSTERING:CONTINUATION interaction did not reach significance.

Table 2

Regions of interest for SC–DO sentences: *introduction*, *main verb*, *complementizer THAT*, *ambiguous noun phrase*, *disambiguation*, *spillover*, and *conclusion*. *Disambiguation* was defined as the first word following the ambiguous noun phrase, which reveals whether the structure is an SC or DO; *spillover* was the first word following this region.

	INTRO	VERB	THAT	AMBIG NP	DISAMBIG	SPILL	CONCLUSION
SC	The chemist	observed	(that)	the mixture	was	far	thicker than expected.
DO	She had	forgotten		the procedure	during	her	month of vacation.

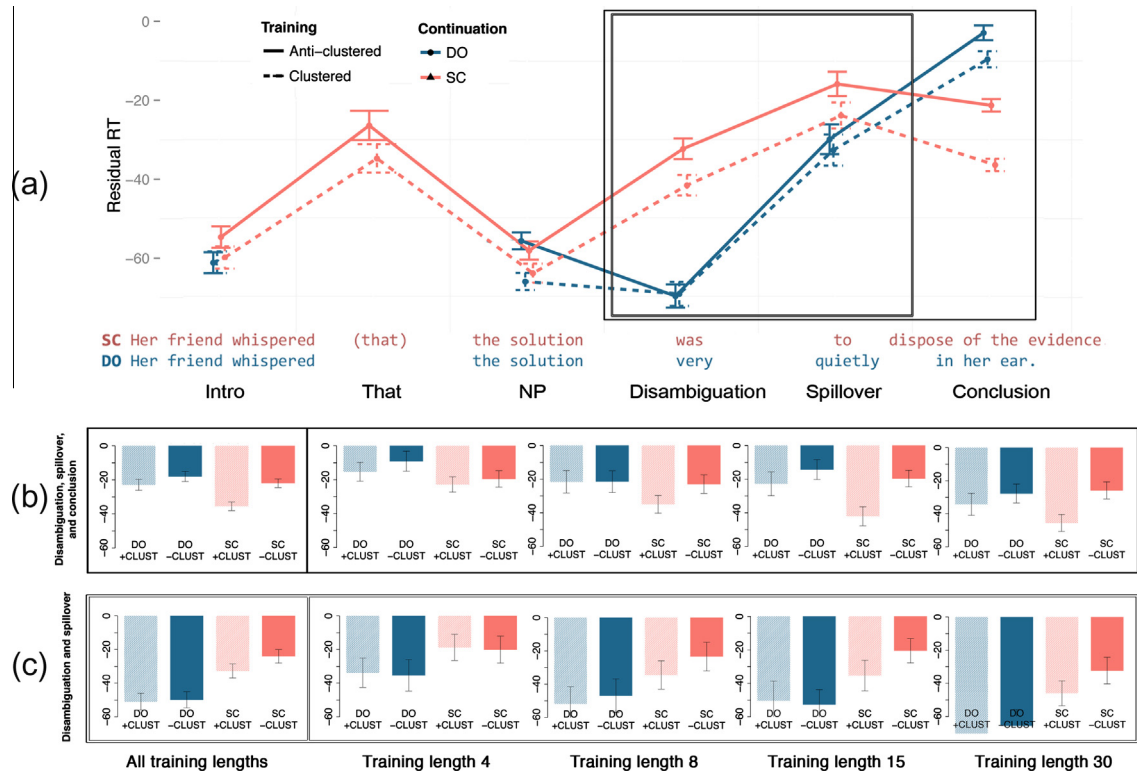


Fig. 3. (a) Region-by-region residual reading times on the second sentence of critical test vignettes in Experiment 1, conditioned on training condition and sentence continuation (all training lengths). (b) Residual reading times aggregated over Disambiguation, Spillover, and Conclusion regions. (c) Residual reading times aggregated over Disambiguation and Spillover regions.

by-subject random slopes for CONTINUATION, and by-item random slopes for LENGTH, CLUSTERING, and their interaction. Results of the linear model are presented in Table 3.

Most crucially for our account of adaptation to the clustering properties of the environment, we predicted an interaction of CLUSTERING during training and CONTINUATION, such that reading times are faster in the SC, clustered condition. Participants trained in the clustering condition should especially strongly expect SCs as the second sentences of test vignettes, which all begin with SC sentences, so reading times should be especially low for SC continuations given clustered training. This effect is significant ($\beta = -8.92, t = -2.77$). To guard against the possibility that our effect was being driven by some interaction of *that*-presence and clustering, we also fit an analogous model excluding items with *that*-present SCs. The critical CLUSTERING:CONTINUATION interaction remained significant in this model: $\beta = -9.92, t = -2.67$.

As a further exploration, we also evaluate a main effect of training LENGTH and interactions of the above effects with length: increased training length could lead to faster reading times as a result of adaptation to the task and greater learning of the patterning of syntactic structures in the environment. No main effect of length emerged ($\beta = -1.21, t = -0.17$), and the interaction of length and continuation was not significant ($\beta = 8.04, t = 1.60$): participants did not speed up for either continuation with increased training length. The three-way interaction of length with

Table 3

Result summary for model of reading times on final three regions of second sentence of critical test vignettes in Experiment 1: coefficient estimates β , standard errors SE (β), and t values.

Predictor	Coef. β	SE(β)	t
Intercept	-23.15	7.79	-2.97
LENGTH	-1.21	7.02	-0.17
CONTINUATION [$DO = -0.5, SC = 0.5$]	-4.26	10.12	-0.42
CLUSTERING [$anti-clustering = -0.5, clustering = 0.5$]	2.08	4.78	0.44
LENGTH: CONTINUATION	8.04	5.05	1.60
LENGTH: CLUSTERING	-19.97	9.70	-2.06
CLUSTERING: CONTINUATION	-8.92	3.22	-2.77
LENGTH: CLUSTERING: CONTINUATION	-12.34	7.32	-1.69

clustering experience and sentence continuation was not significant ($\beta = -12.34, t = -1.69$), although the trend is in the expected direction, such that SCs are (numerically) read more rapidly given increased length of clustered training experience. We return to analysis of length effects in Section 5.4.5.

We also observed an effect for which we had made no prediction. We had no specific expectation that clustered training experience should lead to faster or slower reading times overall than anti-clustered training, and indeed observed no main effect of clustering ($\beta = 2.08, t = 0.44$). However, clustering interacts with length, such that participants speed up with increased length of

clustered training but not anti-clustered training ($\beta = -19.97, t = -2.06$). We return to this point in Section 8.

In order to better understand the locus of the crucial CLUSTERING: CONTINUATION interaction, we fit another model identical to the one above, but excluded the sentence-final Conclusion region, so that only Disambiguation and Spillover were included (Fig. 3c). The results of this model differed from those of the original model in several ways (Table 4): we now observe a main effect of LENGTH ($\beta = -16.10, t = -2.17$), such that participants speed up with increased training. We also observe a theoretically non-meaningful, marginal main effect of CONTINUATION ($\beta = 25.56, t = 1.98$) and interaction with LENGTH ($\beta = 14.60, t = 2.34$): participants read SC continuations more slowly than DO continuations, and this pattern strengthens with increased training length. Effects of CONTINUATION are not theoretically meaningful since the lexical content in SC and DO continuations is totally distinct. Finally, we evaluate the crucial interaction of CLUSTERING and CONTINUATION: for the regions considered in this model, neither this interaction ($\beta = -7.14, t = -1.66$) nor the three-way interaction with LENGTH ($\beta = -7.61, t = -0.87$) reach significance, although they trend in the expected direction, such that comprehenders read repeated SC continuations faster given (increasing amounts of) experience where these continuations cluster. In sum, then, reliable effects of an expectation for repetition are observed immediately at disambiguation as well as at the conclusion of the sentence, though the effect was only marginal in an extended disambiguating region including Disambiguation and Spillover but not Conclusion.

5.4.4. Utility of complementizer *that* as a disambiguation cue

In addition to examining reading times for SC and DO continuations given distinct clustering training experiences, another way to probe comprehender expectation for SCs is to examine the effect of the disambiguating complementizer *that* in SC sentences. Recall that half of SC sentences included *that* and half did not. Given low expectations for an SC continuation, presence of *that* can serve as an important disambiguation cue, such that reading times on the final disambiguation region late in the sentence are substantially increased in its absence. However, if expectations for an SC are high, presence of *that* is less crucial, since comprehenders are not entertaining the possibility of a DO continuation as strongly. Thus for SC reading times, we predict an interaction of *that* presence and training clustering experience, such that absence of *that* slows reading times only given anti-clustered experience (and thus weaker expectation for an SC in the second sentence).

Region-by-region reading time results for SC continuations of second sentences in critical vignettes during the test phase, conditioned on the presence or absence of *that*, are shown in Fig. 4a. A linear mixed-effects model of reading times on the Disambiguation region⁸ was fit with centered log-transformed training LENGTH, training CLUSTERING experience (clustered or anti-clustered), and THAT presence included as fixed effects. Random effects included subject- and item-specific intercepts, by-subject random slopes for THAT, and by-item random slopes for LENGTH, CLUSTERING, and their interaction. Results of the linear model are presented in Table 5.

As expected, we observe a main effect of *that* presence, so that reading times on the final disambiguation region are higher if *that* is absent ($\beta = 10.09, t = 2.32$). The predicted interaction of *that* presence and clustering training experience was not significant ($\beta = -1.70, t = -0.28$). This suggests that comprehenders do not particularly strongly expect DO continuations as opposed to SCs or other structures, even though they could have in principle.

Table 4

Result summary for model of reading times on Disambiguation and Spillover regions of second sentence of critical test vignettes in Experiment 1: coefficient estimates β , standard errors $SE(\beta)$, and t values.

Predictor	Coef. β	$SE(\beta)$	t
Intercept	-51.29	9.60	-5.34
LENGTH	-16.10	7.40	-2.17
CONTINUATION [DO = -0.5, SC = 0.5]	25.56	12.90	1.98
CLUSTERING [anti-clustering = -0.5, clustering = 0.5]	1.50	5.15	0.29
LENGTH: CONTINUATION	14.60	6.24	2.34
LENGTH: CLUSTERING	13.16	10.31	-1.28
CLUSTERING: CONTINUATION	-7.14	4.30	-1.66
LENGTH: CLUSTERING: CONTINUATION	-7.61	8.80	-0.87

This result could reflect a rational response to the statistics of the training environment, in which no DO continuations occurred, as well as expectations for DOs given particular verbs. In line with this explanation, the main effect of *that* presence could result at least in part from spillover effects, since *that* is a short, high-frequency word. As with the model of SC and DO reading times above, length of training produced no significant main effect, and further did not participate in interactions with the effects described here (see Table 5 for details). Likewise, there was no main effect of clustering of training experience ($\beta = -7.77, t = -1.38$).

5.4.5. Effects of training length

We predicted that the effects of learning of the syntactic clustering properties of the environment should be apparent only after some non-trivial length of exposure to the environment. Thus over time, clustered training should confer an increased advantage for processing of SCs following SCs. We plot this in panels (a) and (b) of Fig. 9 as mean RT given anti-clustered training minus mean RT given clustered training, for both SC continuations and DO continuations of second sentences in test (where all first sentences are SCs). (Panel (c) is discussed in Section 7.2.)

Our prediction is borne out; the processing advantage for SCs given clustered training is (numerically) most evident at the longer training lengths. We also observe a (smaller) late advantage for DOs given clustered training. In other words, clustered training leads to general speedups in testing (a point we address in Section 8), but the advantage grows to be much greater for SCs than DOs, after starting at roughly equivalent magnitude at short training lengths. It is noteworthy that no effects of training length appear to be linear: the SC advantage for clustered training initially grows rapidly, then more slowly after training length 4, and not at all after training length 15 (and in fact slightly decreases here, possibly as a result of participant fatigue, leading to decreased observable effects of learning). The DO advantage is similarly non-linear. The log-transformed training lengths used in the linear models above resulted in marginally stronger effects than their non-transformed counterparts, but did not change any patterns of significance. The finding that these learning effects are more roughly logarithmic than linear in time is consistent with the implicit learning of sequential patterns reported in Cleeremans and McClelland (1991) and Alexandre (2010), in which gains in processing speed were most dramatic in early blocks of the experiment.

5.5. Discussion

In Experiment 1, we investigated whether comprehenders adapt their expectations for structural repetition to the syntactic clustering properties of the environment in recent experience. Participants who were trained in a *clustering* environment, in which SC continuations always occurred in clusters of two, came to more

⁸ In line with the linear models in the previous section, we also fit analogous models to each of the three final regions as well as a model including all three regions, but did not find a significant CLUSTERING:THAT interaction.

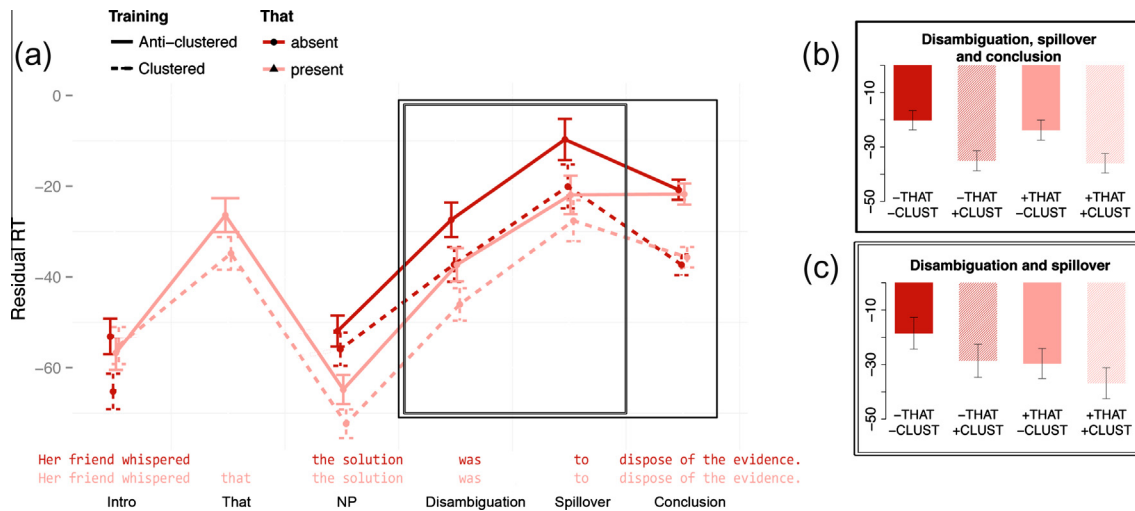


Fig. 4. (a) Region-by-region residual reading times on the second sentence of critical test vignettes of structure SC.SC in Experiment 1, conditioned on training condition and presence of disambiguating complementizer *that*. (b) Residual reading times aggregated over Disambiguation, Spillover, and Conclusion regions (all training lengths). (c) Residual reading times aggregated over Disambiguation and Spillover regions (all training lengths).

strongly expect an additional SC immediately following an initial occurrence than participants who were trained in an *anti-clustering* environment in which SC sentences never occurred before or after other SC sentences. Evidence that comprehenders tracked these between-sentence clustering properties comes from a significant interaction between sentence continuation and clustering experience in training, such that reading times on the disambiguation of temporarily ambiguous SC–DO sentences were especially low for SC continuations given clustered training experience. Further, the processing advantage for SCs following SCs conferred by clustered training became numerically greater than the advantage for DOs following SCs the longer the training period of clustered experience, indicating that knowledge of clustering properties is a function of experience.

Standard structural priming accounts do not predict this differential facilitation for repeated structure based on the clustering properties of syntactic structures in recent experience. On these accounts, given that both groups were exposed to equal total numbers of SCs in the training phase, facilitation for an SC given an immediately preceding SC should be the same for both groups. On our expectation adaptation account, however, clustering properties of the environment are learned and inform expectations about repetition of syntactic structure.

6. Experiment 2: Adaptation within longer contexts

While the results of Experiment 1 are consistent with the view that comprehenders track the clustering properties of syntactic structures in recent experience, an alternative explanation is possible. In particular, the paired nature of the stimulus sentences may have perceptually emphasized clustering or anti-clustering, and participants may have simply learned the pair structure of the experiment rather than the linguistic properties of the environment. In order to investigate this possibility, and to simultaneously test whether our finding is robust to longer, more natural contexts with fewer overall SC structures, we designed a second experiment. This experiment was identical to the first one, but the two-sentence vignettes were now embedded within vignettes ranging between three and five sentences in length. This design thus minimizes any ‘perceptual bookending’ present in Experiment 1.

Table 5

Result summary for model of reading times on disambiguation region of second sentence of critical test vignettes of structure SC.SC in Experiment 1: coefficient estimates β , standard errors $SE(\beta)$, and t values.

Predictor	Coef. β	$SE(\beta)$	t
Intercept	–33.17	11.82	–2.81
LENGTH	–4.65	7.86	–0.59
THAT [<i>present</i> = –0.5, <i>absent</i> = 0.5]	10.09	4.34	2.32
CLUSTERING [<i>anti-clustering</i> = –0.5, <i>clustering</i> = 0.5]	–7.08	5.09	–1.40
LENGTH: THAT	–12.81	8.65	–1.48
LENGTH: CLUSTERING	–18.10	10.97	–1.65
CLUSTERING: THAT	–1.70	6.10	–0.28
LENGTH: CLUSTERING: THAT	8.78	12.20	0.72

6.1. Participants

A total of 148 new Mechanical Turk participants were included (see Section 5.1).

6.2. Materials

Materials were adapted from those used in Experiment 1 by adding one or more structurally unrelated (Other) sentences to the beginning and/or end of the two-sentence vignettes in both the training and testing phases, so that all vignettes ranged between three and five sentences (see Fig. 2). An example is provided below (Other.SC.Other):

- (1) The muffins Natalie baked turned out extremely dry and hard. She had forgotten the procedure had called for adding water. She ended up having to throw them out.

Approximately 80% of vignettes comprised three sentences, 16% comprised four sentences, and 4% comprised five sentences. Further, approximately equal numbers of critical sentence pairs were positioned at the beginning, middle, or end of their embedding vignette. Training lengths included 4, 8, or 15 critical vignette pairs, resulting in equal numbers of total SCs. These training lengths included 1, 2, or 5 filler vignettes, respectively. Because of the substantially higher number of Other structures in this experiment relative to Experiment 1, the ratio of SC to Other sentences was considerably lower, at 1:3. We took two measures to

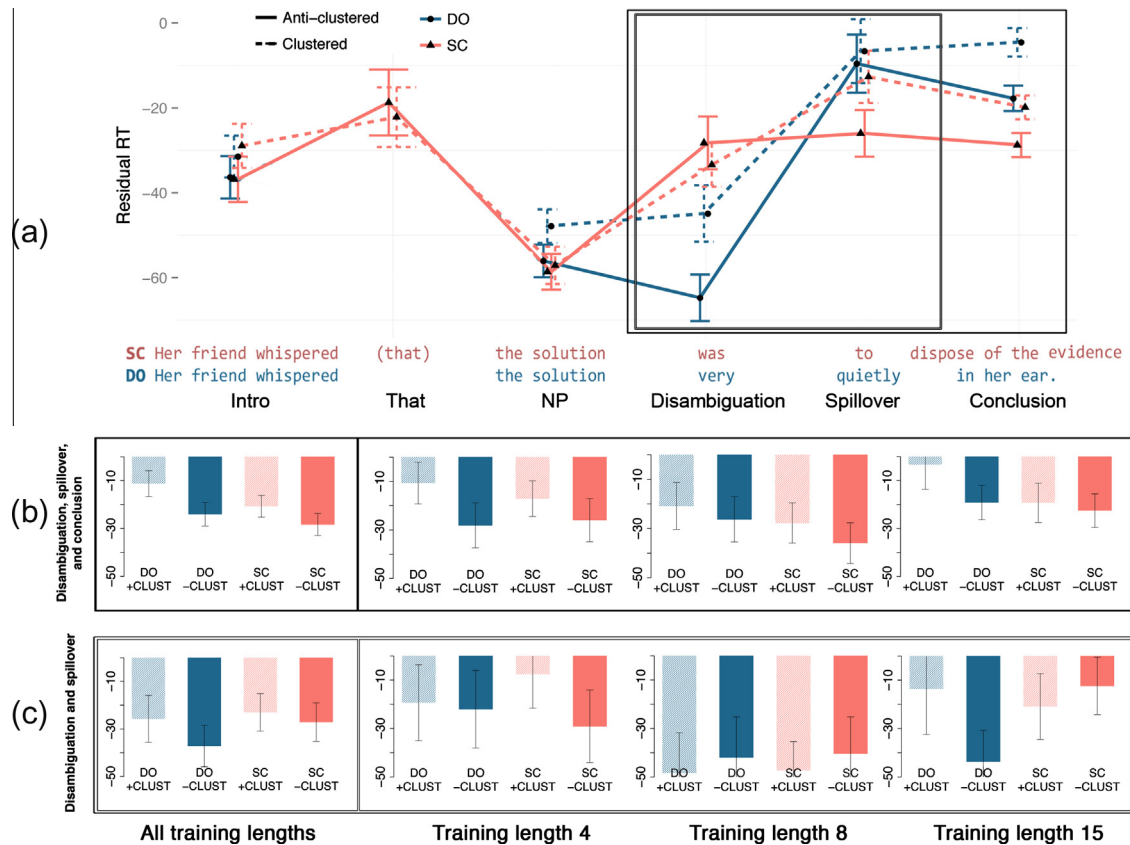


Fig. 5. (a) Region-by-region residual reading times on the second sentence of critical test vignettes in Experiment 2, conditioned on training condition and sentence continuation (all training lengths). (b) Residual reading times aggregated over Disambiguation, Spillover, and Conclusion regions. (c) Residual reading times aggregated over Disambiguation and Spillover regions.

mitigate the substantial lengthening of the experiment resulting from these longer vignettes. In addition to omitting the 30 vignette-pair condition, we reduced the test phase to half the length of the test phase in Experiment 1, presenting each participant with a counterbalanced half of the test items (while maintaining the 2:2:1 SC:SC:SC:DO:Other:Other ratio of Experiment 1).

6.3. Procedure

The procedure was identical to that of Experiment 1 (Section 5.3).

6.4. Results

6.4.1. Exit survey & reading time results

As in Experiment 1, no participant showed evidence of awareness of the manipulation in the exit survey (Section 5.4.1). Again, all RTs that were abnormally low (under 100 ms) or abnormally high (over 5000 ms) were excluded; these represented 2.1% of the data. Outliers more than 3 standard deviations from the overall mean for each region were subsequently removed, which resulted in an additional 0.2% data loss. Remaining RTs were residualized in the same way as in Experiment 1 (Section 5.4). Region-by-region reading time results for the second sentences in critical vignettes during the test phase are shown in Fig. 5a, with breakouts of the CLUSTERING \times CONTINUATION in Fig. 5b and c as was done in Fig. 3 for Experiment 1. Likewise, Fig. 6 shows reading times for SC continuations with and without the complementizer *that*, as was done in Fig. 4 for Experiment 1. Linear mixed-effects models analogous to those in Experiment 1 were fit to the reading time data.

The fundamental prediction for Experiment 2 is the same as for Experiment 1: participants who experienced clustered training should exhibit, at some point in the final three regions of the second critical sentence of testing vignettes, facilitated processing for repeated SCs, and this effect should be apparent only at longer training lengths. However, several incidental differences between the design of the two experiments could orthogonally affect the results. First, the additional sentences in the Experiment 2 vignettes result in a different overall discourse event structure, which could affect where in the sentence the clustering learning effect influences processing (discussed in detail in Section 6.5). Recall, however, that as the expectation-based theory does not predict the specific locus of processing difficulty within the later regions of the sentence (Section 5.4.2), this issue is orthogonal to the critical predictions. Second, effects of training length may manifest themselves in different ways in this experiment, since (i) longer vignettes may produce fatigue earlier in the experiment; (ii) higher proportions of Other:Other sentences could lead to more gradual learning of clustering structure, and (iii) only three, rather than four, training lengths were tested. While all of these caveats should be borne in mind, they are orthogonal to our key prediction of facilitated processing of repetition following sufficient length of clustered experience.

6.4.2. Test sentence processing by training condition and continuation

As in Experiment 1 (Section 5.4), we again first checked for effects of clustering experience on processing of the second (SC or DO) sentence in test vignettes by fitting models to both (i) all three final regions (Fig. 5b) and (ii) the two penultimate regions, such that Conclusion was excluded (Fig. 5c), but again making no

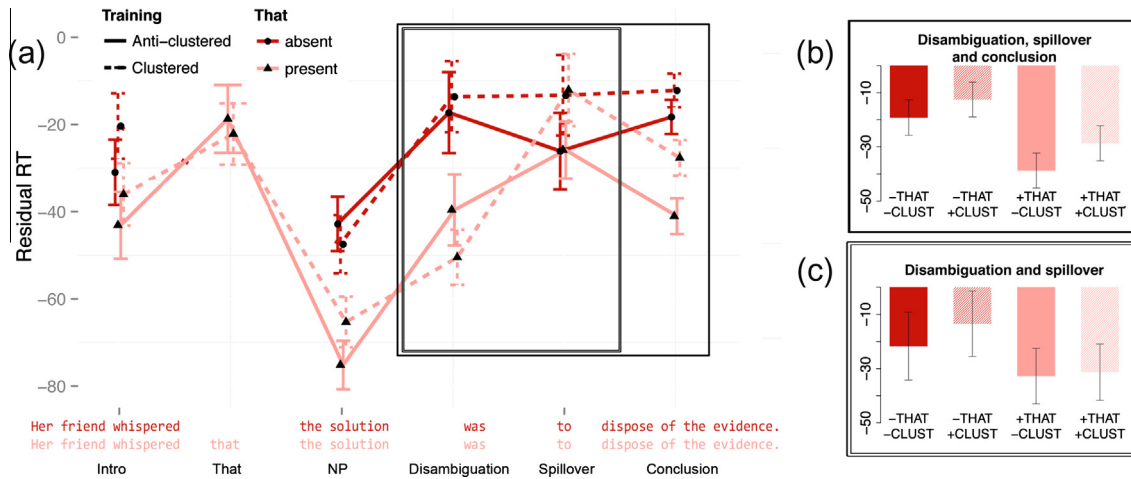


Fig. 6. (a) Region-by-region residual reading times on the second sentence of critical test vignettes of structure SC.SC in Experiment 2, conditioned on training condition and presence of disambiguating complementizer *that*. (b) Residual reading times aggregated over Disambiguation, Spillover, and Conclusion regions (all training lengths). (c) Residual reading times aggregated over Disambiguation and Spillover regions (all training lengths).

prediction for the specific regions where a clustering learning effect would manifest itself.⁹ The qualitative pattern here, especially when considering the Disambiguation and Spillover analysis region, is similar to that of Experiment 1: there is a numeric interactive pattern such that reading times for SC continuations are slow in anti-clustered training, and fast in clustered training, and this pattern grows as training length increases.¹⁰ In statistical analysis, fixed effects predictors were again centered log-transformed training LENGTH, training CLUSTERING experience (clustered or anti-clustered), and CONTINUATION (SC or DO). Random effects included subject- and item-specific intercepts, by-subject random slopes for CONTINUATION, and by-item random slopes for LENGTH, CLUSTERING, and their interaction. We first consider the model including all three final regions (Table 6). No effects were significant in this model.

We next fit a similar model but, as before in Experiment 1, excluded the Conclusion region (Fig. 5c). Results are presented in Table 7. In this model, we observe a significant three-way interaction between CLUSTERING, CONTINUATION, and LENGTH ($\beta = -3.76$, $t = -2.10$): the more extensive a participant's experience in which SC continuations cluster, the more rapidly repeated SCs are subsequently processed. No other effects were significant in this model. As in Experiment 1, we also fit an analogous model excluding items with *that*-present SCs. The critical CLUSTERING:CONTINUATION:LENGTH interaction emerged only as a numeric trend ($\beta = -3.55$, $t = -1.70$). However, given the loss of power resulting from excluding data, and independent evidence that the effect in Experiment 2 is in general smaller and more difficult to detect than the one in Experiment 1, we do not believe there is strong evidence that the critical effect is being driven by some interaction of *that*-presence and clustering.

⁹ As in Experiment 1, we also fit analogous models to individual regions, but in this experiment, the CLUSTERING:CONTINUATION interaction did not reach significance in these models except a model fit to the Disambiguation region of only the longest (15 vignette-pair) training condition ($\beta = -18.00$, $t = -2.3$). Other effects in individual regions were not significant.

¹⁰ Although SCs numerically appear to be slower following clustered training in the analysis region comprising all three final regions (Fig. 5b), these trends are not statistically significant, and are a side effect of the theoretically uninteresting main effect of clustering that is observed in Experiment 1 but not 2. A main effect of clustering does not affect our critical interaction result, and indeed we have no predictions for or against it: such an effect reflects the overall kinds of text readers are encountering, and since these are different in clustered and anti-clustered conditions, this result does not have a meaningful interpretation.

Table 6

Result summary for model of reading times on final three regions of second sentence of critical test vignettes in Experiment 2: coefficient estimates β , standard errors SE (β), and t values.

Predictor	Coef. β	SE(β)	t
Intercept	-27.01	6.24	-4.34
LENGTH	0.64	0.54	1.19
CONTINUATION [$DO = -0.5$, $SC = 0.5$]	-5.31	10.73	-0.49
CLUSTERING [$anti-clustering = -0.5$, $clustering = 0.5$]	13.34	8.85	1.51
LENGTH: CONTINUATION	-0.10	0.80	-0.12
LENGTH: CLUSTERING	-0.24	1.00	-0.24
CLUSTERING: CONTINUATION	1.37	12.28	0.11
LENGTH: CLUSTERING: CONTINUATION	-0.84	1.36	-0.62

Table 7

Result summary for model of reading times on Disambiguation and Spillover regions of second sentence of critical test vignettes in Experiment 2: coefficient estimates β , standard errors SE(β), and t values.

Predictor	Coef. β	SE(β)	t
Intercept	-29.40	6.85	-4.29
LENGTH	0.27	0.63	0.43
CONTINUATION [$DO = -0.5$, $SC = 0.5$]	0.35	12.14	0.03
CLUSTERING [$anti-clustering = -0.5$, $clustering = 0.5$]	10.53	10.44	1.01
LENGTH: CONTINUATION	1.03	1.03	1.00
LENGTH: CLUSTERING	-0.03	1.16	-0.02
CLUSTERING: CONTINUATION	19.33	16.54	1.17
LENGTH: CLUSTERING: CONTINUATION	-3.76	1.80	-2.10

6.4.3. Utility of complementizer *that* as a disambiguation cue

We again also evaluated the effect of the complementizer *that* on SC processing, predicting that participants in the clustering condition should have less need to rely on this cue for disambiguation, since their expectations for SCs should already be higher than those of participants in the anti-clustered condition. Region-by-region reading time results for SC continuations of second sentences in critical vignettes during the test phase, conditioned on the presence or absence of *that*, are shown in Fig. 6a. A linear mixed-effects model analogous to the one in Experiment 1 (Section 5.4.4) was fit; results are presented in Table 8.

As in Experiment 1, we again observe a main effect of *that* presence, so that reading times on the final disambiguation region are higher if *that* is absent ($\beta = 37.52$, $t = 2.96$). The predicted

Table 8

Result summary for model of reading times on disambiguation region of second sentence of critical test vignettes of structure SC.SC in Experiment 2: coefficient estimates β , standard errors $SE(\beta)$, and t values.

Predictor	Coef. β	$SE(\beta)$	t
Intercept	−34.87	10.90	−3.20
LENGTH	0.80	0.80	1.00
THAT [<i>present</i> = −0.5, <i>absent</i> = 0.5]	37.52	12.69	2.96
CLUSTERING [<i>anti-clustering</i> = −0.5, <i>clustering</i> = 0.5]	10.03	14.02	0.72
LENGTH: THAT	−1.27	1.36	−0.93
LENGTH: CLUSTERING	−1.73	1.56	−1.11
CLUSTERING: THAT	12.46	24.15	0.52
LENGTH: CLUSTERING: THAT	0.41	2.65	0.16

Table 9

Explanation of variables in graphical model.

Variable	Explanation
V_1	SC–DO verb in Sent. 1?
C_1	Continuation \in {SC, DO, Oth} of Sent. 1
V_2	SC–DO verb in Sent. 2?
C_2	Continuation \in {SC, DO, Oth} of Sent. 2
V_3	SC–DO verb in Sent. 3? (Expt. 2)
C_3	Continuation \in {SC, DO, Oth} of Sent. 3 (Expt. 2)
ρ_0	Base prob. of SC–DO verb
ρ_m	Prob. of SC–DO verb, given $C_1 = m$
θ_0	Base prob. of SC, given curr. $V = \text{SC–DO}$
θ_m	Prob. of SC, given $V_2 = \text{SC–DO}$, $C_1 = m$
α	Mixing parameter
m	Previous sentence continuation \in {SC, DO, Oth, None}
n	Vignette number

interaction of *that* presence and clustering training experience was not significant ($\beta = 12.46$, $t = 0.52$) and did not interact with LENGTH ($\beta = 0.41$, $t = 0.16$), again likely due to low expectations for DO continuations as a rational response to the statistics of the training environment, in which no DOs occurred. Consistent with the results of Experiment 1, no other effects were significant.

6.5. Discussion

Experiment 2 was designed to replicate Experiment 1 in the context of longer vignettes without the ‘perceptual bookending’ conferred by the pairs of sentences comprising the stimuli for that experiment. Even in the context of 3-, 4-, and 5-sentence vignettes in this second experiment, we observe that comprehenders adapt their expectations for structural repetition to the syntactic clustering properties of the environment. Participants who were trained in a *clustering* environment, in which SC continuations always occurred in clusters of two, again came to more strongly expect an additional SC immediately following an initial occurrence than participants who were trained in an *anti-clustering* environment in which SC sentences never occurred before or after other SC sentences. As in Experiment 1, this learning effect is most apparent at longer training lengths (Fig. 9, Section 5.4.5).

The effect manifests itself in slightly different ways in this second experiment. Some differences are likely explicable in theoretically uninteresting ways. First, in Experiment 1 but not Experiment 2, clustering leads to faster processing overall, although because the texts are different across clustering conditions and experiments, this result does not have a meaningful interpretation. Second, while in both experiments the clustering learning effect is clearly evident at medium and long, but not short, training lengths, training length has a significant linear interaction with the clustering learning effect in Experiment 2 but not Experiment 1. However, this may result from the fact that Experiment 2

did not have a 30-vignette training condition: in Experiment 1, this longest condition resulted in a slight decrease in observed learning effects, possibly due to fatigue, resulting in overall nonlinear effects of training length; see Section 5.4.5.

Other differences between the results of the two experiments may be of more theoretical interest. While our theory does not predict or depend on the exact location of clustering-based facilitation within the sentence (Section 5.4.2), we observe that the effect extends to the sentence-final Conclusion region in Experiment 1, but not Experiment 2. Recall that in Experiment 1 but not Experiment 2, this region is always the final region in the vignette. We speculate that this may indicate interaction between (clustering-derived) syntactic expectations and discourse wrap-up. In particular, learning syntactic clustering properties may give comprehenders a strong cue to overall discourse event structure, thus facilitating wrap-up on encountering the expected syntactic evidence in Experiment 1. In contrast, in Experiment 2, the Conclusion region may not enjoy this advantage because there is often at least one more sentence remaining, meaning that comprehenders may be entertaining more complex, less certain hypotheses about both syntactic recurrence patterns and overall event structure. We return to the relationship between syntactic clustering and discourse processing in Section 8.

Perhaps the strongest and most characteristic difference between the results of the two experiments is that the clustering learning effect is weaker in Experiment 2—both numerically and in the qualitative ways already mentioned. One potential explanation for this difference is that the substantially higher proportion of sentences with Other structures leads to relatively fewer opportunities to collect evidence of SC clustering, plausibly leading to more gradual learning of clustering structure. In the next section of this paper, we develop a Bayesian belief-updating model investigating this difference in learning properties, and showing that the optimal degree of sensitivity to these between-sentence contingencies is indeed learnable through experience.

7. A Bayesian belief-update model of adaptation to clustering in the environment

In this section, we model comprehender learning of the syntactic clustering properties of the environment through Bayesian belief update, and show how expectation for repetition of structures—the classic priming effect—can result from this kind of rational environmental adaptation. More specifically, the model serves as a proof-of-concept that the relative importance of clustering properties in the environment, as well as the particular shape of these clustering properties, are learnable through experience, and that the resulting syntactic expectations qualitatively mirror human behavioral data. While our theoretical claims in general, and model in particular, are cast at Marr’s (1982) computational level of analysis, related implementations exist at the algorithmic and physical levels of analysis; see Yu and Cohen (2008) for details.

We model syntactic comprehension as a problem of Bayesian inference, wherein comprehenders combine their prior linguistic experience with the input in the current linguistic environment in order to infer posterior probability distributions over syntactic structures as they read sentences. In particular, they must infer the probability of an SC continuation given an ambiguous SC–DO verb and the syntactic structure of the previous sentence. The high-level structure of the model is straightforward: comprehenders track both the general probability of SCs and the clustering properties of SCs in their experience, and learn the optimal mixture of these information sources (which could theoretically be

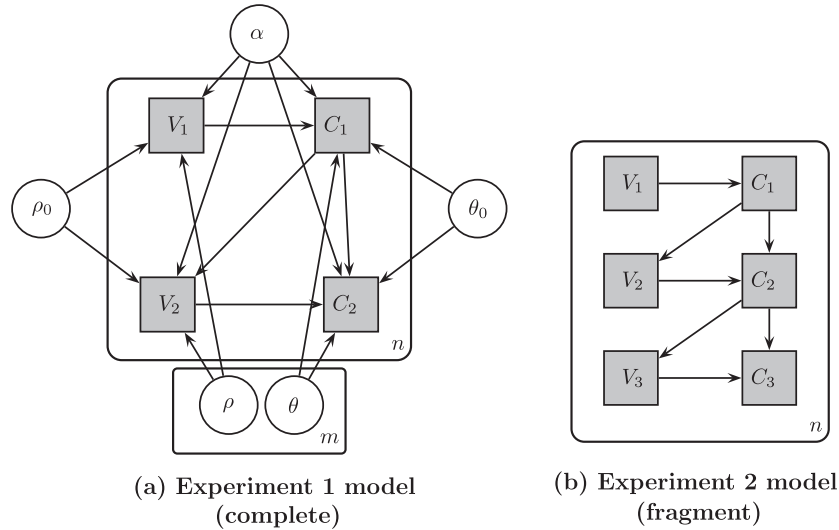


Fig. 7. Graphical representation of the Bayesian belief-update model (variable explanations in Table 9; full specification in Section 7.1) (a) Complete model fit to Experiment 1 data. (b) Fragment of analogous model for Experiment 2: n plate in the case of a three-sentence vignette. Nodes are added analogously in the cases of four- and five-sentence vignettes. Following the structure of the model in (a), the following dependences always hold: ρ_0 and ρ_m to all V ; θ_0 and θ_m to all C ; and α to all V and C .

categorical reliance on a single one of these data sources) in inferring the posterior probability of an SC.

This type of *additive mixture model* is a natural choice given our theoretical claim that comprehenders learn both base rates of syntactic structures (evidenced in previous findings in expectation adaptation) and local context effects (evidenced in the current experiments) (Section 3.1). This two-level model also provides a good fit to classic, non-linguistic sequential learning: a similar mixture model developed in Wilder, Jones, and Mozer (2009), building on Yu and Cohen (2008), shows that tracking both the overall frequency and the repetition pattern of prior occurrences yields the best fit to human behavior in the sequential forced-choice tasks of Soetens et al. (1985) and Cho et al. (2002). Additive mixture models also hold a deep theoretical relevance to language, since they discover the optimal levels of granularity at which to make generalizations (in the case of our experimental environments, unigrams and bigrams of syntactic structures), a critical and ubiquitous task in natural language processing and learning.

7.1. Model specification

The model is presented in graphical form in Fig. 7. We begin with the case of Experiment 1, where stimuli are two-sentence vignettes, where each sentence can contain either an SC structure, DO structure, or Other unrelated structure. Let there be n vignettes in a training phase. The model iterates through each of these n vignettes, combining prior expectations for the probability of each of these three structures with the actual outcome of the current vignette in order to infer a posterior distribution over the structures. Initially, the prior consists of only previous linguistic experience, but with each subsequent iteration, the posterior distribution inferred in the most recent vignette serves as the prior for the current vignette. The final posterior distribution at the end of training is assumed to inform comprehender expectations during the test phase.

Each sentence consists of a verb V , which may either be an SC–DO verb or an Other verb subcategorizing for an unrelated structure, and a continuation C , which may be one of SC, DO, or Other. In this model, each verb and continuation may in principle be influenced by the continuation of the previous sentence. Thus we define

a set m of possible continuations for the previous sentence, consisting of SC, DO, Other, and None (for the case where the current sentence is the first sentence in the vignette, so that there is no previous sentence continuation).

We begin by modeling the outcome of the main verb V_1 in the first sentence of the n th vignette. This outcome can either be an SC–DO subcategorizing verb, or not. If it is not, it is clear at this point that the sentence continues with an Other structure. If it is, the sentence remains ambiguous between SC and DO continuations until final disambiguation. Since the main verb can turn out either to subcategorize for SC–DO continuations or not, let V_1 be Bernoulli distributed with success parameter (SC–DO verb outcome) defined as a mixture between the general unigram probability ρ_0 of SC–DO verbs and the probability $\rho_{m=none}$ of an SC–DO verb given that there is no previous sentence:

$$V_1 \sim \text{Bern}(\alpha \cdot \rho_{m=none} + (1 - \alpha) \cdot \rho_0) \quad (6)$$

where α is a mixing parameter ranging between $[0, 1]$ such that high α implies greater dependence on the preceding context (that is, sensitivity to clustering), and low α implies greater dependence on the general probability (base rate) of SC–DO verbs (low sensitivity to clustering). The model learns α from the data. We assume that comprehenders initially have no particularly strong or weak prior expectation for the value of α . Therefore we place a uniform Beta-distributed prior on α , in which each value is equally likely:

$$\alpha \sim \text{Beta}(0.5, 2) \quad (7)$$

For ρ_0 , we assume the comprehender's prior unigram expectation for an SC–DO verb mirrors the general frequency of SC–DO verbs in natural language. Thus we place a Beta prior on ρ_0 with mean μ_{ρ_0} and sum of shape parameters ν :

$$\rho_0 \sim \text{Beta}(\mu_{\rho_0}, \nu) \quad (8)$$

We estimate μ_{ρ_0} as 0.127, the relative frequency of sentences containing SC–DO verbs in the Brown Corpus.¹¹ Since the concentration of this value—the strength of the belief that the probability of an

¹¹ In particular, we count the summed occurrences of the 84 SC–DO verbs listed in Appendices B.1.1 and B.2.1 of this paper in Roland et al.'s (2007) patterns `sc1-sc4` for SCs and patterns `st1-st5`, `tr1-tr5`, and `dt` for DOs.

SC–DO verb in natural language is equal to μ_{ρ_0} —is not recoverable from the corpus, the value of ν was selected to optimize model fit to behavioral data; see Section 7.2 for details on model fitting. We likewise place a Beta prior on $\rho_{m=none}$, the probability of an SC–DO verb in a vignette-initial sentence, with mean $\mu_{\rho_{m=none}}$ and sum of shape parameters ν , estimating $\mu_{\rho_{m=none}}$ as 0.142, the relative frequency of an SC–DO verb in a document-initial sentence in the Brown corpus.

Next, given the outcome of the main verb in the first sentence of the vignette, we model its continuation C_1 , which can be SC, DO, or Other. If the verb V_1 is non-SC–DO, the continuation C_1 must be Other. However, if V_1 is an SC–DO verb, C_1 may continue either as an SC or as a DO. Thus in this case C_1 is Bernoulli distributed with success parameter defined as a mixture between the general unigram probability θ_0 of SC continuations given an SC–DO verb and the probability $\theta_{m=none}$ of an SC continuation given a SC–DO verb and no previous sentence. We multiply this quantity by the outcome of the verb V_1 (where SC–DO = 1 and Other = 0):

$$C_1 \sim \text{Bern}(V_1 \cdot (\alpha \cdot \theta_{m=none} + (1 - \alpha) \cdot \theta_0)) \quad (9)$$

where μ_{θ_0} is estimated from the Brown Corpus as 0.416 and $\mu_{\theta_{m=none}}$ as 0.407, which are, respectively, the unigram probability of an SC continuation and the probability of an SC continuation in a document-initial sentence, given one of the verbs used in this study (see Appendix A).

The main verb and continuation of the second sentence of the vignette are modeled similarly, but m takes on distinct values, since the continuation of the previous sentence may be SC, DO, or Other. The main verb of the second sentence V_2 is Bernoulli distributed with success parameter defined as a mixture between general probability of SC–DO verbs ρ_0 and the probability ρ_m of an SC–DO verb given the outcome of C_1 in the previous sentence. The ρ_m have Beta priors:

$$\rho_m \sim \text{Beta}(\mu_{\rho_m}, \nu) \quad (10)$$

estimated from Brown as $\mu_{\rho_{SC}} = 0.188$, $\mu_{\rho_{DO}} = 0.123$, and $\mu_{\rho_{Other}} = 0.122$. Thus V_2 is distributed as:

$$V_2 \sim \text{Bern}(\alpha \cdot \rho_m + (1 - \alpha) \cdot \rho_0) \quad (11)$$

Finally, the continuation C_2 of the second sentence is similarly Bernoulli distributed according to a mixture of the general probability of SC outcomes given SC–DO verbs, and the probability of SC outcomes given SC–DO verbs and the continuation of the preceding sentence:

$$C_2 \sim \text{Bern}(V_2 \cdot (\alpha \cdot \theta_m + (1 - \alpha) \cdot \theta_0)) \quad (12)$$

where the Beta priors for θ_m are estimated from Brown as $\mu_{\theta_{SC}} = 0.52$, $\mu_{\theta_{DO}} = 0.410$, and $\mu_{\theta_{Other}} = 0.410$.

In order to adapt this model for the three- to five-sentence vignettes of Experiment 2, we add nodes according to the structure in Fig. 7(b), so that plate n contains the same number of iterations of V and C as there are sentences in the vignette. Note that, for simplicity, this expanded model retains the bigram dependency structure of the original model: in the case of a three-sentence vignette, for example, *direct* dependencies hold from Sentence 1 to 2 and from 2 to 3, but not from Sentence 1 to 3.

7.2. Model implementation and fitting

We implemented the belief-update model in JAGS (Plummer, 2003), using Markov Chain Monte Carlo simulation to infer posterior distributions over the probability of SC continuations in second sentences of test vignettes, given clustered or anti-clustered training at various training lengths (see Section 6). For each version of

the model, we specified a burn-in period of 1000 iterations followed by 10,000 iterations. In order to avoid potential autocorrelation, we thinned the Markov chain by taking one of every 20 samples after burn-in.

In order to compare our model's output (comprehender expectations in the form of probability distributions) with the behavioral data collected in each of the two reading time experiments, we employ SURPRISAL as our linking hypothesis (Hale, 2001; Levy, 2008). Processing difficulty (operationalized as reading time) of a structure T is correlated linearly with its surprisal (Levy, 2008), which is defined as its negative log-probability:

$$\text{difficulty} \propto -\log P(T) \quad (13)$$

For each experiment, in order to fit the model's one free parameter ν , the concentration of all Beta distributions except (7), for each training length we regressed the differences between RT speedups for SCs versus DOs plotted in Fig. 9(a) against the differences between surprisal differentials for SCs versus DOs plotted in Fig. 9(b). We then searched the parameter space between 1 and 50 using grid search at 0.1 intervals for the ν that yielded the lowest sum of squared residuals (and, therefore, the maximum likelihood) in this regression. Using this methodology, ν was set at 1.4 for Experiment 1 and 3.2 for Experiment 2.¹²

7.3. Model predictions

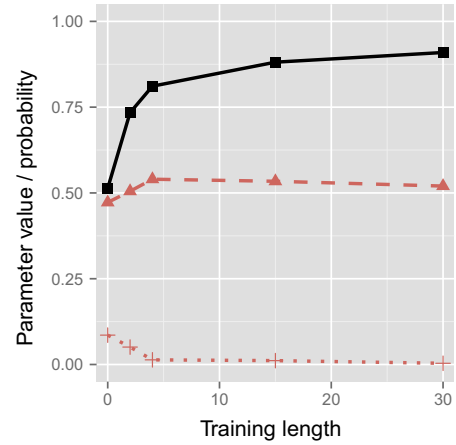
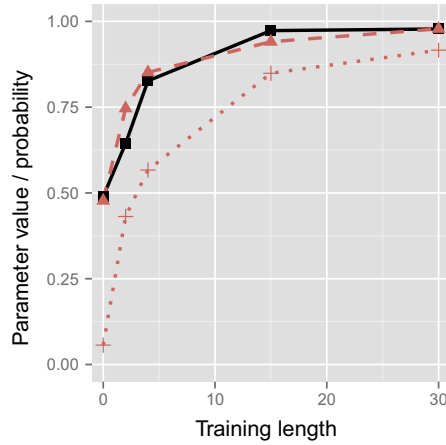
Fig. 8 displays the model predictions for the probability of SC continuations in each sentence of test vignettes, as well as the learned value of α on which this probability is based. Point estimates for each model parameter were obtained by averaging the model estimates from the thinned Markov chain. We first consider its predictions for Experiment 1. Under clustered training, the model learns an increasingly high value for α , implying increasing dependence on the clustering structure of SC sentences rather than the general base rate of SCs, whereas given anti-clustered training, α increases more gradually. This is unsurprising due to filler trials (which consist of Other.Other structures), which result in both Other.SC and Other.Other vignettes in anti-clustered training. Since either an SC or Other structure may thus follow Other first sentences, clustering is not as reliable under anti-clustered training as it is under clustered training, where vignettes are exclusively of the form SC.SC or Other.Other.

Crucially, the model learns to increasingly strongly expect an SC following an SC (that is, in second sentences of test vignettes) given longer amounts of clustered training, but not given anti-clustered training. The initial increase once the first training data are seen is due to learning a high base rate of SC continuations, since no DOs occur in training. In Experiment 1, then, the model learns a behavior expected under the hypothesis that comprehenders track clustering patterns of syntactic structures and use these to inform their syntactic expectations.

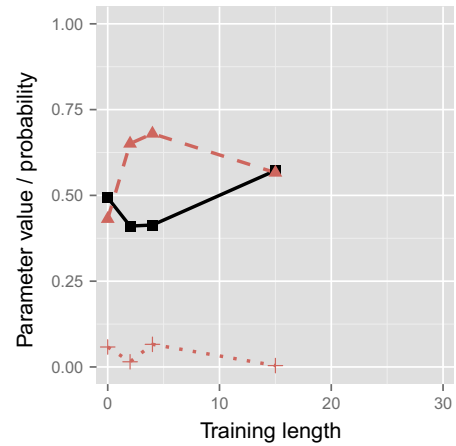
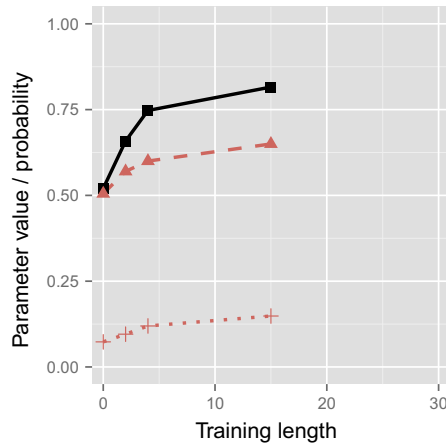
In Experiment 2, the model's predictions under clustering are similar to those in Experiment 1, albeit, unsurprisingly, weaker. Under anti-clustering, evidence for any kind of (anti-)clustering structure is so sparse that the model initially learns a decreasing value of α , but ultimately weakly learns that some clustering structure exists. In Experiment 2, then, the model does learn to expect repeated SCs more strongly under clustering than anti-clustering, but this differential is much smaller and learned more slowly than in Experiment 1. This is not surprising, since the stimuli in Experiment 2 provides relatively fewer opportunities to observe clustering structure. The substantially higher ratio of Other sentences

¹² We assume, following e.g. Yu and Cohen (2008), that participants assume a non-stationary environment, so that ν can be relatively small, reflecting heavier weighting of recent experience (i.e., experience within the experiment).

Expt. 1



Expt. 2



Clustered training

Anti-clustered training

Learned value of α
 $P(C_2=SC|V_2=SCDO, C_1=SC)$
 $P(C_2=SC|C_1=SC)$

Fig. 8. Model predictions for test vignettes.

introduces a form of noise and thereby weakens learning, the effect of which is already evident in the anti-clustered condition in Experiment 1 but exacerbated in both conditions in Experiment 2. In sum, then, this modeling result is consistent with an explanation of Experiment 2's overall weaker behavioral effects as resulting from learning differences associated with a higher proportion of Other structures, as we suggested in Section 6.5.

In order to more formally compare model predictions to behavioral results, for each experiment we transform the model output to surprisal values and plot these alongside reading time data in Fig. 9. These values are based on conditional probability of an SC continuation in the second sentence of testing vignettes, given an SC continuation in the first sentence and an SC–DO verb in the second sentence. As length of training increases, the model predicts an increasingly strong processing advantage in clustered training for SC continuations, and an increasingly strong disadvantage for DO continuations, qualitatively matching behavioral data. The best-fit values for ν in each model, 1.4 for Experiment 1 and 3.2 in Experiment 2, reflect less flexibility and a smaller learning effect in Experiment 2, an expected result given the overall less robust behavioral results Experiment 2. We limit the evaluation of the model's predictions to this qualitative discussion, since the model is intended simply as a demonstration that the optimal degree of sensitivity to clustering properties is learnable through experience (but dependent on the degree of noise therein).

7.4. Discussion

In this section, we presented a Bayesian belief-update model demonstrating how facilitated processing of repeated structure may result from rational adaptation to the environment. In particular, the model successfully learns that clustering properties indeed predict upcoming syntactic structures, and to weight this cue accordingly in developing syntactic expectations. The model predicts that clustering properties are weighted more heavily with increasing experience, implying adaptation to the statistics of the current environment, and that a processing advantage for repeated structures (given clustered training) similarly grows over time with increased experience. This prediction qualitatively matches behavioral data, in which comprehenders indeed show this processing advantage most clearly at longer training lengths. (The model does not predict the lack of additional learning at the longest training length, which may be influenced by participant fatigue and is consistent with previously observed non-linearity of learning, as we discuss in Section 5.4.5.) The model further predicts behavioral data by revealing a weakened learning effect given the noisier stimuli of Experiment 2, in which the far greater ratio of filler sentences appears to constrain learning. The ability to learn the importance of clustering properties from experience lends plausibility to our account of repetition priming as a result of rational adaptation to the environment.

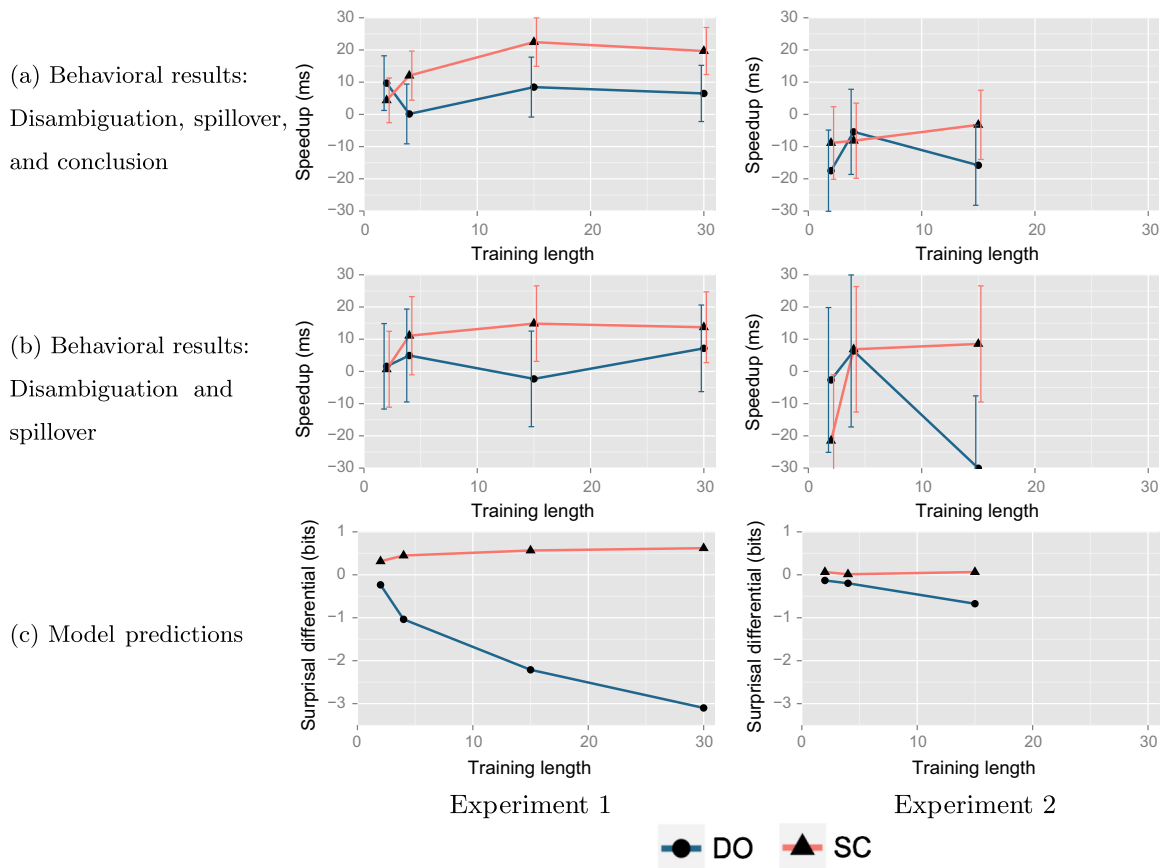


Fig. 9. Processing advantage of clustered training (versus anti-clustered training) for final three regions of second sentence of test vignettes, given an SC continuation in the first sentence and an SC–DO verb in the second sentence. (a and b) Reading times at test, given anti-clustered experience, minus reading times at test, given clustered experience. Error bars reflect 95% confidence intervals of by-subject means. (c) Model-predicted surprisal at test, given anti-clustered experience, minus surprisal at test, given clustered experience. See Sections 7.2 and 7.3 for model prediction details.

8. General discussion

In this paper, we developed a computational-level account of structural priming in comprehension, proposing that it is rational to expect repetitions on the basis of the temporal clustering properties of syntactic structures in natural language. Sensitivity to this higher-order statistical property of language, we argue, is an optimal behavior given the challenge of efficient online language processing, in which comprehenders must continually update expectations at multiple levels of structure while negotiating considerable environmental variability. In this section, we discuss the implications of a rational approach to comprehension priming, including its particular relevance for language processing.

8.1. Repetition & rationality

What is gained through a rational approach to repetition priming? At a high level, this approach leads to a unified functional motivation for observed behaviors and cognitive architectures, in which these comprise an optimal means of achieving an agent's goals given its environment. *A priori*, there may be little reason to expect, for example, an architecture in which residual activation of a representation is briefly maintained following the occurrence of a stimulus: perhaps this behavior is in fact more costly than immediately discarding the representation. Yet given an environment in which stimuli are likely to recur at short intervals, an architecture that keeps representations briefly active is rational (Anderson, 1990; Anderson & Schooler, 1991). Conversely, this

architecture would be extremely maladaptive in a world in which stimuli never clustered in time. While a feedback loop in which clustering linguistic structure in the world, and a processing strategy geared for clustering, mutually reinforce one another seems plausible, independent factors such as topicality and discourse-functional goals also appear to contribute to clustering structure in the world (e.g. DuBois, 2010; Healey, Purver, & Howes, 2014), motivating an explanation in which the structure of the world *shapes* comprehension strategies. As we discuss below, rational expectation for repetition is extraordinarily well-suited to the temporal structure of natural language, and an appeal to this account may allow for substantial simplification of theories of particular aspects of language processing.

8.2. Language & rational expectation for repetition

Language is a natural and uniquely significant testing ground for rational accounts of repetition priming. While comprehenders' ability to develop rational expectations based on higher-order sequential contingencies is relatively well-established in explicit sequential prediction tasks (Section 3.3.2), the categories of stimulus used in these studies—typically, semantically devoid sequences of letters or colors—do not have any inherent clustering property, and it is thus uncertain whether the strategies comprehenders deploy in these tasks generalize to more naturalistic, ecologically valid sequence processing. Language, in contrast, is an organic and inherent system of sequential dependencies tending naturally toward clustering of repeated elements, and thus allows for a

qualitatively novel generalization of a rational account of expectation for repetition.

Indeed, the rational approach leads to particular advantages within theories of language processing itself. For instance, the so-called *lexical boost effect*, in which syntactic priming is briefly strengthened in the presence of lexical overlap between prime and target sentences, has led to the proposal of ‘dual-mechanism’ accounts of priming, in which short-term, lexically-influenced structural priming operates through residual activation, while longer-term structural priming operates through implicit learning (Hartsuiker et al., 2008; Reitter et al., 2011). A computational-level appeal to sensitivity to environmental clustering properties may reveal a common utility for these contrasting mechanisms, or even motivate a single-mechanism hypothesis: since lexical content clusters in time (Section 3.2.1), the relatively fast decay of the lexical boost effect may simply be an optimal response to lexical clustering in the environment (Jaeger & Snider, 2013:74). Generalizing this logic, temporal decay of syntactic priming strength more generally may find explanation in the clustering structure of the environment: the longer the interval since the most recent occurrence of a structure, the less likely that structure is to be needed again, and the less strongly it should be activated (in the spirit of Anderson & Schooler’s (1991) *need probability* hypothesis; Section 3.3.2) (Jaeger & Snider, 2013:74). Our results are encouraging for both of these possibilities.

8.3. Context, flexibility, & rationality

Our results expand the empirical scope of *rapid expectation adaptation*, in which comprehenders rapidly and rationally adapt their linguistic expectations to converge on the statistics of the current environment (Section 2.2). We have shown that such adaptation is sensitive to higher-order temporal contingencies between sentences, supporting a role for multiple levels of linguistic context in rapid implicit learning throughout discourses. Multilevel context is consistent with evidence from other domains, in which sensitivity to both the base rate of a stimulus as well as higher-order sequential patterning are jointly necessary to account for human behavior (Thomson et al., 2012; Wilder et al., 2009).

The rational behavior we have described involves flexibly adjusting expectations for repetition as a function of the structure of the current environment. However, such flexibility may not be available in all scenarios. For tasks with very rapid timescales, for instance, it may not be feasible to learn and deploy knowledge of the clustering properties of the current environment in real time. Given this inflexibility, a nevertheless rational default behavior may be to prepare for immediate repetitions, since clustering is a ubiquitous temporal dynamic of phenomena in the world (Section 1). Hypothesized mechanisms such as residual activation are highly compatible with this behavior, and this behavior is indeed observed in these situations: at extremely short response-stimulus intervals at early stages of their experiment, Soetens et al. (1985) and Cho et al. (2002) find that participants respond more quickly to immediately repeated stimuli even if their occurrence violates an established pattern, but at RSIs over 100 ms, they respond more quickly to stimuli that fit the pattern, regardless of whether they are repetitions of the immediately preceding trial. Similarly, Traxler et al. (2000) find that early eye fixation measures on repeated lexical items in sentences are fast, but later reading time measures are slow if the repetition is not naturally licensed by the discourse. These behaviors, rational within, perhaps, the physical constraints of the agent, may indicate a role for *bounded rationality* in comprehension priming (Simon, 1955).

While we crucially observe environmental flexibility with regard to expectations for clustering of syntactic structures, the extent of such flexibility remains an open question. Specifically,

in naturalistic language processing, do comprehenders adapt equally well to *a priori* implausible distributions—that is, distributions that differ sharply from prior expectations? Such an ability would underscore the power of expectation adaptation as a means to efficiently process a wide variety probable and improbable language input, a constant demand of naturalistic comprehension. In this vein, for example, Farmer et al. (2011) show that comprehenders learn experiment-specific verb subcategorization biases, despite apparent absence of an obvious, plausible motivation for such a correlation. In the domain of syntactic clustering, specific questions remain as to (1) whether plausibility of clustering—perhaps informed by factors such as discourse coherence—affects learning of the clustering distribution itself, and (2) whether an even more naturalistic (and thus plausible) clustering distribution, in which clusters do not occur exactly as pairs as they do in the present experiments, would lead to differences in learning. For example, it seems plausible that syntactic clusters may comprise three or four sentences, possibly interrupted by a small number of other sentences (our corpus study in Section 4 presents suggestive evidence), and learning of a distribution of such clusters may be affected by competing pressures such as, one hand, its increased complexity, and on the other hand, its potentially greater prevalence in natural language.

Investigating the processing of syntactic clusters beyond simple pair structures will also provide critical additional evidence for how and why comprehenders learn syntactic recurrence distributions. An alternate explanation for the evidence adduced in this study is that participants simply learned that SCs occur in pairs. Learning to specifically expect pairs is distinct from learning generally higher probabilities of stimuli in the aftermath of one or more occurrences, and it remains an open question to what extent either or both of these processes are at play both in our experimental results and in comprehension priming in naturalistic sentence processing. Indeed, Kim, Mauner, and Koenig (2007) argue for dissociation between learned pairwise expectations and general facilitation of repetitions, on the basis of experimental results in which DOs facilitated processing of following SCs when there was a clear pairwise relationship signaled by a repeated verb. Indeed, pairwise learning provides a plausible alternative explanation for our overall weaker learning effects in Experiment 2, where pair structure was less obvious, and future work should explicitly investigate this hypothesis. Even under this scenario, our experiments demonstrate for the first time that comprehenders can be sensitive to higher-order statistics of syntactic recurrence distributions, a crucial prerequisite for the possibility that they leverage this learning to efficiently process the widely-observed naturalistic clustering of linguistic forms.

9. Conclusion

In this paper, we proposed that facilitated processing for repeated syntactic structure is a result of rational expectation for repetition, given that tokens of particular syntactic structures cluster together in natural language. We showed that comprehenders rapidly adapt their expectations for clustering of syntactic structures to the statistics of the current environment, and that formulating these expectations on the basis of a mixture of both (i) the base rate of occurrence of the structure and (ii) the clustering properties of the structure is necessary to optimize syntactic prediction. Standard structural priming accounts that do not track temporal clustering properties of syntactic structures do not fully account for the behavior observed in the experiments reported here. The current proposal, in contrast, accounts for classic priming effects in comprehension as well as behavior in novel environments by positing that comprehenders track long-range distributions of

syntactic structures in both prior and recent experience, by means of a potentially general mechanism for environmental adaptation.

Acknowledgements

We are grateful to audiences at the 26th Annual CUNY Conference on Human Sentence Processing and the 88th Annual Meeting of the Linguistic Society of America, members of the UCSD Computational Psycholinguistics Lab, Florian Jaeger, Andrew Kehler, and Robert Kluender, and three anonymous reviewers for insightful feedback on this work. We also thank Cody McCormack and Agatha Ventura for assistance with stimulus preparation. Any remaining errors and omissions are our own. This work was supported by an NSF Graduate Fellowship to MM and by NSF grant 0953870, NIH grant R01HD065829, and an Alfred P. Sloan Research Fellowship to RL.

Appendix A. Baseline for within-document clustering

In this appendix, we detail our procedure for computing the expected number of immediate SC repetitions given random permutation of the ordering of sentences with a corpus document (Section 4). We conceptualize the first part of the problem as a ball-and-urn problem, in which the number of ways to drop k balls among n dividers is computed. In our case, the balls are (sequences of) SC sentences, and the dividers are (sequences of) non-SC sentences (so that the number of non-SCs in a document is equal to $n - 1$). Consider a document containing 4 SCs and 3 non-SCs, which contains r unique partitions of SC sentences (let s represent an SC sentence, and $.$ represent one or more non-SC sentences):

Partition	k	
ssss	1	(A.1)
s.sss	2	
s.ss.s	3	
ss.ss	2	
s.s.s.s	4	

For each of these partitions, the number of ways to distribute k SC sequences among n non-SC dividers is given by

$$\binom{n}{k} \quad (\text{A.2})$$

We next compute the number of permutations within each of these partitions. For example, the partition containing sequences s and ss has three permutations, $s.s.s$, $ss.s.s$, and $s.s.s.s$. The number of permutations for a partition is given by

$$\frac{k!}{\prod_i (t_i!)} \quad (\text{A.3})$$

where t_i is the number of tokens of sequence type i (such that in the case of the above partition containing sequences s and ss , the t_i are 1 and 2, corresponding to one token of ss and two tokens of s).

Combining expressions (A.2) and (A.3), the expected value of the number of repetitions of SCs in a document is then equal to

$$\sum_{j=1}^r \left(\binom{n}{k} \frac{k!}{\prod_i (t_i!)} N_j \right) \cdot p_j \quad (\text{A.4})$$

where r is the number of unique partitions, N_j is the number of immediate SC repetitions present in partition j , and p_j is the proportion of all possible permutations reflected by partition j .

Appendix B. Stimulus materials

Bracketed material reflects extended vignettes used in Experiment 2.

B.1. Training phase

B.1.1. Training phase critical items

SC.SC, SC.OTH and OTH.SC items

1. *accept, know*

SC.SC	[The problems on the math test had seemed completely impossible.] Sara accepted her score was the lowest she'd ever received on a test. Her teacher knew the answers were difficult to work out. [She ended up giving lots of extra credit.]
SC	[The problems on the math test had seemed completely impossible.] Sara accepted her score was the lowest she'd ever received on a test. She was disappointed in herself, but eager to do better on the next test. [Besides, her teacher ended up giving lots of extra credit.]
OTH	[Liz's homework was taking hours, and she wanted to finish before soccer practice.] She eventually gave up on her algebra homework, since the problems didn't seem to have any solution. Her teacher knew the answers were difficult to work out. [She always tried to challenge her students.]

2. *motion, warn*

SC.SC	The director motioned the vehicle should be purchased with city funds. Her assistant warned the taxpayers would not approve of her plan. [After all, the citizens would much rather renovate the city park than replace a patrol car.]
SC	The director motioned the vehicle should be purchased with city funds. The other board members vetoed the decision, since the city was in debt. [After all, the citizens would much rather renovate the city park than replace a patrol car.]
OTH	The mayor wanted to upgrade her mansion using tax revenues. Her assistant warned the taxpayers would not approve of her plan. [They hated corruption at city hall.]

3. *expect, doubt*

SC.SC	Mary had expected the light could be explained by the scientists. But her friend doubted it was a natural phenomenon. [Little did they know, Mary's little brother was just pulling a silly prank.]
SC	Mary had expected the light could be explained by the scientists. Ever since childhood, she had always been a very logical person. [Little did she know, her little brother was just pulling a silly prank.]
OTH	Stella and her friend witnessed something unusual hovering in the night sky. But her friend doubted it was a natural phenomenon. [She had recently watched a documentary about UFOs.]

4. *observe, forget*

- sc.sc The chemist observed the mixture was far thicker than expected. She had forgotten the procedure had called for adding water. [Unfortunately, she had to start over and try again.]
- sc. The chemist observed the mixture was far thicker than expected. She wondered if she had used the right concentration of chemicals. [Unfortunately, she had to start over and try again.]
- OTH. The muffins Natalie baked turned out extremely dry and hard. She had forgotten the procedure had called for adding water. [She ended up throwing them out.]
- sc

5. *suggest, promise*

- sc.sc [Aaron was having trouble with his grades.] His teacher suggested a tutor could help him with his physics homework. His parents promised a reward was in store for him if his grades improved.
- sc. [Aaron was having trouble with his grades.] His teacher suggested a tutor could help him with his physics homework. He had been struggling with it since the beginning of the school year.
- OTH. [Tom was having trouble in his literature class, and he was falling behind.] His New Year's resolution was to spend an extra hour studying in the library each day. His parents promised a reward was in store for him if his grades improved.
- sc

6. *judge, fear*

- sc.sc Stephen judged no one could see him at such a distance. His accomplice feared the police would spot him anyway, putting both of them at risk. [The crime they committed put them in danger of arrest.]
- sc. Stephen judged no one could see him at such a distance. He was in the perfect hiding spot for playing hide-and-seek with his brother. [But his brother was peeking the whole time.]
- OTH. Calvin hid his gun and crouched behind a truck in the dark alley. His accomplice feared the police would spot him anyway, putting both of them at risk. [The crime they committed put them in danger of arrest.]
- sc

7. *teach, perceive*

- sc.sc Rodolfo's class was taught evolution has been a controversial issue in politics. He perceived the tension was due to the complexities in politics and science.
- sc. Rodolfo's class was taught evolution has been a controversial issue in politics. His teacher was very fair-minded and tried not to offend anyone.
- OTH. Elliot's parents disagreed about global warming, but tried to keep their arguments from him. He perceived the tension was due to the complexities in politics and science.
- sc

8. *notice, sense*

- sc.sc The politician noticed an error had made its way into her speech. She sensed her mistake could cost her the trust of the voters.
- sc. The politician noticed an error had made its way into her speech. She blushed and apologized immediately after the press conference.
- OTH. The candidate didn't stand up for the national anthem, and the audience looked at her disapprovingly. She sensed her mistake could cost her the trust of the voters.
- sc

9. *maintain, regret*

- sc.sc [Miguel was wondering about the savings account.] His financial advisor maintained the account was not subject to taxation. Miguel regretted his investments were in a different bank without that kind of account.
- sc. [Matthew had some tax questions about his savings.] His financial advisor maintained the account was not subject to taxation. Matthew was not so sure, and went to another firm for a second opinion.
- OTH. [Miguel was wondering about the savings account.] As the bank manager pointed out, the account had a very high interest rate. Miguel regretted his investments were in a different bank without that kind of account.
- sc

10. *charge, suspect*

- sc.sc Christina charged Brent's new laptop was stolen property and should be reported. She suspected his friend was also aware of the situation.
- sc. Christina charged Brent's new laptop was stolen property and should be reported. She threatened to file the report herself if he was unwilling to do it.
- OTH. Mario was hiding beer in his room, and it was very obvious to his concerned mother. She suspected his friend was also aware of the situation.
- sc

11. *coach, overhear*

- sc.sc Frank coached the team should throw a long pass late in the football game. The other team overheard the decision was a desperate last resort.
- sc. Frank coached the team should throw a long pass late in the football game. He had a tendency to favor risky choices, but his team won many games.
- OTH. The marketing team's new plan to win the contest was to give away free samples. The other team overheard the decision was a desperate last resort.
- sc

12. *guess, learn*

- SC.SC The student guessed the answer must require using $E = mc^2$. Eventually she learned the formula had been discovered by Einstein.
- SC The student guessed the answer must require using $E = mc^2$. She just wasn't sure how to use it to solve the problem she was working on.
- OTH. Try as she might, the physics student couldn't solve the atomic energy problem. Eventually she learned the formula had been discovered by Einstein.

13. *see, deny*

- SC.SC Marco saw the burglary had not been captured on video. He denied misconduct had taken place that night.
- SC He saw the burglary had not been captured on video. No one could accuse him of committing the crime without video evidence.
- OTH. No one was around when Tyson snuck into the room full of confidential records. He denied misconduct had taken place that night.

14. *swear, prove*

- SC.SC The cult member swore the oath had not been broken by him. The leader proved his alibi could not possibly be true.
- SC The cult member swore the oath had not been broken by him. It must have been the newest member, who nobody trusted.
- OTH. At the United Nations meeting, the military advisor described his activities on the night of the attack. The leader proved his alibi could not possibly be true.

15. *emphasize, hint*

- SC.SC Kyle emphasized creativity would be the key to solving the riddle. He hinted the answer might have to do with art.
- SC Kyle emphasized creativity would be the key to solving the riddle. He enjoyed giving vague clues about the puzzles he created.
- OTH. Everyone was wondering what activity Sergio had designed for the preschool children. He hinted the answer might have to do with art.

16. *worry, guarantee*

- SC.SC [The rally downtown was escalating.] The angry protestors worried the bystanders would call the police. They guaranteed violence would not break out in the street.

SC [The rally downtown was escalating.] The angry protestors worried the bystanders would call the police. They didn't want to be arrested, so they gradually dispersed.

OTH. [It was the upset of the decade in baseball.] The sports fans were getting rowdy after their team lost the World Series. They guaranteed violence would not break out in the street.

17. *find, hear*

- SC.SC The team of scientists found the solution would be expensive and unworkable. Their director heard the news was not very promising for his research program.
- SC The team of scientists found the solution would be expensive and unworkable. They would have to restart their research program using new ideas.
- OTH. The research and development team was laid off due to budget cuts. Their director heard the news was not very promising for his research program.

18. *estimate, note*

- SC.SC The technician estimated the cost of repairs would be higher than expected. Mariel noted the expense was more than the cost of a new car.
- SC The mechanic gave Mariel the bill for installing a new transmission in her car. Mariel noted the expense was more than the cost of a new car.
- OTH. The mechanic gave Mariel the bill for installing a new transmission in her car. Mariel noted the expense was more than the cost of a new car.

19. *argue, dispute*

- SC.SC The lawyer argued the case was a waste of the court's time. Her opponent disputed the conclusion was so simple and straightforward.
- SC The lawyer argued the case was a waste of the court's time. The judge had a different opinion, and listened to both sides carefully.
- OTH. According to the governor, lowering taxes would solve the state's economic crisis. Her opponent disputed the conclusion was so simple and straightforward.

20. *move, copy*

- SC.SC [Rosa was getting fed up.] She moved the board should place a time limit on speeches. The notetaker copied all speeches would be subject to new regulations. [That way, the meetings would be shorter and more productive.]
- SC [Rosa was getting fed up.] She moved the board should place a time limit on speeches. The

directors had been wasting too much time giving long, rambling speeches. [That way, the meetings would be shorter and more productive.]

- OTH. [Changes were coming the city council.] The committee imposed a new publicity policy for all public commentaries made during meetings. The note-taker copied all speeches would be subject to new regulations. [It was one more rule, but would hopefully simplify things.]
- SC
-

21. *advise, conclude*

- SC.SC [Jasmine was a graduate student in political science at Stanford.] The famous professor DeWillis advised Jasmine should critique her own writing. She concluded her essay was too long and unfocused.
- SC. [Jasmine was a graduate student in political science at Stanford.] The famous professor DeWillis advised Jasmine should critique her own writing. The strategy would transform her into a better and more independent writer.
- OTH
- OTH. [After waiting until the last minute to start her assignment, Anna couldn't procrastinate any longer.] She spent the whole afternoon hunched over her computer working on the assignment. She concluded her essay was too long and unfocused.
- SC
-

22. *decide, predict*

- SC.SC [Historically, young voters had not made much impact.] But in 2008, voters under 30 decided the election was worth participating in. They predicted the winner would face important challenges over the years.
- SC. [Historically, young voters had not made much impact.] But in 2008, voters under 30 decided the election was worth participating in. They organized energetic campaigns and voted in very high numbers.
- OTH
- OTH. [The race for governor of New York was on everyone's mind.] The political analysts started crunching numbers months before the election. They predicted the winner would face important challenges over the years.
- SC
-

23. *rule, announce*

- SC.SC [All was not well in the kingdom of Trimaris. An evil usurper had seized the throne 20 years ago.] The tyrant ruled the villagers would pay higher taxes each year. He announced the change would take place immediately. [The villagers revolted against his demands, refusing to pay the taxes.]
- SC. [All was not well in the kingdom of Trimaris. An evil usurper had seized the throne 20 years ago.] The tyrant ruled the villagers would pay higher taxes each year. Many wondered if they could
- OTH
-

escape his clutches, but had nowhere to go. [The villagers revolted against his demands, refusing to pay the taxes.]

- OTH. [A new restaurant opened up on the same street as Fred's Mexican Cafe. Of course, this created a lot of competition for him.] Unfortunately, he could no longer offer free chips and salsa at his restaurant. He announced the change would take place immediately. [Long-time customers were disappointed, but stayed loyal to Fred's.]
- SC
-

24. *declare, confess*

- SC.SC Vincent declared his affection was exclusively for Connie and no one else. She confessed her love had begun to weaken over the years.
- SC. Vincent declared his affection was exclusively for Connie and no one else. His gesture was sweet, but Connie didn't love him back.
- OTH
- OTH. None of their friends were really sure how Amanda felt about Stuart. She confessed her love had begun to weaken over the years.
- SC
-

25. *repeat, answer*

- SC.SC The angry customer repeated his request had been ignored by the employees. They finally answered his complaint would be taken seriously by the company.
- SC. The angry customer repeated his request had been ignored by the employees. He left the restaurant and started to organize a boycott.
- OTH
- OTH. Rafael sent a detailed complaint letter to the phone company and waited six weeks for a response. They finally answered his complaint would be taken seriously by the company.
- SC
-

26. *establish, continue*

- SC.SC The lawyer established the firm had violated environmental regulations. He continued the case was clearly an act of corporate wrongdoing.
- SC. The lawyer established the firm had violated environmental regulations. The owner offered him a bribe to keep the violations secret.
- OTH
- OTH. The lawyer started by pointing out examples of corruption at the top levels of the company. He continued the case was clearly an act of corporate wrongdoing.
- SC
-

27. *appreciate, recognize*

- SC.SC Tammy appreciated her guests had arrived on time to the potluck. She recognized Lucy and Ben had left work early to be there.
-

(continued on next page)

- SC. Tammy appreciated her guests had arrived on time to the potluck. She couldn't stand latecomers who would sneak in without bringing food to share.
- OTH. Deborah was in charge of running the awards ceremony, and tonight was the big night. She recognized Lucy and Ben had left work early to be there.
-

28. *read, recall*

- SC.SC. Walter read some Harry Potter books had been burned by protestors in his city. He recalled the incident had turned extremely violent. [From that moment on, he made a habit of locking his books in a vault for protection.]
- SC. Walter read some Harry Potter books had been burned by protestors in his city. They wanted to shield their children from stories about witchcraft. [From that moment on, he made a habit of locking his books in a vault for protection.]
- OTH. Hans had accidentally rear-ended another driver on Fourth Street a while back. He recalled the incident had turned extremely violent. [Fortunately, no one was hurt.]
-

29. *say, debate*

- SC.SC. Tim said his prayers would help his team win the football game. Debate Some of his fans debated the point was valid and ignored his ritual. [Naturally, their lack of faith cost them the game.]
- SC. Tim said his prayers would help his team win the football game. A few of the spectators joined his prayers, but others simply didn't care. [Naturally, their lack of faith cost them the game.]
- OTH. Robert always put coins in his pockets before championship golf games because it brought good luck. Some of his fans debated the point was valid and ignored his ritual. [But whatever it was, Robert always had a great golf game.]
-

30. *conceal, believe*

- SC.SC. [The Nguema regime was under investigation.] The government concealed the weapons had been moved from where the report said they were. The inspectors believed the report was fake and should not be trusted.
- SC. [The Nguema regime was under investigation.] The government concealed the weapons had been moved from where the report said they were. They needed to continually hide them from the weapons inspectors.
- OTH. [Paul was under suspicion of being a drug dealer.]
- SC. But according to the police report, there were no drugs hidden anywhere in Paul's house. The inspectors believed the report was fake and should not be trusted.
-

OTH.OTH items

1. Grace went for an iced coffee after her busy morning shift. She purchased the biggest size available at the coffee shop. [Once the caffeine kicked in, she was ready to face the rest of her day.]
2. The dog chased the cat that stole his treat. The cat ended up at the top of a tall tree, unable to climb back down. [Firemen had to rescue the cat and deliver it back to its family.]
3. [A young boy helped clean up the mess left in the cafeteria by some rowdy students who started a food fight.] The janitor thanked the boy who picked up the trash. The boy blushed at the compliment and continued on his way. [He was happy to have helped.]
4. Film critics from all of the major newspapers responded poorly to the movie. Actors from the movie hid in shame after the reviews. [It was Twilight all over again but with more blood and paler vampires.]
5. Clock makers no longer enjoy the success that they once did. It is now viewed mostly as a hobby, attracting a small base of customers. [To some, however, it is still considered quite the art form.]
6. The book that was reviewed by the author of another book did well. Many readers respected the author's critical interpretation more than other book reviewers.
7. [Not many people have the ability to run or jog for long periods of time, and they are always looking for alternative ways to exercise.] Mountain bike riding is a great way to stay in shape. It increases endurance and strength for its participants and can cut stress levels.
8. Kara quickly took out her camera and timed her shot perfectly. The beautiful moment was forever frozen in time on her film roll.
9. The shop owner called a meeting for all of her employees. They knew she was not happy with their performance during the past week.
10. According to the traditional view, a college education increases lifetime earnings. However, many economic experts no longer agree with this assessment.
11. Rolando's favorite things to play with as a child were balloons. Even as an adult, he couldn't resist the occasional water balloon fight. [Eventually, his children picked up the same hobby, and they enjoyed playing together on the weekends.]
12. Patrick walked away from the store in a hurry, looking over his shoulder nervously. He didn't want his mom to see what he had bought her for her fortieth birthday.
13. Many of the most popular movies were shunned by critics. Critics often look for things that average movie goers might not pay attention to.
14. Dean and Roger laughed hysterically at Jessica's joke for a full minute after she made it. Jessica was surprised at their reaction, but pleased to be so popular.
15. The prevalence of marriage has been on the decline in the United States. Many younger people prefer long term relationships to marriage.
16. Rebecca wanted to throw a party for all of her friends and coworkers. She asked her friend for permission to use her beautiful beach house to host the party.
17. Statistically speaking, playing the lottery is very unlikely to result in a win. However, many people play every week anyway.
18. [Although he was nervous to go to the prom without a date, he bravely put on his tuxedo and faced his fears. He stood against the wall for most of the night, staring out onto the

dance floor.] The girl who the boy liked danced gracefully. He watched in awe but couldn't find the courage to join her on the dance floor.

19. As a young boy, Lenny loved to play with his toy trucks. When he grew up, he became a real truck driver.
20. The Austin Fire Department is more competitive than Harvard. Only the absolute best applicants are accepted to join the fire department.
21. There was nothing Enrique's dog loved more than chewing on bones. One day, Enrique bought his dog the biggest bone the store had to offer.
22. Many of the best runners from both high schools in the city joined the team. The coach had made them an offer they couldn't refuse.
23. [Apart from eating well and exercising, there are other important ways to stay healthy.] Experts agree on the importance of drinking plenty of water regularly. Eight full glasses of water a day is their usual recommendation.
24. Baxter made a resolution to lose thirty pounds by cutting down his food intake. However, dinner had always been his favorite meal of the day. [And tonight his mom made him a delicious meatloaf with mashed potatoes.]
25. [Xavier's coach always commended him on his strong back swing, and even made him team captain.] He loved to play tennis for at least an hour every day, and had developed a high level of skill. His dad had taught him when he was a very young boy.
26. Many people choose engineering degrees because of the earnings potential. Sometimes this leads to a change of major midway through college. [For this reason, it is important to choose a major that relates to one's interests, and not necessarily just for potential of income.]
27. [Ever since the creation of Netflix, Cassie had not gone out on many dates.] She was in the habit of watching her favorite show every Friday with a big bowl of popcorn. But lately, they seemed to be showing nothing but reruns. [Suddenly, inspiration struck. She created an online dating profile.]
28. The computer programmer worked furiously through the night on the project. It was due the next morning, and his boss did not respond well to failure. [Even after many years of practice, he had not yet mastered the skill of working well under pressure. Thankfully, he had a full pot of coffee ready, and was determined to finish.]
29. In some developing nations, getting clean water is still a very difficult problem. Many charities are working to combat this health problem.
30. The soldier was gone on deployments to faraway countries for many months at a time. Her husband worked hard to cope with her absence. [For them, long distance was a struggle, but it was always worth it when they were reunited. They would always go out for a big homecoming meal to celebrate.]

B.1.2. Training phase fillers

1. Raising children is a serious responsibility. That is why many people wait until later in life to start a family. [After all, it takes both commitment and stability to support a family.]
2. Movie night was a cherished tradition for the Martins. It was cancelled at once when Chris couldn't come. [He had a paper due the next day and had to take a rain check.]
3. The baseball cap with the Giants logo on it fit far too tightly. So Cliff put it back on the rack in disappointment.

4. [Dustin received the cast list for the school musical. Gemma did not get the part.] After Dustin accidentally forwarded her the email, Gemma strode into the room with malice in her eyes.
5. [Parking was free on the weekends, but Friday counted as a weekday.] Rebecca wanted to cry when she saw the parking ticket on her windshield. She was terrified of the amount she had to pay for a stupid mistake.
6. Once he got to the edge of the playground, Samuel hopped over the railing. It seemed to be in the way for no good reason.
7. Without a doubt, being a student had its advantages for Albert. For example, he used his student status to get out of jury duty.
8. Many of the world's nations abolished slavery hundreds of years ago. Unfortunately, the practice still exists in some areas of the world today. [In fact, human trafficking is still taking place right under our noses.]
9. Back to school shopping was Heather's favorite thing to do at the end of the summer. Her mother liked to go with her so she could enjoy a few days of shopping as well. [However, Heather wanted to be able to go alone with her friends.]
10. Paul spent the entire morning playing with his presents with great joy. His parents had known exactly what he wanted when they went to buy presents. [They were very excited for the baby's first Christmas, and they were glad it went well.]
11. Daniel had a bad case of strep throat and ended up turning the assignment in a little late. Luckily his professor was very considerate and gave him a pass. [He warned him, however, to make sure it would not happen again.]
12. Bringing his mother a glass of water gave Eric satisfaction. He loved to help out around the house, even if it was just little things.
13. Riding his motorcycle was the one thing that made Clyde really love life. There was just no other experience quite like gliding down country roads on his Harley.
14. Michelle was a professional dancer, but she also loved figure skating. Her two interests usually had a lot of overlap, and she used the same techniques in both.
15. Kayla was applying to law school during her senior year of college. Unfortunately her best friend was going to medical school.

B.2. Testing phase

B.2.1. Testing phase critical items SC.SC items

-
1. *concede, claim*
Samir conceded the race had been staged from the beginning. Even so, he claimed the trophy should be his to keep at the end of the race. [Of course, the winner disagreed, and Samir went home empty-handed.]
 2. *reveal, follow*
[It was finally time for the performer's final act.] He revealed the cat was simply an electronic replica. It followed the mouse must have been electronic as well.
 3. *write, anticipate*
The retiring columnist wrote his last piece would include

(continued on next page)

a farewell to his readers. Everyone anticipated the column would be long and sentimental. [To everyone's surprise, he kept it short and sweet. It was the perfect ending to a long, thriving career.]

4. *disclose, surmise*
[Marty was a highly respectable employee at a law firm.] He had disclosed his salary was higher than the other employees' salaries. Everyone had surmised this was due to his years of experience.
5. *comprehend, infer*
[In her research, Professor Goodall come across a mysterious poem by Anpao. Every scholar before her had been baffled by it.] She comprehended the poem was full of intricate hidden meanings. She inferred its message was subtle and complex, requiring years of interpretation.
6. *indicate, feel*
Brian indicated the location of the bruise was his upper left arm. The nurse felt the area would heal on its own given enough time. [Brian was a little bummed because a cast would have looked cool.]
7. *propose, confirm*
During the meeting, Kelly proposed the solution should be more cost-effective. The treasurer confirmed the expense was greater than necessary for the company. [Buying iPads for every employee did seem like an excessive expense.]
8. *imply, mention*
[Asha was keeping a close eye on Kenji.] His behavior strongly implied an affair with Jenny was secretly developing. He mentioned his personal plans were keeping him too busy to go out with his friends. [That just didn't seem like Kenji.]
9. *signal, explain*
[Analysts are still fascinated by voting patterns in 2008.] The election results signaled a generational shift was taking place in the United States. One analyst explained the change had been expected for quite some time.
10. *lecture, demand*
The grumpy old man lectured small children were too spoiled these days. He demanded stricter rules should be enforced at all times. [The children's parents, however, angrily disagreed.]
11. *print, admit*
[The big trial finally ended, and the public was waiting impatiently for the details of the ruling.] The New York Times printed the court's decision was unanimous, but one judge had dissented. The editor admitted the error was an embarrassing oversight on his part.
12. *assert, urge*
[Determined to solidify power, the government took radical measures.] The dictator asserted his authority would be maintained through force. His advisor urged peaceful solutions should be the top priority for his government.

 SC.DO items

1. *understand, remember*
[The bank robber was not stupid.] He understood the consequences would be severe if he was caught. He remembered the robbery very vividly and in great detail.
2. *check, discover*
Monroe checked his wallet was still in his pocket after the roller coaster ride. He discovered a credit card while rummaging through it. [He wasn't sure whose it was.]
3. *assume, realize*
The ambitious advisor assumed the presidency would finally be his if the president resigned. He realized his goal during a scandal over leaked documents. [However, the public immediately became suspicious of his motives.]
4. *imagine, deduce*
[Andersen was as clever as he was creative.] He imagined a knife had been hidden in the butler's closet. He deduced the murderer during his daily walk through the park.
5. *require, publish*
The editor required the article must appear the day after election day. The newspaper published the winner too early, with only some votes counted. [Naturally, the public was surprised when the official tallies yielded different results.]
6. *proclaim, divulge*
[Everyone was waiting for the outcome of the competition.] The Olympic judge proclaimed the winner was the swimmer from the United States. He divulged his secret strategy afterward in an exclusive interview. [To many people, it was a shocker: taking ballet lessons.]
7. *figure, think*
The math student figured the sum must be greater than 100. Her friend had thought the opposite before their study session on Friday. [As a result, they went into class confused, afraid that they may be given an incomplete grade.]
8. *acknowledge, order*
[In their haste, the graphic design firm put the movie title in the wrong place.] They acknowledged their error would appear on the movie poster. The lead actor ordered the poster anyway for his studio, despite the misprint.
9. *confide, whisper*
Ruby confided her secret had been really bothering her for a long time. Her friend whispered the only solution very quietly in her ear and gave her a big hug. [Ruby came to her senses and planned to come clean.]
10. *protest, shout*
[Changes to the tax rate were announced at the city council meeting.] Valerie protested the policy was not fair to the wealthiest of the taxpayers. Someone else shouted her opinion before Valerie's turn to speak.

11. *insure, boast*
The construction company owners insured the townhouses would never flood again. They boasted their latest safety upgrades because of the owners' concerns. [That seemed to reassure most people.]
12. *advocate, report*
[Cesar Chavez was a controversial activist in the 60s and 70s.] He advocated farm workers should be paid a fair wage. The newspapers reported the movement before its peak of success.

B.2.2. Testing phase fillers

1. [Ruben knew exactly what he wanted for his fifth birthday.] Getting a new coloring book and box of fresh crayons filled Ruben with joy. Everything about coloring books was perfection in his mind.
2. Drew took the steak knife from the cupboard. He was getting ready to cook his specialty meal for his girlfriend's parents. [Unfortunately, his hands were slippery from washing his hands and he dropped the knife into the trash can.]
3. The globe sat in its typical spot in the corner of the classroom. Dust was covering most of North America. [Ever since the school board banned teaching geography, there seemed to be no use for it.]
4. [Nicole lost track of time while catching up with an old friend.] She ran to her car as quickly as she could. She couldn't afford to be late to work again or she would be fired.
5. Hunter sat in his attic watching the kids play in the rain. His mom never let him go outside in bad weather.
6. Blake was a great skater, and he spent most Saturdays at the skate park perfecting his skill. With a few more years of practice, he would have the chance to go pro.

References

- Alexandre, J. D. (2010). Modeling implicit and explicit processes in recursive sequence structure learning. In *32nd annual meeting of the Cognitive Science Society*. Portland, OR.
- Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLoS ONE*, 4, 76–78.
- Anderson, J. (1990). *The adaptive character of thought*. Lawrence Erlbaum.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- Anderson, N. H., & Whalen, R. E. (1960). Likelihood judgments and sequential effects in a two-choice probability learning situation. *Journal of Experimental Psychology*, 60, 111.
- Anderson, R. B., Tweney, R. D., Rivardo, M., & Duncan, S. (1997). Need probability affects retention: A direct demonstration. *Memory & Cognition*, 25, 867–872.
- Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Bahrack, H. (1979). Maintenance of knowledge: Questions about memory we forget to ask. *Journal of Experimental Psychology: General*, 108, 296–308.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bigand, E., Tillmann, B., Poulchin-Charronnat, B., & Manderlier, D. (2005). Repetition priming: Is music special? *The Quarterly Journal of Experimental Psychology Section A*, 58, 1347–1375.
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology General*, 129, 177–192.
- Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 468–481.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 234–272.
- Cho, R. Y., Nystrom, L. E., Brown, E. T., Jones, A. D., Braver, T. S., Holmes, P. J., et al. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective, & Behavioral Neuroscience*, 2, 283–299.
- Church, K. W. (2000). Empirical estimates of adaptation: The chance of two Noriegas is closer to $p/2$ than $\hat{p}2$. In *Proceedings of the 18th international conference on computational linguistics* (Vol. 1, pp. 180–186).
- Church, K. W., & Gale, W. A. (1995). Poisson mixtures. *Natural Language Engineering*, 1, 163–190.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235–253.
- DeRoeck, A., Sarkar, A., & Garthwaite, P. (2007). Even very frequent function words do not distribute homogeneously. In *Recent advances in natural language processing IV* (pp. 267–276). Philadelphia: John Benjamins Publishing Company.
- Drummond, A. (2012). Ibex: A web interface for psycholinguistic experiments. <<https://github.com/addrummond/ibex>>.
- Dubey, A., Sturt, P., & Keller, F. (2005). Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 827–834). Association for Computational Linguistics.
- DuBois, J. W. (2010). Towards a dialogic syntax. Ms., University of California, Santa Barbara.
- Farmer, T. A., Fine, A. B., & Jaeger, T. F. (2011). Implicit context-specific learning leads to rapid shifts in syntactic expectations. In *The 33rd annual meeting of the cognitive science society*. Boston, MA.
- Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 37, 578–591.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, 8, e77661.
- Fine, A. B., Qian, T., Jaeger, T. F., & Jacobs, R. A. (2010). Is there syntactic adaptation in language comprehension? In *Proceedings of the 2010 workshop on cognitive modeling and computational linguistics* (pp. 18–26).
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, 15, 1–16.
- Goodman, J. T. (2001). A bit of progress in language modeling. <<http://research.microsoft.com/en-us/um/people/joshuago/longcombine-short-draftonly.ps>>.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 311–347.
- Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34, 365–399.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1–8).
- Hartsuiker, R. J., 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, 214–238.
- Healey, P. G. T., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PLoS ONE*, 9, e98598.
- Heller, J., Pierrehumbert, J., & Rapp, D. (2010). Predicting words beyond the syntactic horizon: Word recurrence distributions modulate on-line long-distance lexical predictability. In *Architectures and mechanisms for language processing (AMLaP) 2010*. York, UK.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127, 57–83.
- Jones, M., & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 626–640.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20, 137–194.
- Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124, 66–71.
- Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, 133, 450–467.
- Kaschak, M. P., Loney, R. A., & Borreggine, K. L. (2006). Recent experience affects the strength of structural priming. *Cognition*, 99, B73–B82.
- Kim, S., Mauner, G., & Koenig, J. (2007). What is contingent about contingency learning? Evidence from structural priming in comprehension. *Poster presented at the 20th annual CUNY conference on human sentence processing*, La Jolla, CA.
- Kleinschmidt, D., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. *ACL HLT 2011*, 10.
- Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, 19, 332–338.
- Kuhn, R. (1988). Speech recognition and the frequency of recently used words: A modified Markov model for natural language. In *Proceedings of the 12th conference on computational linguistics* (Vol. 1, pp. 348–350).
- Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 647–652).

- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177.
- Levy, R. (2013). Memory and surprisal in human sentence comprehension. In *Sentence processing* (pp. 78–114). Hove: Psychology Press.
- Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman and Company.
- Mitchell, D. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In *New methods in reading comprehension research* (pp. 9–89). Hillsdale, NJ: L. Erlbaum Associates.
- Peterson, L., Hillner, K., & Saltzman, D. (1962). Supplementary report: Time between pairings and short-term retention. *Journal of Experimental Psychology*, 64, 550–551.
- Pickering, M. J., & 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. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134, 427–459.
- Pickering, M. J., & Traxler, M. J. (2004). Syntactic priming in comprehension. In *CUNY sentence processing conference, College Park, MD*.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003)*. March (pp. 20–22).
- Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, 35, 587–637.
- Reitter, D., Moore, J. D., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *28th annual meeting of the Cognitive Science Society* (pp. 685–690).
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57, 348–379.
- Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains: From simple arithmetic to relative-clause attachment. *Psychological Science*, 22, 1319–1326.
- Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128, 302–319.
- Soetens, E., Boer, L., & Hueting, J. (1985). Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 598–616.
- Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, 1, 113–150.
- Thomson, D. R., D'Ascenzo, M., & Milliken, B. (2012). Learning what to expect: Context-specific control over intertrial priming effects in singleton search. *Memory & Cognition*, 41, 533–546.
- Thothathiri, M., & Snedeker, J. (2008). Give and take: Syntactic priming during spoken language comprehension. *Cognition*, 108, 51–68.
- Tooley, K. M., & Traxler, M. J. (2010). Syntactic priming effects in comprehension: A critical review. *Language and Linguistics Compass*, 4, 925–937.
- Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, 29, 581–595.
- Traxler, M. J., & Tooley, K. M. (2008). Priming in sentence comprehension: Strategic or syntactic? *Language and Cognitive Processes*, 23, 609–645.
- Vazquez, A., Oliveira, J., Dezs, Z., Goh, K.-I., Kondor, I., & Barabási, A.-L. (2006). Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73.
- Wells, J., Christiansen, M., Race, D., Acheson, D., & Macdonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250–271.
- Wilder, M., Jones, M., & Mozer, M. (2009). Sequential effects reflect parallel learning of multiple environmental regularities. *Advances in Neural Information Processing Systems*, 22, 2053–2061.
- Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2013). Linguistic variability and adaptation in quantifier meanings.
- Yu, A., & Cohen, J. D. (2008). Sequential effects: Superstition or rational behavior? In *NIPS* (pp. 1873–1880).