Lexical Interference Effects in Sentence Processing: Evidence From the Visual World Paradigm and Self-Organizing Models

Anuenue Kukona, Pyeong Whan Cho, James S. Magnuson, and Whitney Tabor
University of Connecticut, and Haskins Laboratories, New Haven, Connecticut

Psycholinguistic research spanning a number of decades has produced diverging results with regard to the nature of constraint integration in online sentence processing. For example, evidence that language users anticipatorily fixate likely upcoming referents in advance of evidence in the speech signal supports rapid context integration. By contrast, evidence that language users activate representations that conflict with contextual constraints, or only indirectly satisfy them, supports nonintegration or late integration. Here we report on a self-organizing neural network framework that addresses 1 aspect of constraint integration: the integration of incoming lexical information (i.e., an incoming word) with sentence context information (i.e., from preceding words in an unfolding utterance). In 2 simulations, we show that the framework predicts both classic results concerned with lexical ambiguity resolution (Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979), which suggest late context integration, and results demonstrating anticipatory eye movements (e.g., Altmann & Kamide, 1999), which support rapid context integration. We also report 2 experiments using the visual world paradigm that confirm a new prediction of the framework. Listeners heard sentences like "The boy will eat the white . . ." while viewing visual displays with objects like a *white cake* (i.e., a predictable direct object of "eat"), *white car* (i.e., an object not predicted by "eat," but consistent with "white"), and distractors. In line with our simulation predictions, we found that while listeners fixated *white cake* most, they also fixated *white car* more than unrelated distractors in this highly constraining sentence (and visual) context.

Keywords: anticipation, artificial neural networks, local coherence, self-organization, sentence processing

Linguistic structure at the phonological, lexical, and syntactic levels unfolds over time. Thus, a pervasive issue in psycholinguistics is the question of how information arriving at any instant (e.g., an "incoming" word) is integrated with the information that came before it (e.g., sentence, discourse, and visual context). Psycholinguistic research spanning a number of decades has produced diverging results with regard to this question. On the one hand, language users robustly anticipate upcoming linguistic structures (e.g., they make eye movements to a *cake* in a visual scene on hearing "The boy will eat the . . ."[1]; Altmann & Kamide, 1999; see also Chambers & San Juan, 2008; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Knoeferle & Crocker, 2006, 2007), suggesting that they rapidly integrate information from the global context in order to direct their eye movements to objects in a visual display that satisfy contextual constraints. On the other hand, language users also seem to activate information that only indirectly relates to the global context but by no means best satisfies contextual constraints (e.g., "bugs" primes "SPY" even given a context such as "spiders, roaches, and other bugs"; Swinney, 1979; see also Tanenhaus, Leiman, & Seidenberg, 1979).

These findings pose a theoretical challenge: They suggest that information from the global context places very strong constraints on sentence processing, while also revealing that contextually inappropriate information is not always completely suppressed. Crucially, these results suggest that what is needed is a principled account of the balance between context-dependent and context-independent constraints in online language processing. In the current research, our aims were as follows: first, to show that the concept of *self-organization* provides a solution to this theoretical challenge; second, to describe an implemented self-organizing neural network framework that predicts classic findings concerned with the effects of context on sentence processing; and third, to test

[1] Throughout, when discussing stimuli from experiments we use quotation marks to indicate linguistic information (e.g., words, sentences, etc.) and italics to indicate visual display information (e.g., objects).

a new prediction of the framework in a new domain. The concept of self-organization refers to the emergence of organized, group-level structure among many, small, autonomously acting but continuously interacting elements. Self-organization assumes that structure forms from the bottom up; in the case of language processing, the assumption is that responses that are consistent with some part of the bottom-up input are gradiently activated. Consequently, it predicts bottom-up interference from context-conflicting responses that satisfy some but not all of the constraints. At the same time, self-organization assumes that the higher order structures that form in response to the bottom-up input can entail expectations about likely upcoming inputs (e.g., upcoming words and phrases). Thus, it also predicts anticipatory behaviors. Here we implemented two self-organizing neural network models that address one aspect of constraint integration in language processing: the integration of incoming lexical information (i.e., an incoming word) with sentence context information (i.e., from the preceding words in an unfolding utterance).

The rest of this article comprises four parts. First, we review psycholinguistic evidence concerned with effects of context on language processing. Second, we describe a self-organizing neural network framework that addresses the integration of incoming lexical information (i.e., an incoming word) with sentence context information (i.e., from preceding words in an unfolding utterance). We show that the framework predicts classic results concerned with lexical ambiguity resolution (Swinney, 1979; Tanenhaus et al., 1979), and we extend the framework to address anticipatory effects in language processing (e.g., Altmann & Kamide, 1999), which provide strong evidence for rapid context integration. Third, we test a new prediction of the framework in two experiments in the visual world paradigm (VWP; Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

Rapid, Immediate Context Integration

Anticipatory effects in language reveal that language users rapidly integrate information from the global context, and rapidly form linguistic representations that best satisfy the current contextual constraints (based on sentence, discourse, and visual constraints, among others). Strong evidence for anticipation comes from the VWP, which presents listeners with a visual context, and language about or related to that context. Altmann and Kamide (1999) found that listeners anticipatorily fixated objects in a visual scene that were predicted by the selectional restrictions of an unfolding verb. For example, listeners hearing "The boy will eat the . . ." while viewing a visual display with a *ball*, *cake*, *car*, and *train* anticipatorily fixated the edible *cake* predicted by the selectional restrictions of "eat." By contrast, listeners hearing "The boy will move the . . ." in a context in which all items satisfied the selection restrictions of "move" fixated all items with equal probability.

Kamide, Altmann, and Haywood (2003) also demonstrated anticipatory effects based on the combined contextual constraints from a subject and verb. Listeners hearing "The girl will ride the . . ." while viewing a visual display with a (young) *girl*, *man*, *carousel*, and *motorbike* made more anticipatory fixations during "ride" to *carousel*, which was predicted by "girl" and "ride," than to *motorbike*, which was predicted by "ride" but not "girl." Thus, anticipatory eye movements at the verb were directed toward items that were predicted by the contextual constraints (e.g., *carousel*), rather than toward items that were closely related to the unfolding utterance (e.g., the rideable *motorbike*), but not predicted by the context (for further discussion, see the General Discussion).

Closely related work on syntactic ambiguity resolution provides further support for immediate integration of contextual constraints. Tanenhaus et al. (1995) presented listeners with temporary syntactic ambiguities like "Put the apple on the towel in the box." They found that these sentences produced strong garden paths in "one-referent" visual contexts, which included objects like an *apple* (on a towel), an *empty towel* (with nothing on it), a *box*, and a *pencil*. On hearing "towel," they found that listeners tended to fixate the *empty towel*, in preparation to move the *apple* there. However, there was no evidence for garden paths in "two-referent" visual contexts, which included a second *apple* (on a napkin). On hearing "towel," they found that listeners rapidly fixated the *apple* (on a towel), and that fixations to the *empty towel* were drastically reduced (there were roughly as many fixations to the *empty towel* as with unambiguous control sentences [e.g., "Put the apple that's on the towel in the box"]). These findings suggest that "on the towel" was immediately interpreted as specifying which apple to move, rather than where to move an apple, when it could be used to disambiguate between referents in the visual context. Thus, eye movements were immediately constrained by the contextual constraints, such that fixations were not launched (i.e., relative to unambiguous controls) toward objects that were closely related to the unfolding utterance (e.g., *empty towel*), but not predicted by the context.

Related work on spoken word recognition also supports immediate effects of contextual constraints. Dahan and Tanenhaus (2004; see also Barr, 2008b) showed that "cohort effects," or the coactivation of words that share an onset (e.g., Allopenna, Magnuson, & Tanenhaus, 1998), were eliminated in constraining sentence contexts. They found that listeners hearing a neutral context (e.g., Dutch: "Nog nooit is een bok . . ."/"Never before has a goat . . ."; "een bok" ["a goat"] is the subject) showed a typical cohort effect at "bok": Listeners fixated the cohort competitor (e.g., *bot/bone*) of the target noun (e.g., *bok/goat*) more than unrelated distractors (e.g., *eiland/island*). However, listeners hearing a constraining context (e.g., "Nog nooit klom een bok . . ."/"Never before climbed a goat . . ."; again, "a goat" is the subject, and the cohort *bot* [*bone*] is not predicted by "climbed") showed no difference in looks between the cohort competitor and unrelated distractors. Thus, eye movements at noun onset were immediately directed toward objects that were compatible with the contextual constraints (e.g., *bok*), rather than toward items that were closely related to the speech signal (e.g., *bot*, which was related to the onset of the noun), but not predicted by the context. Magnuson, Tanenhaus, and Aslin (2008) found similar results in a word-learning paradigm that combined the VWP with an artificial lexicon (Magnuson, Tanenhaus, Aslin, & Dahan, 2003): There was no phonological competition from noun competitors when the visual context predicted an adjective should be heard next, and no phonological competition from adjective competitors when the visual context did not predict an adjective.

In summary, the findings reviewed in this section support the rapid, immediate integration of contextual constraints. They suggest that at the earliest possible moment (e.g., word onset; Barr, 2008b; Dahan & Tanenhaus, 2004; Magnuson et al., 2008), listen-

ers form linguistic representations that best satisfy contextual constraints. Moreover, these results suggest that listeners also do so *anticipatorily* (e.g., Altmann & Kamide, 1999). Nevertheless, results like those of Dahan and Tanenhaus (2004) are perhaps surprising in that language users do have experience with (often context-specific) uses of language that seem to violate typical expectations (e.g., a child playing with a dog bone could say, "The *bone* is climbing the dog house!"). Thus, their results suggest that verb selectional restrictions are a particularly powerful contextual constraint, which rapidly suppress representations that violate these restrictions.

Bottom-Up Interference

Although findings concerned with anticipation and ambiguity resolution in the VWP seem to suggest that listeners immediately integrate contextual information, and immediately suppress representations that conflict with contextual constraints, closely related work supports a very different conclusion. In a number of settings, language users appear to activate representations on the basis of context-independent, "bottom-up" information from the linguistic signal, such that incoming input (e.g., an incoming word) is allowed to activate linguistic representations as if there were no context, and linguistic representations that are outside the scope of the contextual constraints, or conflicting with the context constraints, are also activated.

Classic work on lexical ambiguity resolution (e.g., Swinney, 1979; Tanenhaus et al., 1979) provides evidence that language users initially activate all senses of ambiguous words, irrespective of the context. In a cross-modal priming task, listeners heard sentences that biased one interpretation of a lexically ambiguous word (e.g., "spiders, roaches, and other bugs," biasing the insect rather than espionage sense of "bugs"; Swinney, 1979). At the offset of the ambiguity, they performed a lexical decision on a visual target word. Relative to unrelated words (e.g., SEW), Swinney found equivalent priming for targets related to both senses of the homophone (e.g., ANT and SPY). Tanenhaus et al. (1979) found similar effects with ambiguous words with senses from different syntactic categories (e.g., "rose" has both a noun "flower" sense and a verb "stood" sense). For example, participants hearing "rose" in either a noun context (e.g., "She held the rose") or a verb context (e.g., "They all rose") showed cross-modal priming for "FLOWER" when it was presented 0 ms after "rose," although "FLOWER" is inappropriate in the verb context. These results suggest that ambiguous words initially activate lexical representations independent of the sentence context (as we discuss later, these priming effects eventually dissipate when contextual constraints are available).

Local coherence effects provide further evidence that language users also form syntactic representations that conflict with contextual constraints (e.g., Tabor, Galantucci, & Richardson, 2004; see also Bicknell, Levy, & Demberg, 2010; Konieczny, Müller, Hachmann, Schwarzkopf, & Wolfer, 2009; Konieczny, Weldle, Wolfer, Müller, & Baumann, 2010). Tabor et al. (2004) compared sentences like "The coach smiled at the player (who was) tossed/thrown the Frisbee" in a word-by-word self-paced reading experiment. Whereas the string "the player tossed the Frisbee" forms a locally coherent active clause (e.g., which cannot be integrated with "The coach smiled at . . ."), "the player thrown the Frisbee"

does not. Critically, they found that reading times on "tossed" were reliably slower than on "thrown," suggesting that the grammatically ruled-out active clause was activated, and interfering with processing. These results suggest that ambiguous phrases activate "local" syntactic representations independent of the "global" sentence context.

Cases where similarity in the bottom-up signal overwhelms a larger context have also been reported in spoken word recognition. Allopenna et al. (1998) found evidence of phonological competition between words that rhyme, despite mismatching phonological information at word onset. For example, listeners hearing "beaker" while viewing a visual display with a *beaker*, *beetle*, *speaker*, and *carriage* fixated *beaker* most, but they also fixated rhyme competitors like *speaker* more than completely unrelated distractors like *carriage* (with a time course that mapped directly onto phonetic similarity over time). Thus, listeners activated rhyme competitors even though they could be ruled out by the context (e.g., the onset /b/).

Findings on the activation of lexical–semantic information also suggest that language users activate information that only indirectly relates to the contextual constraints. Yee and Sedivy (2006) found that listeners fixated items in a visual display that were semantically related to a target word but not implicated by the unfolding linguistic or task constraints. For example, listeners instructed to touch a "lock" while viewing a visual display with a *lock*, *key*, *apple*, and *deer* fixated *lock* most, but they also fixated semantically related competitors like *key* more than unrelated distractors. Even more remarkably, such effects can be phonologically mediated; if *lock* is replaced with *log*, people still fixate *key* more than other distractors upon hearing "log," suggesting "log" activated "lock," which spread semantic activation to "key." Similarly, listeners also fixated *trumpet* when hearing "piano" (Huettig & Altmann, 2005), *rope* when hearing "snake" (Dahan & Tanenhaus, 2005), and *typewriter* when hearing "piano" (Myung, Blumstein, & Sedivy, 2006). These results suggest that the representations that are activated during language processing are not merely those that fully satisfy the constraints imposed by the context (which predicts fixations to the target alone, in the absence of potential phonological competitors). Rather, rich lexical–semantic information that is related to various aspects of the context (e.g., *key* is related to the word "lock" but not the task demand of touching a "lock") is also activated.

Kukona, Fang, Aicher, Chen, and Magnuson (2011) also reassessed evidence for rapid context integration based on anticipation. Their listeners heard sentences like "Toby arrests the . . ." while viewing a visual display with a character called *Toby* and items like a *crook*, *policeman*, *surfer*, and *gardener*. Listeners were told that all sentences would be about things *Toby* did to someone or something, and an image of *Toby* was always displayed in the center of the screen (thus, it was clear that *Toby* always filled the agent role). Critically, their visual displays included both a good patient (e.g., *crook*) and a good agent (e.g., *policeman*) of the verb, but only the patient was a predictable direct object of the active sentence context. Kukona et al. hypothesized that if listeners were optimally using contextual constraints, they should not anticipatorily fixate agents, since the agent role was filled by *Toby*. During "arrests," they found that listeners anticipatorily fixated the predictable patient (e.g., *crook*) most. However, listeners also fixated the agent (e.g., *policeman*) more than unrelated distractors (in fact,

fixations to the patient and agent were nearly indistinguishable at the verb), based on context-independent thematic fit between the verb and agent. Thus, items were gradiently activated in proportion to the degree to which they satisfied the contextual constraints. These results suggest that in addition to anticipating highly predictable items that satisfy the contextual constraints, language users also activate items based on thematic information independent of the sentence context.

In summary, the bottom-up interference effects reviewed in this section suggest that contextually inappropriate information is not always completely suppressed by contextual constraints. Rather, language users activate information that may only be indirectly related to, or even in conflict with, contextual constraints. Critically, these studies provide a very different insight into the representations that language users are activating: They suggest that language users' representations include rich information that is outside what is literally or directly being conveyed by an utterance.

Additionally, there is an important temporal component to these various effects: They tend to be highly "transient." For example, recall that Tanenhaus et al. (1979) found equivalent priming for "FLOWER" in both the noun context "She held the rose" and the verb context "They all rose" when probes were presented 0 ms after "rose." However, when "FLOWER" was presented 200 ms after "rose," they found greater priming in the noun context as compared to the verb context (suggesting initial exhaustive access to items matching the bottom-up input, followed by a later stage of context-based selection). Thus, although these findings suggest that contextually inappropriate information is ultimately suppressed, they nevertheless reveal an important time course component: the transient activation of context-conflicting representations (e.g., the noun sense of "rose" in a verb context like "They all rose").

Simulations and Experiments

Results like those of Swinney (1979) and Tanenhaus et al. (1979) were historically taken as support for encapsulated lexical and syntactic "modules" (e.g., J. A. Fodor, 1983), because they suggested that lexical processes were not immediately subject to the influences of context. However, this kind of modularity has been all but abandoned given more recent evidence for wide-ranging influences of context on language processing (see Rapid, Immediate Context Integration). Thus, these diverging results regarding the impact of contextual constraints on language processing raise an important question: What kind of theory would predict the language system to be both context dependent, such that language users activate representations that best satisfy the contextual constraints, and context independent, such that language users activate representations without respect to the context?

One potential solution to this question involves reassessing what we take to be the relevant contextual time window. For example, following "Toby arrests the . . ." (Kukona et al., 2011), *policeman* only conflicts with the context within a narrow time window (e.g., it is unlikely as the next word). Alternatively, in a longer (e.g., discourse) time window *policeman* may be very predictable, in so much as we often talk about policemen when we talk about arresting (e.g., "Toby arrests the crook. Later, he'll discuss the case with the on-duty policeman"). (By contrast, note that Kukona et al., 2011, interpreted their data as indicating that listeners were

activating information that was not necessarily linguistically predictable.) However, this account is problematic in two senses: First, eye movements to *policeman* quickly decrease following "arrest," although *policeman* presumably becomes more discourse predictable further on from the verb (e.g., it is likely in a subsequent sentence). More importantly, this account also cannot explain other effects of context-independent information. For example, there is not an apparent (e.g., discourse) time window in which "FLOWER" is relevant to "They all rose" (Tanenhaus et al., 1979), "the player tossed the Frisbee" as an active clause is relevant to "The coach smiled at the player tossed the Frisbee" (Tabor et al., 2004), or *speaker* is relevant to "beaker" (Allopenna et al., 1998). Thus, our goal with the current research was to develop a unifying framework that explains a wide range of evidence supporting both bottom-up interference from context-independent constraints and rapid context integration.

Here we argue that the concept of self-organization offers a key insight into this puzzle. The term *self-organization* has been used to describe a very wide range of phenomena. In physics, it is associated with a technical notion, self-organized criticality (Jensen, 1998), which describes systems that are poised between order and chaos. In agent-based modeling, the term is used to describe systems of interacting agents that exhibit systematic behaviors. For example, Reynolds (1987) showed that "coordinated" flocking behavior emerges ("self-organizes") among agents who are following very simple rules (e.g., maintain a certain distance from neighbors and maintain a similar heading). We adopt the following definition: *Self-organization* refers to situations in which many, small, autonomously acting but continuously interacting elements exhibit, via convergence under feedback, organized structure at the scale of the group.

Self-organizing systems in this sense are often specified with sets of differential equations. We use such a formulation in the simulations we describe below. Well-known examples of self-organization from biology include slime molds (e.g., Keller & Segel, 1970; Marée & Hogeweg, 2001), social insects (e.g., Gordon, 2010), and ecosystems (Solé & Bascompte, 2006). In each of these cases, the continuous bidirectional interactions between many small elements cause the system to converge on coherent responses to a range of environmental situations. We claim that language processing works in a similarly bottom-up fashion: The stimulus of the speech stream prompts a series of low-level perceptual responses, which interact with one another via continuous feedback to produce appropriate actions in response to the environment. Previously proposed bottom-up language processing models such as the interactive activation model of letter and word recognition (McClelland & Rumelhart, 1981), TRACE (McClelland & Elman, 1986), and DISCERN (Miikkulainen, 1993) are self-organizing in this sense.

Critically, structure forms from the bottom up in self-organization. Initially, all possible responses that are consistent with some part of the bottom-up input (e.g., phoneme, morpheme, word) are activated. But the response that is most consistent with all the parts of the input will generally get the strongest reinforcement through feedback, and will thus come to dominate, while the other possible responses will be shut down through inhibition. Thus, signs of bottom-up interference (e.g., activation of context-conflicting responses; see Bottom-Up Interference) will be transitory. In fact, if the contextual constraints are very strong, and the

bottom-up signal supporting context-conflicting responses is very weak, then there will be no detectable interference. This latter case is hypothesized to hold in cases of rapid context integration (see Rapid, Immediate Context Integration). Indeed, even in the classic case of lexical ambiguity resolution (e.g., Swinney, 1979; Tanenhaus et al., 1979), the early activation of context-conflicting lexical items can be eliminated with a sufficiently strong context (e.g., closed-class words like "would" do not prime open-class homophones like "wood" in strongly constraining contexts that only predict a closed-class word; Shillcock & Bard, 1993). Here we focus on the integration of incoming lexical information (i.e., an incoming word) with sentence context information (i.e., from preceding words in an unfolding utterance), and we show that adopting a self-organizing perspective unifies a range of findings from the psycholinguistics literature.

The Self-Organizing Neural Network Framework

Artificial neural networks are systems of simple, interacting neuron-like units. Continuous settling recurrent neural networks are systems whose activations change continuously and whose units have cyclic connectivity. These networks have the architectural prerequisites for self-organization as defined above: Their many small units interact via bidirectional feedback connections to converge on a coherent response to their input. Previously, continuous settling recurrent neural networks have been used to model, and to generate predictions about, many aspects of human sentence processing (e.g., Kukona & Tabor, 2011; Spivey, 2007; Spivey & Tanenhaus, 1998; Tabor & Hutchins, 2004; Vosse & Kempen, 2000; see also MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994). Here we describe continuous settling recurrent neural networks that address the integration of incoming lexical information (i.e., an incoming word) with sentence context information (i.e., from preceding words in an unfolding utterance). We focus on two key findings in the literature: evidence for context-independent processing, based on studies of lexical ambiguity resolution (e.g., Swinney, 1979; Tanenhaus et al., 1979), and evidence for rapid context integration, based on studies of anticipatory effects in the VWP (e.g., Altmann & Kamide, 1999).

Prior research suggests that artificial neural networks are well suited to capturing effects of both context-dependent and context-independent information. Tabor and Hutchins (2004) implemented an artificial neural network model of sentence processing, called self-organizing parsing (SOPARSE), in which lexically activated treelets (J. D. Fodor, 1998; Marcus, 2001) self-organize into syntactic parse structures. The interacting lexical elements in SOPARSE act *semiautonomously*: Although each element has a context-independent mode of behavior, its behavior is continuously dependent on its interactions with other elements. Because organized structure (as reflected in activation patterns across the network) emerges out of the interactions of the semiautonomous elements in SOPARSE, the model naturally predicts context-independent processing effects: The semiautonomous elements are free to simultaneously form various (e.g., context-conflicting) syntactic parses, even as they form a syntactic parse that best satisfies the contextual constraints. Consequently, the global coherence of syntactic parses is not enforced in SOPARSE, thus predicting

effects of context-independent information, like local coherence effects (e.g., Tabor et al., 2004).

However, feedback mechanisms in SOPARSE (Tabor & Hutchins, 2004) also allow structures that best satisfy the constraints to inhibit competing, context-conflicting structures over time (i.e., from the point of encountering the ambiguous bottom-up input), thus predicting that context-independent processing effects will be transient (e.g., SOPARSE predicts priming of "FLOWER" immediately after "They all rose," but the absence of priming after a short delay; Tanenhaus et al., 1979).

Simulation 1: Self-Organizing Neural Network Approach to Lexical Ambiguity Resolution

In Simulation 1, we implemented a network that addressed the classic results of Swinney (1979) and Tanenhaus et al. (1979). Again, these authors showed that listeners activated both contextually appropriate and contextually inappropriate senses of lexically ambiguous words (homophones) immediately after hearing them in constraining sentence contexts.

In Simulation 1, our neural network learned to activate sentence-level phrasal representations while processing sentences word by word. The simulation included only essential details of the materials and task: The network was built of simple word input units with feedforward connections to many phrasal output units, which were fully interconnected and continuously stimulated each other. Additionally, the network acquired its representations via a general learning principle. In Simulation 1, the network was trained on a language that included a lexically ambiguous word like "rose," which occurred as both a noun ("The rose grew") and a verb ("The dandelion rose"). The critical question is whether we see activation of the verb sense of "rose" in noun contexts and activation of the noun sense of "rose" in verb contexts.

Our strategy with the current set of simulations was to model the integration of lexical and sentential information given a minimally complex network that was capable of capturing the critical phenomena. Although implementing models that capture a wide variety of phenomena can provide insight into cognitive phenomena, such an approach often requires very complex models that can be very difficult to understand. Consequently, we implemented simulations that were complex enough to make critical behavioral distinctions but also relatively simple to analyze. As a result, the current simulations are considerably less assumption-laden than, for example, SOPARSE (Tabor & Hutchins, 2004). Also in contrast to SOPARSE, the current simulations used simple phonological-form detectors, like Elman's (1990, 1991) simple recurrent network (SRN), rather than linguistic treelets.[2]

Method.

Architecture. The architecture of the network is depicted in Figure 1. The network consisted of a word input layer, with 12 localist word nodes (i.e., one word per node) of various syntactic classes (determiners: "the," "your"; nouns: "dandelion," "tulip"; verbs: "grew," "wilted"; adjectives: "pretty," "white"; adverbs:

---

[2] The current model is also simpler than the SRNs used by Elman (1990, 1991), in that it did not use hidden units. One shortcoming of the current model is that it cannot handle arbitrary long-distance dependencies (whereas the Elman net and its relatives plausibly can; e.g., Tabor, 2000; Tabor, Cho, & Szkudlarek, 2013).

phrasal layer

```
┌─────────────────────────────────┐
│  27 output units (phrases)      │──┐
└─────────────────────────────────┘  ○
              ▲
┌─────────────────────────────────┐
│  12 input units (words)         │
└─────────────────────────────────┘
```
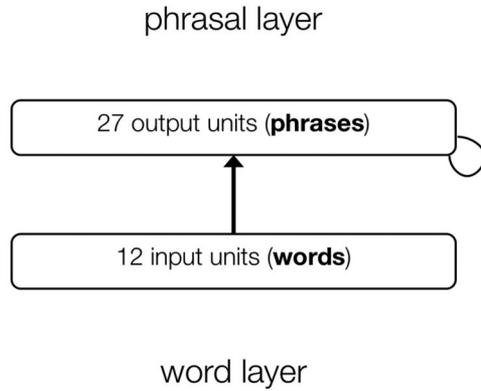
word layer

Figure 1. Simulation 1: Architecture of the self-organizing neural network. The network had feedforward connections from the word input layer to the phrasal output layer and recurrent connections within the phrasal output layer.

"slowly," "quickly"; and a lexically ambiguous noun/verb: "rose"), and a phrasal output layer, with 27 localist noun or verb phrasal nodes (reflecting the grammatical combinations of word nodes; e.g., NP: "the dandelion," "the pretty dandelion," "the white dandelion," etc.; VP: "grew," "slowly grew," "quickly grew," etc.). Thus, the output phrasal layer used distributed representations to a limited degree, in that the network activated both an NP and VP node for each sentence. We included this limited degree of abstraction in order to show that the network can interact effectively with a mix of local and distributed representations (distributed representations are more plausible given complex combinatorial structures, which would alternatively require a very large number of localist representations). The network had feedforward connections from the word input layer to the phrasal output layer, and recurrent connections within the phrasal output layer. Activations in the word input layer corresponded to words the network "heard," and activations across the phrasal output layer corresponded to the network's sentence-level representations.

Processing. Our network is closely related to a number of artificial neural network models in the literature, although it differs in certain key respects. Like the SRNs used by Elman (1990, 1991) to model sentence processing, the network processed sentences word by word: Processing of a word was implemented by setting the activation for a word node to 1 in the word input layer and all other word nodes to 0. Unlike the typical approach with SRNs, word units were turned on for an interval of time, corresponding to the fact that words have temporal durations. In this regard, the network resembles TRACE (McClelland & Elman, 1986), a model of spoken word recognition that takes in input across continuous time (however, in TRACE the input is gradient acoustic/phonological information, rather than categorical localist word inputs). Our localist inputs are a convenient simplification that were sufficient to characterize the phenomena we focus on here (however, graded, distributed representations, which can characterize similarity in phonological and semantic structure, are quite compatible with this framework).

So as to not introduce a duration-based bias, all words processed by the network were of the same duration (40 time steps). Explorations of the network's behaviors suggest that although it required

more training for shorter word durations (i.e., to reach a similar level of performance), the behavior of the network at the end of training remained similar across a wide range of values for the word duration parameter.

The phrasal output layer activations changed according to Equations 1 and 2. In these equations, net_i is the net input to the *i*th unit; a_j and a_i are the activation of the *j*th and *i*th units, respectively; w_{ij} is the weight from the *j*th to *i*th unit; and η_a is a scaling term ($\eta_a = .05$). Equation 2 creates a sigmoidal response as a function of net input (net_i).[3]

$$\text{net}_i = \sum_{j \in \text{units}_{\text{sending}}} a_j w_{ij} \qquad (1)$$

$$\Delta a_i = \eta_a \text{net}_i a_i (1 - a_i) \qquad (2)$$

Training. The network was trained on a language of 162 sentences (e.g., "The pretty dandelion slowly wilted," "Your tulip grew"), reflecting the grammatical combinations of words from the network's language. The set of training sentences is specified by the grammar in Appendix A. The network was trained to activate both the NP and VP node in the phrasal output layer that was associated with each sentence (e.g., NP: "The pretty dandelion" and VP: "slowly wilted" for "The pretty dandelion slowly wilted") as each word in a sentence was heard. The network was trained on a set of 5,000 sentences, which were selected at random from the network's language. We stopped training after 5,000 sentences, at which point the network showed a high level of performance in activating target nodes (performance statistics are reported in Results). We report the average behavior of the network across 10 simulations. Learning was implemented via the delta rule in Equation 3, with weights updated once per word presentation. In this equation, t_i is the target activation of the *i*th unit in the phrasal output layer, and η_w is the learning rate ($\eta_w = .02$).

$$\Delta w_{ij} = \eta_w (t_i - a_i) a_i \qquad (3)$$

Results. To confirm that the network learned the training language, we tested the network on one sentence of each of the four syntactic types (i.e., Det N V; Det Adj N V; Det N Adv V; Det Adj N Adv V) at the end of training. At the offset of each test sentence, the network activated the target NP node more than all nontarget nodes with all test sentences (accuracy: $M = 1.00$, $SD = .00$), and the target VP node more than all nontarget nodes in all but two instances across the 10 simulations (accuracy: $M = .95$, $SD = .11$). Although we examined variations in a range of free parameters (e.g., word durations, number of simulations), we did not need to engage in any formal or informal parameter fitting; the behavior of the network was consistent across a wide range of parameter settings given training to a similar level of performance. Following training, we tested the network with the example sentences "The rose grew" (noun context) and "The dandelion rose" (verb context). Activations of relevant NP and VP nodes are plotted in both the noun and verb contexts in Figures 2A and 2B. Again, the critical question is whether we see greater activation in the noun context of the VP nodes associated with the verb "rose"

---

[3] Equation 2 combines the excitatory and inhibitory equations used in other language processing models (e.g., interactive activation model of letter and word recognition, McClelland & Rumelhart, 1981; TRACE, McClelland & Elman, 1986) into a single equation, without decay terms.
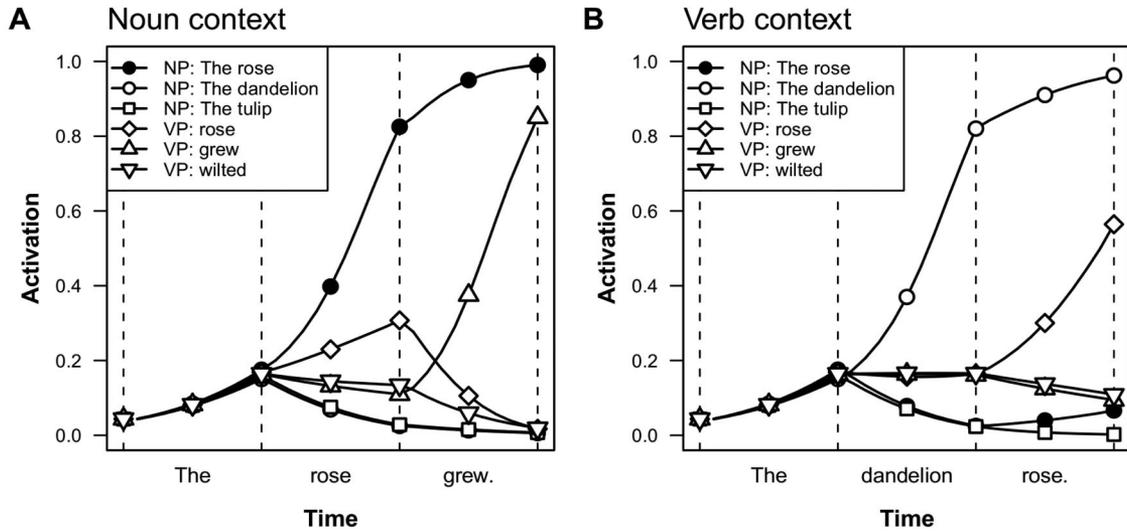
Figure 2.   Simulation 1: Activations of selected phrasal output nodes in the noun context "The rose grew" (A) and the verb context "The dandelion rose" (B). NP = noun phrase; VP = verb phrase.

(e.g., "VP: rose") as compared to VP nodes not associated with the verb "rose" (e.g., "VP: grew," "VP: wilted"), and whether we see greater activation in the verb context of the NP nodes associated with the noun "rose" (e.g., "NP: rose") as compared to NP nodes not associated with the noun "rose" (e.g., "NP: dandelion," "NP: tulip").[4]

First, in both contexts we observed greatest activation of the appropriate NP and VP structures at the end of each sentence: In Figure 2A, the network activated "NP: The rose" and "VP: grew" most when hearing "The rose grew," and in Figure 2B, the network activated "NP: The dandelion" and "VP: rose" most when hearing "The dandelion rose." These results suggest that the network learned the structure of the language. Second, we also found evidence for bottom-up interference from context-independent constraints in both the noun and verb contexts. In Figure 2A, the network activated "VP: rose" more than both "VP: grew" and "VP: wilted" when hearing "rose" in a noun context. In Figure 2B, the network activated "NP: The rose" more than "NP: The tulip" when hearing "rose" in a verb context (note that "NP: The dandelion" is activated more than "NP: The rose" because the network just heard "The dandelion").

Discussion.   The network highly activated phrasal structures that were appropriate for the sentence context (e.g., "NP: The rose" and "VP: grew" when hearing "The rose grew"). However, in line with Tanenhaus et al. (1979), the network also activated context-conflicting structures based on context-independent lexical information (e.g., "VP: rose" given "rose" in a noun context, and "NP: rose" given "rose" in a verb context). These results reveal that the network was gradiently sensitive to both context-dependent constraints, based on the sentence context, and context-independent constraints, based on context-independent lexical information, consistent with a large body of evidence supporting both rapid context integration (see Rapid, Immediate Context Integration) and bottom-up interference (see Bottom-Up Interference).

The transitory activation of both senses of the ambiguous word "rose" stems from the self-organizing nature of the trained net-

work. The learning rule causes positive connections to form between the input unit for "rose" and all output units involving "rose." It also causes inhibitory connections to form among all the NP output nodes (because only one NP percept occurs at a time) and among all the VP output nodes (for the corresponding reason). Consequently, even though the VP nodes associated with "rose" (e.g., "VP: rose") are activated when the input is "rose" in a noun context, this errant activation is shut down once the appropriate VP node (e.g., "VP: grew") becomes sufficiently activated to inhibit competitors (e.g., "VP: rose").[5]

Simulation 2: Self-Organizing Neural Network Approach to Anticipatory Effects

In Tanenhaus et al. (1979), the contextual information logically ruled out words from particular syntactic classes (e.g., verbs in a noun context and nouns in a verb context). However, their contexts were not completely constraining in the sense that the word classes that were not ruled out by the contextual constraints still contained massive numbers of words (cf. Shillcock & Bard, 1993). By contrast, in studies like Altmann and Kamide (1999) and Dahan

---

[4] Note that the model was not trained on semantic category information (i.e., it was not trained that the nouns "dandelion," "rose," and "tulip" all belong to the semantic category of "flowers"), and thus it would not be expected to show semantic priming effects (i.e., between nouns).

[5] For clarity, it is worth noting that there is another timescale at which the present network model is self-organizing: the timescale of the learning. In this case, the successive events of word processing that occur over many sentences during the training phase cause the weights to move into relationships in which the network coherently responds to its environment. This convergence of the weight relationships occurs via feedback among the weights themselves: Adjustment of one weight often requires complementary adjustment of other weights. We do not discuss the self-organizing aspects of weight change further because this article is focused on online processing as opposed to learning (for other recent modeling work addressing interactions between learning and processing, see, e.g., Dell & Chang, in press; Fine & Jaeger, 2013; Jaeger & Snider, 2013).

and Tanenhaus (2004), the contextual information ruled out all but a single outcome (e.g., the *cake* in the scene following "The boy will eat the . . ."). Critically, in this completely constraining context (i.e., in which there was a single outcome), Dahan and Tanenhaus (2004; see also Barr, 2008b; Magnuson et al., 2008) found that context-independent activation of cohort competitors was eliminated (e.g., the cohort competitor *bot* [*bone*], which was not predicted by the verb "klom" ["climbed"], was not fixated more than unrelated distractors in the context of "Nog nooit klom een bok . . ." ["Never before climbed a goat . . ."]). In Simulation 2, we tested the network in a similarly constraining anticipatory VWP context. Again, the simulation included only essential details of the materials and task: The network was built of simple word input units with feedforward connections to object output units, which interacted with one another via recurrent connections.

In Simulation 2 (and the experiments that follow), we used contexts involving strong verb selectional restrictions that were closely based on the materials of Altmann and Kamide (1999). We presented the network (followed by human listeners) with sentences like "The boy will eat the white cake" while viewing visual displays with objects like a *white cake*, *brown cake*, *white car*, and *brown car*. The direct object of the sentence (e.g., *white cake*) could be anticipated at the adjective, based on the contextual constraints from "eat" and "white" (predicting *white cake*, and not any of the other objects in the visual display). However, the visual display also included a "locally coherent" competitor (e.g., inedible *white car*), which fit the context-independent lexical constraints from the adjective (e.g., "white") but was not predicted by the contextual constraints of the verb selection restrictions of "eat"). Critically, we tested for the anticipation of objects predicted by the contextual constraints (e.g., *white cake*), and for the activation of objects based on context-independent lexical constraints (e.g., *white car*, which was ruled out by "eat," but consistent with "white"). In Experiments 1 and 2, which follow, we tested the predictions of Simulation 2 with human language users.

Method.

Architecture. The architecture of the network is depicted in Figure 3. The network consisted of a language input layer, with six localist word nodes, and a visual output layer, with four localist object nodes. As in Simulation 1, the network had feedforward connections from the input layer to the output layer and recurrent connections within the output layer. Activations in the language input layer corresponded to words the network heard, and activations across the visual output layer corresponded to a probability distribution of "fixating" each object in the visual context. Thus, the network modeled the central tendency of fixation behaviors (e.g., average proportions of fixations) rather than individual eye movements.

Processing. Processing was identical to that of Simulation 1. Additionally, activations across the four object nodes in the visual output layer were interpreted as fixation probabilities, p_i, after normalization via Equation 4.

$$p_i = \frac{a_i}{\sum_{j \in \text{units}_{\text{output}}} a_j} \qquad (4)$$

Training. The network was trained on a set of four verb-consistent sentences (e.g., "Eat [the] white/brown cake," "Drive
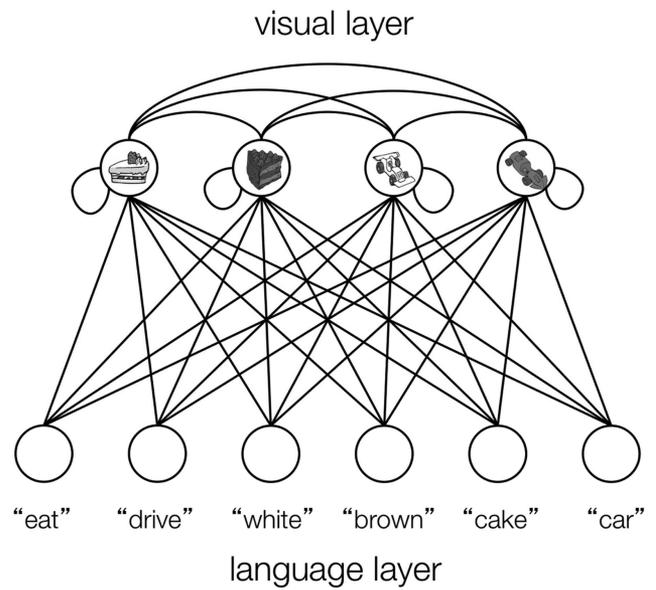

Figure 3. Simulation 2: Architecture of the self-organizing neural network. The network had feedforward connections from the language input layer to the visual output layer and recurrent connections within the visual output layer.

[the] white/brown car"). To simplify the structure of the language, we used imperative sentences that lacked an overt subject noun phrase (as well as a direct object determiner). However, these sentences were comparable to the materials in Altmann and Kamide (1999) because the subject in those sentences could not be used to anticipate the direct object (e.g., "The boy . . ." does not predict edible vs. inedible). The network was trained to activate the object node in the visual output layer that was the direct object of each sentence (e.g., *white cake* for "Eat [the] white cake"), as each word in a sentence was heard by the network (e.g., as the word nodes for "eat," "white," and "cake" in the language input layer were turned on word by word). As in Simulation 1, learning was implemented via the delta rule. Training epochs consisted of training on each sentence once in a randomized order. The network was trained for 250 epochs.

Results. To confirm that the network learned the training language, we tested the network on each of the four training sentences at the end of training. At the offset of each test sentence, the network activated the target object node more than all nontarget nodes with all test sentences (accuracy: $M = 1.00$, $SD = .00$). At various points during training, we also tested the network with the example sentence "Eat [the] white cake." Simulated proportions of fixations to objects in Epochs 1, 50, 100, 150, and 250 are plotted in Figures 4A–4E. The total sum squared error in the network across training is also plotted in Figure 4F. Again, the critical question is whether we see greater activation following "Eat [the] white . . ." of *white car*, the locally coherent competitor, compared to *brown car*, an unrelated distractor.

Following Epoch 1, the fixation pattern was as follows: Simulated proportions of fixations to the cakes diverged from the cars at "eat," and simulated fixations to *white cake* diverged from *brown cake* at "white." Critically, the network also fixated *white*
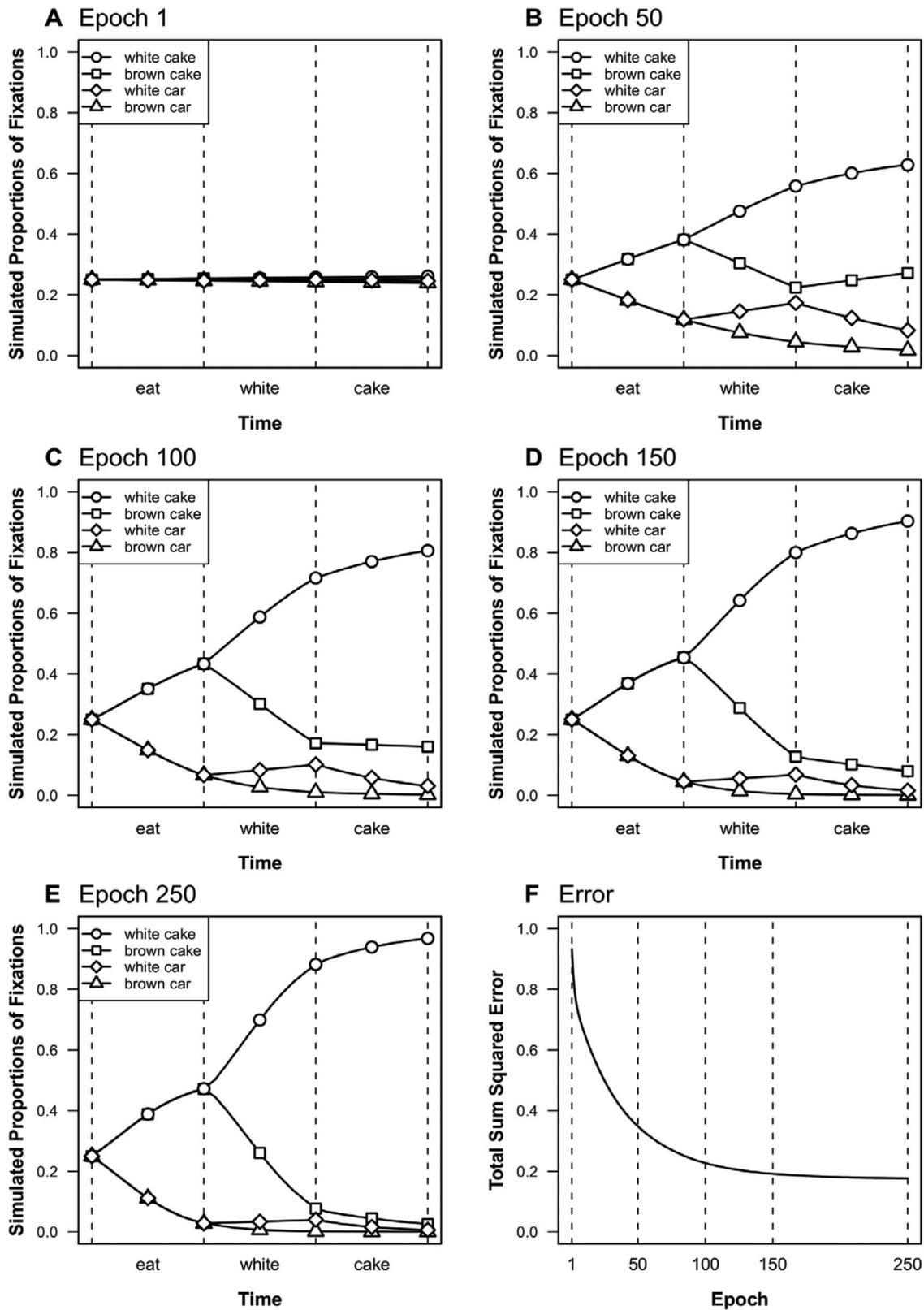
Figure 4. Simulation 2: Simulated proportions of fixations to objects in the visual layer at Epoch 1 (A), 50 (B), 100 (C), 150 (D), and 250 (E), for the example sentence "Eat [the] white cake," and the total sum squared error in the network across training (F).

car more than *brown car* at "white" (and "cake"). Further training of the network had the effect of increasing looks to the *white cake* and decreasing looks to the remaining objects. Additionally, the magnitude of the difference between *white car* and *brown car* diminished with further training (for further discussion, see Evaluation of the Self-Organizing Neural Network Approach).[6]

Discussion. The network anticipated target objects based on the combined contextual constraints from the verb and adjective (e.g., *white cake* given "Eat [the] white . . ."). The network's incremental sensitivity to the contextual constraints is consistent with prior demonstrations of anticipatory eye movements in the VWP (e.g., Altmann & Kamide, 1999; and specifically, anticipatory effects based on multiword constraints; e.g., Kamide, Altmann, & Haywood, 2003). However, the network also activated context-conflicting structures based on context-independent lexical information from the adjective (e.g., the locally coherent competitor *white car* given "Eat [the] white . . ."). As in Simulation 1, these results reveal that the network was gradiently sensitive to both context-dependent constraints, based on the sentence context, and context-independent constraints, based on context-independent lexical information, despite the completely constraining context. Again, these gradient effects are compatible with a large body of evidence supporting both rapid context integration (see Rapid, Immediate Context Integration) and bottom-up interference (see Bottom-Up Interference). In Experiments 1 and 2, we directly test the predictions of Simulation 2 with human language users.

Evaluation of the Self-Organizing Neural Network Approach

Simulations 1 and 2 reveal that the network robustly predicts gradient effects of context-dependent and context-independent information in language processing, thus providing a coherent account of results showing both rapid context integration and bottom-up interference. In both simulations, the network learned to balance context-dependent and context-independent information via a general learning principle. The networks are self-organizing by the definition stated above: Bottom-up stimulation causes the output units to continuously interact via bidirectional feedback connections and to converge on a coherent response to each environmental stimulus. Unlike with some other self-organizing models (e.g., interactive activation model of letter and word recognition, McClelland & Rumelhart, 1981; TRACE, McClelland & Elman, 1986; DISCERN, Miikkulainen, 1993), the converged state is not a distributed representation with many activated units. Unlike some other self-organizing phenomena, the converged state is also not an elaborate, temporally extended pattern of activity, as in the case of foraging for nectar in environmental locations where flowers are abundant (e.g., bee colony self-organization; Hölldobler & Wilson, 2008). Instead, the converged state in the current networks (i.e., as reflected in the training outputs) is a two-on bit vector specifying the syntactic structure of a sentence (Simulation 1) or a localist vector specifying a locus of eye fixation (Simulation 2). Despite the natural intuition that self-organization involves the formation of very complex (possibly temporally extended) structures, we are suggesting here that the crucial property that distinguishes self-organizing systems from other kinds of systems is that they respond coherently to their environment via continuous

feedback among bidirectionally interacting elements. The current networks fit this description. On the one hand, the current networks also have simple converged states. However, this is because we specified a simple encoding for their output behaviors. We did not use these simple encodings because we think that human responses to language stimuli are simple, or because we believe that a two-layer featural network is adequate; hidden units and multiple, recurrently connected layers are likely needed. Rather, we made these assumptions to highlight crucial features of self-organization, and their relation to key data points, while using a network that can be straightforwardly analyzed. Nevertheless, Simulation 1 provides an indication of how the framework can scale up to languages of greater complexity, and Simulation 2 reveals that the framework captures results from a number of experimental paradigms (e.g., VWP).

Helpfully, consideration of the current simulations leads to an answer to the question of why human language users show bottom-up interference effects: The simulations suggest that they may be a residue of the learning process. In the network, context-independent effects depended on first-order associations between inputs (e.g., words) and outputs (e.g., phrases/objects), whereas context-dependent effects depended on higher order associations across sequences of inputs. Crucially, the network must first be able to detect first-order input–output associations, before it can detect higher order associations across sequences of inputs. Therefore, context-independent processing of inputs occurs first during training (thus, bottom-up interference from *white car* is greatest early in training; see Figure 4B), and context-dependent processing develops as a modulation of context-independent effects later in training (thus, fixations to *white car* diminish with further training).

In connection with prior models, the framework also provides coverage of a wide variety of phenomena that support bottom-up interference from context-independent constraints. In addition to the results of Swinney (1979) and Tanenhaus et al. (1979), we highlighted two closely related findings in the introduction: local coherence effects (e.g., activation of "the player tossed the Frisbee" when reading "The coach smiled at the player tossed the Frisbee"; Tabor et al., 2004) and rhyme effects (e.g., activation of *speaker* when hearing "beaker"; Allopenna et al., 1998). As we reviewed, Tabor and Hutchins's (2004) SOPARSE model captures local coherence effects. Relatedly, Magnuson et al. (2003) used an SRN (Elman, 1990, 1991) to capture rhyme effects. Their model used localist phonetic feature inputs and localist lexical outputs. Critically, both SOPARSE and SRNs are compatible with our self-organizing neural network framework: They share the core property that coherent structures emerge via repeated feedback interactions among multiple interacting elements.

However, our simulations also raise a number of important issues. One critical simplification is the network's feedforward connections. For example, activations in Simulation 2 flowed from the language layer to the object layer but not back from the object layer to the language layer. The implication is that language

---

[6] The magnitude of the difference between *white car* and *brown car* initially increased in size during training: The average difference in fixations was roughly zero in Epoch 1 (see Figure 4A) and grew to a peak around roughly Epoch 33 (shortly before Figure 4B).

processing influences eye movements but that visual context does not influence language processing, counter to a number of findings (e.g., Tanenhaus et al., 1995). In fact, the self-organizing framework is fully compatible with visual to language feedback connections. For example, Kukona and Tabor (2011) modeled effects of visual context on the processing of lexical ambiguities using an artificial neural network that included visual to language feedback connections. Critically, they observed closely related bottom-up interference effects even in the context of these visual to language feedback connections.

A second issue raised by the current simulations concerns training. In fact, the magnitude of the difference between *white car* and unrelated distractors in Simulation 2 diminished over training, such that if training were allowed to continue for an infinite length of time, the effect would disappear altogether (given limitations on the numerical precision of our computer-based simulations, the effect actually disappears before then). Thus, a critical question is, at which point in the network's training should its behavior be mapped onto human language users? Our predictions are drawn from the network when it is in a transient state, and there is room for error to improve (even if to a very small degree; e.g., prior to Epoch 250), rather than when it is in an asymptotic state, and error has stabilized (e.g., beyond Epoch 250). This strategy closely follows that of Magnuson, Tanenhaus, and Aslin (2000), who similarly ended training in their SRN before it learned the statistics of its training language perfectly. They suggest that ending training before asymptote provides a closer "analog to the human language processor," given the much greater complexity of human language (i.e., than typical training languages) and the much greater amount of training required to reach asymptote (p. 64; see also Magnuson et al., 2003).

Moreover, the network also fixates *white car* more than unrelated distractors over essentially its entire training gradient (with the exception of Epoch 1): When total sum squared error is anywhere between its maximum and minimum (see changes in error along the *y*-axis in Figure 4F), we observe the same qualitative pattern, which changes only in magnitude (e.g., see Figures 4B–4E). Thus, our predictions do not reflect precise training criteria; rather, our predictions reflect the behavior of the network over all but a small (initial) piece of its training gradient. Additionally, the pattern of results we observed over the course of training provides predictions about human language development and individual differences, which we return to in the General Discussion.

Finally, it is important to note that the current predictions stem from the self-organizing nature of the network in combination with constraints on its training. Certainly, we are not simply drawing our predictions from the network when it is untrained and exhibiting largely random behavior (e.g., Epoch 1; see Figure 4A). Rather, the network simultaneously shows bottom-up interference while also showing a massive anticipatory effect (e.g., Epoch 250; see Figure 4E), suggesting that it has considerable knowledge of the structure of its language even while showing the crucial interference effect (i.e., not unlike human language users). Moreover, the current predictions are not a direct and unambiguous consequence of training to less-than-asymptote. For example, an alternative approach might plausibly predict a very similar pattern to that of Dahan and Tanenhaus (2004; e.g., no *white car* vs. *brown car* difference, suggesting no bottom-up interference), but with

fewer looks to targets, and more looks to all nontargets (i.e., irrespective of whether or not they are locally coherent), early in training. Crucially, the current predictions depend on the assumption that responses that are only consistent with parts of the bottom-up input also get activated.[7]

Experiments

In Experiments 1 and 2, we tested for bottom-up interference in contexts involving strong verb selectional restrictions (e.g., Altmann & Kamide, 1999; Dahan & Tanenhaus, 2004). Our experiments followed directly from Simulation 2: Human language users heard sentences like "The boy will eat the white cake" while viewing a visual display with objects like a *white cake*, *brown cake*, *white car*, and *brown car* (see Figure 5). The direct object (e.g., *white cake*) could be anticipated based on the combined contextual constraints from "eat" and "white." Critically, the visual display also included a locally coherent competitor (e.g., *white car*), which fit the lexical constraints from "white," but was not predicted by the selection restrictions of "eat." Finally, the visual display also included a verb-consistent competitor (e.g., *brown cake*, which was predicted by "eat" but not "white"), to test for the impact of the joint verb plus adjective constraints, and an unrelated distractor, which was either similar (Experiment 1: e.g., *brown car*) or dissimilar (Experiment 2: e.g., *pink train*) to the locally coherent competitor.

In both experiments, we tested two key questions. First, do listeners anticipatorily fixate predictable target objects (e.g., *white cake*) that best satisfy the contextual constraints (e.g., as compared to all other objects)? And second, do listeners fixate locally coherent competitor objects (e.g., *white car*) based on context-independent lexical constraints (e.g., as compared to unrelated distractors)?[8] Simulation 2 predicts most anticipatory fixations following the adjective to the predictable target. However, the simulation also predicts more fixations to the locally coherent competitor than unrelated distractors, on the basis of context-independent lexical information. By contrast, if verb selectional restrictions completely rule out representations that are inconsistent with those constraints, then we expect anticipatory fixations

---

[7] An anonymous reviewer asked us to clarify how the network's behavior would differ given a language with different word ordering. The structure of the language in Simulation 1 is based closely on English, in which adjectives precede nouns (e.g., "white cake"). This is not true of all languages. In French the order is reversed ("gâteau blanc"). Critically, our simulation results would thus differ for languages like French: Note that in English, following "Eat the white," *white car* is consistent with one word (e.g., "white") and inconsistent with one word ("eat"). In French, following an utterance like "Tu manges gâteau blanc," *white car* is consistent with one word (e.g., "blanc"/"white") but inconsistent with two words (e.g., "manges"/"eat" and "gateau"/"cake"). Thus, fixations to *white car* at "blanc" ("white") would be predicted to be smaller in French, given the greater number of contextual constraints.

[8] Note that the *white car* effects we are testing for are also anticipatory, in that we expect them while sentences are still unfolding. Consistent with this prediction, a large body of research (e.g., Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Sedivy, 2003; Sedivy, Tanenhaus, Chambers, & Carlson, 1999) reveals that listeners rapidly use adjective information to anticipate objects in the VWP. For example, listeners hearing sentences like "Pick up the yellow banana" while viewing a visual display with objects like a *yellow banana*, *yellow notebook*, *green pea*, and *pink eraser* fixated the yellow objects more than the nonyellow objects shortly after hearing "yellow."
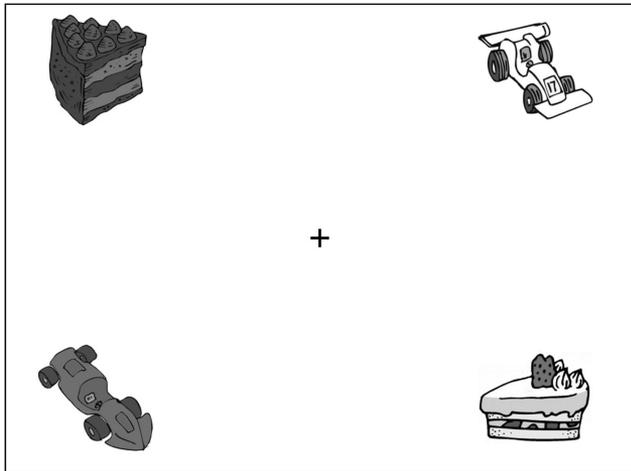
Figure 5. Example visual display from Experiment 1. Participants heard predictive sentences like "The boy will eat the white cake." Visual displays included a target (e.g., *white cake*), which was predicted by the joint contextual constraints from the verb and adjective, and a locally coherent competitor (e.g., *white car*), which fit the lexical constraints of the adjective but was not predicted by the verb selectional restrictions of the sentence context.

following the adjective to the predictable target, and no difference in fixations between the locally coherent competitor and unrelated distractors, consistent with the results of Dahan and Tanenhaus (2004).

Experiment 1

Method.

Participants. Twenty University of Connecticut undergraduates participated for partial course credit. All participants were native speakers of English with self-reported normal hearing and self-reported normal or corrected-to-normal vision.

Materials. We modified 16 predictive sentences from Altmann and Kamide (1999) to include a postverbal adjective (e.g., "The boy will eat the white cake"). Each sentence was associated with a verb-consistent direct object (e.g., cake), a verb-inconsistent competitor (e.g., car), and two color adjectives (e.g., white/brown). The full item list is presented in Appendix B.[9] Each sentence was also associated with a visual display, which included four objects in the corners of the visual display and a fixation cross at the center. The four objects reflected the crossing of verb-consistent/inconsistent objects with the color adjectives (see Figure 5).

A female native speaker of English recorded sentences using both adjectives in a sound-attenuated booth. Recordings were made in Praat with a sampling rate of 44.1 kHz and 16-bit resolution. The speaker was instructed to produce the sentences naturally. We used clip-art images of the direct objects and competitors for our visual stimuli.

Design. We used a 2 × 2 design with verb consistency (consistent and inconsistent) and adjective consistency (consistent and inconsistent) as factors, and with all four conditions represented in each trial. For participants hearing "The boy will eat the white cake," for example, the visual display included a verb-consistent and adjective-consistent *white cake*, a verb-consistent and adjective-inconsistent *brown*

cake, a verb-inconsistent and adjective-consistent *white car*, and a verb-inconsistent and adjective-inconsistent *brown car*. Thus, participants viewed 16 objects of each type across the experiment (i.e., one of each type per visual display). Half the participants heard the other associated color adjective (e.g., "The boy will eat the brown cake"). In this case, the adjective consistency labeling was reversed (e.g., *brown car* was verb-inconsistent and adjective-consistent). The target adjective in sentences, the location of objects in the visual display, and the order of sentences were randomized for each participant. Because objects rotated randomly between adjective-consistent and -inconsistent conditions, differences in looks between the adjective conditions are not likely to be due to differences in saliency.

Procedure. Participants initiated each trial by clicking on a fixation cross in the center of the computer screen. Then the visual display for the trial was presented for a 500-ms preview, and it continued to be displayed as the associated sentence was played over headphones. Participants were told that they would be hearing sentences about people interacting with various things, and they were instructed to use the mouse to click on the object that was involved in each interaction (i.e., the direct object of the sentence). A trial ended when participants clicked on an object in the visual display, at which point it disappeared and was replaced by a blank screen with a fixation cross at the center. The experiment began with four practice trials with feedback, and there were no additional filler trials. We used an ASL R6 remote optics eye tracker sampling at 60 Hz with a head-tracking device (Applied Scientific Laboratories, Bedford, MA) and E-Prime software (Version 1.0; Psychology Software Tools, Inc., Pittsburgh, PA). The experiment was approximately 10 min in length.

Mixed-effects models. Eye movements from Experiments 1 and 2 were analyzed with linear mixed-effects models (lme4 in R). We focused our analyses on the onset of the direct object noun (e.g., "cake"), where we expected listeners' eye movements to reflect information from the verb and adjective, but not direct object noun. Eye movements were submitted to a weighted empirical-logit regression (Barr, 2008a). First, we computed the number of trials with a fixation to each object for each condition for each participant or item (in order to allow for participant and item analyses) at the onset of the direct object noun. Second, we submitted the aggregated data (i.e., across participants or items) to an empirical-logit transformation. Third, the transformed data were submitted to mixed-effects models with fixed effects of condition and random intercepts for participants/items (e.g., participant models used the following syntax: outcome ~ condition + (1| participant)). We did not include random slopes because there was only one data point (i.e., in the aggregated data) for each condition for each participant/item (e.g., see Barr, Levy, Scheep-

---

[9] Although some competitors began with the same phonological segment (e.g., "cake" and "car"), this is generally not considered sufficient phonological overlap to result in robust competition; the standard definition of "cohort" competitors is overlap in the first two segments or in the onset and first vowel (e.g., Magnuson, Dixon, Tanenhaus, & Aslin, 2007). Moreover, the analyses we report below focus on fixations prior to bottom-up information from the noun. Thus, onset competition effects from the noun are neither expected nor capable of contaminating our analyses of primary interest. For the sake of completely ruling out any such possibility, we also report analyses that exclude all items with first-segment overlap (i.e., Items 1, 10, 12, and 15; see Appendices A and B and Footnotes 10 and 11).

ers, & Tily, 2013). We computed p values for our fixed effects by using a model comparison approach: We tested for a reliable improvement in model fit (-2 log-likelihood) against an otherwise identical model that did not include a fixed effect of condition.

Results. The average proportions of fixations to each object type in the visual display are plotted in Figure 6A. Fixations were time locked to the onsets of the verb, determiner, adjective, and noun. The full plotted window extends from the onset of the verb to 200 ms past the offset of the direct object noun. Our critical comparison involves looks to the verb-inconsistent objects (e.g., cars) following the adjective modifier, which are plotted across a zoomed-in window in Figure 6B, to magnify potential differences between conditions.

We submitted eye movements to separate mixed-effects models to test for anticipation, by comparing looks to the verb-consistent objects (e.g., cakes), and for context-independent processing, by comparing looks to the verb-inconsistent objects (e.g., cars). We did not directly compare verb-consistent and -inconsistent objects because objects did not rotate between verb conditions (e.g., cars were never the target), and thus differences could be due to saliency. Both models included a fixed effect of adjective consistency (consistent vs. inconsistent). Below we report means and standard errors by participants for both the proportions of fixation data and (empirical-logit) transformed data.

The analysis of verb-consistent objects (e.g., *white* vs. *brown cake*, when hearing "The boy will eat the white . . .") revealed a reliable effect of adjective consistency in both the participant analysis, estimate = 2.88, $SE = 0.15$, $\chi^2 = 64.48$, $p < .001$, and item analysis, estimate = 2.90, $SE = 0.17$, $\chi^2 = 45.82$, $p < .001$, with reliably more fixations to adjective-consistent objects (proportions of fixations: $M = 0.71$, $SE = 0.03$; transformed: $M = 0.91$, $SE = 0.16$) as compared to adjective–inconsistent objects (proportions of fixations: $M = 0.06$, $SE = 0.01$; transformed: $M = -2.61$, $SE = 0.18$). The analysis of verb-inconsistent objects (e.g., *white* vs. *brown car*, when hearing "The boy will eat the white . . .") revealed a reliable effect of adjective consistency in both the participant analysis, estimate = 0.72, $SE = 0.19$, $\chi^2 = 11.27$, $p < .001$, and item analysis, estimate = 0.78, $SE = 0.23$, $\chi^2 = 8.85$, $p < .01$, with reliably more fixations to adjective-consistent objects (proportions of fixations: $M = 0.06$, $SE = 0.01$; transformed: $M = -2.52$, $SE = 0.16$) as compared to adjective-inconsistent objects (proportions of fixations: $M = 0.02$, $SE = 0.01$; transformed: $M = -3.23$, $SE = 0.12$).[10]

Discussion. Listeners anticipated target objects based on the joint contextual constraints from the verb and adjective (e.g., *white cake* given "The boy will eat the white . . ."). This incremental impact of contextual constraints on anticipatory eye movements is consistent with Simulation 2 and prior findings from the VWP (e.g., Kamide, Altmann, & Haywood, 2003). However, as predicted by Simulation 2, listeners also fixated locally coherent competitors more than unrelated distractors, based on context-independent lexical information from the adjective (e.g., *white car* vs. *brown car* given "The boy will eat the white . . ."). These results reveal that listeners were gradiently sensitive to both context-dependent constraints, based on the sentence context, and context-independent constraints, based on context-independent lexical information, despite the completely constraining context.

However, closer examination of the materials from Experiment 1 raises a potential issue: In our visual displays, the locally

coherent competitor (e.g., *white car*) was always in a "contrast set" with a similar object of a different color (e.g., *brown car*). Sedivy (2003) found that eye movement behaviors in the VWP were modulated by the presence of a contrast set. For example, listeners hearing "Pick up the yellow banana" in the context of a *yellow banana* and *yellow notebook* were equally likely to look to both items during the adjective. However, when another item of the same type as the target was also included (e.g., *brown banana*), which created a contrast set (among the *bananas*), fixations favored the *yellow banana* over the *yellow notebook* in that same adjective window. Sedivy's results suggest that listeners rapidly map adjectives (from the unfolding language) onto specific items in the visual display that are in a contrast set (presumably to disambiguate between them). Consequently, listeners in Experiment 1 may have fixated the locally coherent competitor simply because it happened to always be in a contrast set. To address this concern, locally coherent competitors were not in a contrast set in Experiment 2.

Experiment 2

We tested for bottom-up interference in contexts similar to those in Experiment 1, but in which the verb-inconsistent objects were not in a contrast set. Our predictive contexts involving strong verb selectional restrictions were unchanged, except that we replaced the verb- and adjective-inconsistent objects (e.g., *brown car*) with a completely unrelated distractor (e.g., *pink train*). Again, Simulation 2 predicts more looks to the locally coherent competitor (e.g., *white car*) with adjective-consistent sentences ("The boy will eat the white . . .") as compared to adjective-inconsistent sentences ("The boy will eat the brown . . ."), even in the absence of a contrast set.

Method.

Participants. Thirty University of Connecticut undergraduates participated for course credit. All participants were native speakers of English with self-reported normal hearing and self-reported normal or corrected-to-normal vision who had not participated in Experiment 1.

Materials. We used materials identical to those used in Experiment 1, except that we replaced one of the verb-inconsistent objects (e.g., *brown car*) in each visual display with a completely unrelated distractor (e.g., *pink train*), which was unrelated to the verb and the two adjectives associated with each sentence (see Figure 7). Sentences were recorded as in Experiment 1, except that they were recorded by a male native speaker of English. The full item list is presented in Appendix C.

Design. We used a one-factor design of adjective consistency (consistent and inconsistent). In the adjective-consistent condition, the locally coherent competitor (e.g., *white car*) in the visual display was consistent with the adjective modifier of the direct object in the sentence (e.g., "The boy will eat the white . . ."). In the adjective-inconsistent condition, this same competitor was inconsistent with the adjective modifier of the direct object (e.g., "The boy will eat the brown . . ."). In both conditions, visual

---

[10] Exclusion of cohort competitors revealed a similar pattern among verb-inconsistent objects: a reliable effect of adjective consistency in both the participant analysis, estimate = 0.73, $SE = 0.19$, $\chi^2 = 14.16$, $p < .001$, and item analysis, estimate = 0.86, $SE = 0.30$, $\chi^2 = 6.31$, $p < .05$.
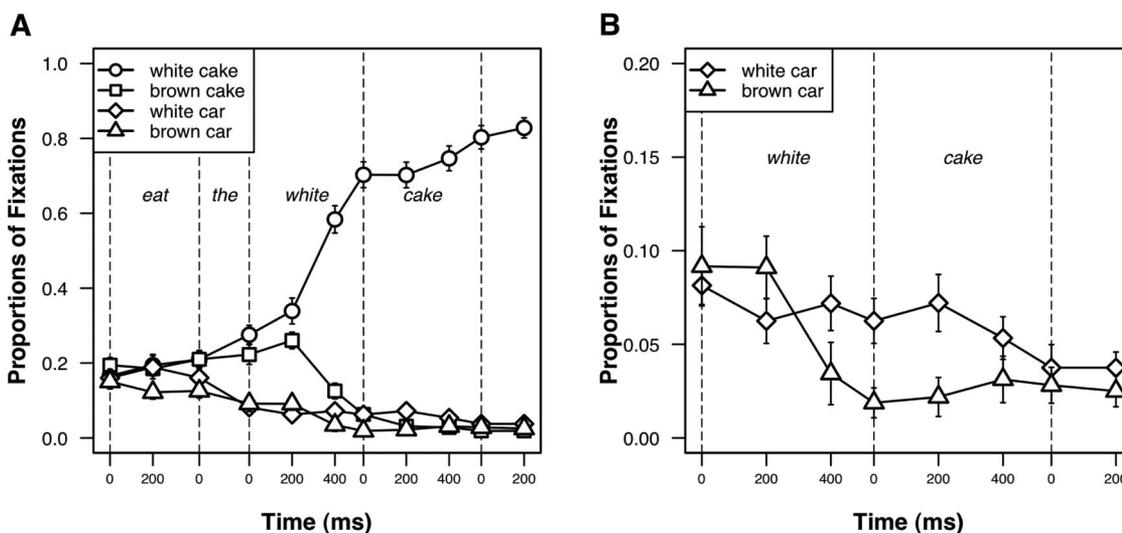
Figure 6. Experiment 1: Average (standard error) proportions of fixations to the objects in the visual display between mean verb onset and mean direct object noun offset (A), and to the verb-inconsistent objects in a zoomed-in window (proportions plotted to .20) between adjective onset and mean direct object noun offset (B), for the example sentence "The boy will eat the white cake." Fixations were resynchronized at the onset of each word and extend to the average offset of each word.

displays also included two verb-consistent objects of the same type, each consistent with one of the associated adjectives (e.g., *white cake* and *brown cake*), and a completely unrelated distractor (e.g., *pink train*), which was unrelated to both the verb and the associated adjectives. In our analyses, we compared looks to the same object (e.g., *white car*) in two language contexts, so differences in looks between the contexts cannot be attributed to differences in saliency (this differs from the design in Experiment 1, where saliency was controlled by rotating a given object across, e.g., locally coherent competitor and unrelated distractor conditions). We created two lists to rotate objects across the adjective-

Figure 7. Example visual display from Experiment 2. Participants heard predictive sentences like "The boy will eat the white cake." Unlike in Experiment 1, the unrelated distractor (e.g., *pink train*) was not in a competitor set with the locally coherent competitor (e.g., *white car*).

consistent and -inconsistent conditions, with half of each type on each list. Participants were randomly assigned to a list, and the location of objects in the visual display and the order of sentences was randomized for each participant.

We created 16 additional filler items to counterbalance biases related to the direct object target, which was always paired with a like object (e.g., *white cake* and *brown cake*) in critical trials. In filler trials, the visual display contained two objects of the same type and two unpaired objects, and the target of the trial was always one of the unpaired objects (the target was also predictable, as the only object related to the verb in the sentence, mirroring the critical conditions).

Procedure. Procedures were identical to those in Experiment 1, except that the experiment was slightly longer (approximately 15 min instead of 10 min, due to the additional filler items).

Results. The average proportions of fixations to each object type in the visual display are plotted in Figure 8A. Our critical comparison involves looks to the locally coherent competitor (e.g., *white car*) with adjective consistent (e.g., "white") versus inconsistent (e.g., "brown") sentences, which is plotted across a zoomed-in window in Figure 8B. We did not directly compare the verb-inconsistent objects (e.g., *white car* vs. *pink train*), as in Experiment 1, because objects did not rotate across these conditions, and differences could thus be due to saliency. Rather, to control for effects of saliency, we compared looks to the same object (e.g., *white car*) in the two language contexts (e.g., "white" vs. "brown").

As in Experiment 1, we focused our analyses on the onset of the direct object noun (e.g., "cake"), and we submitted eye movements to a weighted empirical-logit regression (Barr, 2008a). Again, our linear mixed-effects models included fixed effects of condition and random intercepts for participant/items. The analysis of verb-consistent objects (e.g., *white* vs. *brown cake*, when hearing "The
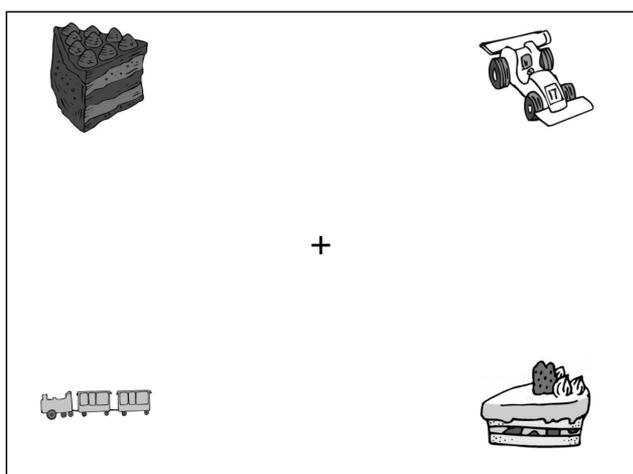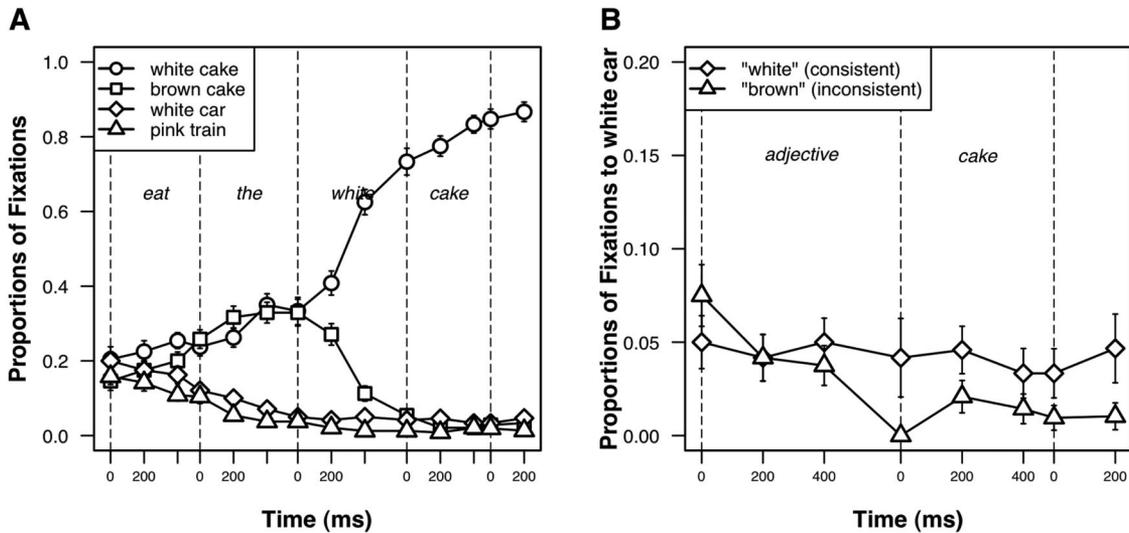
Figure 8. Experiment 2: Average (standard error) proportions of fixations to the objects in the visual display between mean verb onset and mean direct object noun offset with adjective-consistent sentences (A), and to the locally coherent adjective competitor object in the adjective-consistent versus adjective-inconsistent condition in a zoomed-in window (proportions plotted to .20) between adjective onset and mean direct object noun offset (B). In the adjective-consistent condition, the locally coherent competitor (e.g., *white car*) was consistent with the adjective in the sentence (e.g., "The boy will eat the white cake"), whereas in the adjective-inconsistent condition, this competitor was inconsistent with the adjective in the sentence (e.g., "The boy will eat the brown cake"). Fixations were resynchronized at the onset of each word and extend to the average offset of each word.

boy will eat the white . . .") revealed a reliable effect of adjective consistency in both the participant analysis, estimate $= 2.66$, $SE = 0.17$, $\chi^2 = 84.37$, $p < .001$, and item analysis, estimate $= 2.85$, $SE = 0.22$, $\chi^2 = 41.62$, $p < .001$, with reliably more fixations to adjective-consistent objects (proportions of fixations: $M = 0.72$, $SE = 0.03$; transformed: $M = 0.93$, $SE = 0.15$) as compared to adjective–inconsistent objects (proportions of fixations: $M = 0.05$, $SE = 0.02$; transformed: $M = -2.36$, $SE = 0.13$). The analysis of the locally coherent competitor (e.g., *white car*) with adjective consistent (e.g., "white") versus inconsistent (e.g., "brown") sentences revealed a reliable effect of adjective consistency in both the participant analysis, estimate $= 0.42$, $SE = 0.17$, $\chi^2 = 9.60$, $p < .01$, and item analysis, estimate $= 0.67$, $SE = 0.24$, $\chi^2 = 8.53$, $p < .01$, with reliably more fixations to the locally coherent competitor with adjective consistent sentences (proportions of fixations: $M = 0.04$, $SE = 0.02$; transformed: $M = -2.53$, $SE = 0.13$) as compared adjective inconsistent sentences (proportions of fixations: $M = 0.01$, $SE = 0.01$; transformed: $M = -2.75$, $SE = 0.06$).[11]

Discussion. Listeners again anticipated target objects based on the joint contextual constraints from the verb and adjective (e.g., *white cake* given "The boy will eat the white . . ."). This incremental impact of contextual constraints on anticipatory eye movements is consistent with that in Experiment 1, Simulation 2, and prior findings from the VWP (e.g., Kamide, Altmann, & Haywood, 2003). However, as predicted by Simulation 2, listeners again also fixated locally coherent competitors more with consistent rather than inconsistent adjectives, based on context-independent lexical information from the adjective (e.g., *white car* given "The boy will eat the white . . ." vs. "The boy will eat the

brown . . ."). These results reveal that listeners were gradiently sensitive to both context-dependent constraints, based on the sentence context, and context-independent constraints, based on context-independent lexical information, despite the completely constraining context and the lack of a contrast set.

One potential alternative explanation of the *white car* effect is that listeners were using low-level visual cues about color from their periphery. Consequently, they may have erroneously launched eye movements to locally coherent competitors because they had sufficient visual acuity to know they were white, but they did not have sufficient visual acuity to know whether they were edible. In part, the rapid, massive reduction in fixations to cars during "eat" does suggest that listeners had good visual acuity in general (i.e., to know that the cars were cars and thus inedible). Moreover, the current results are also consistent with a number of closely related findings (e.g., Borovsky, Elman, & Fernald, 2012; Kamide, Altmann, & Haywood, 2003; Kukona et al., 2011), which do not depend on low-level visual cues to color. For example, Kukona et al. (2011) found that listeners fixated *policeman* (a locally coherent competitor like *white car*) more than the distractors *gardener* and *surfer* when hearing "Toby will arrest the," although all of these objects are quite visually distinct from *crook*, the predictable direct object.

---

[11] Exclusion of cohort competitors revealed a similar pattern among verb-inconsistent objects: a reliable effect of adjective consistency in both the participant analysis, estimate $= 0.32$, $SE = 0.15$, $\chi^2 = 8.32$, $p < .01$, and item analysis, estimate $= 0.64$, $SE = 0.30$, $\chi^2 = 5.19$, $p < .05$.

General Discussion

A growing body of research reveals that language users rapidly form linguistic representations that best satisfy the contextual constraints provided by the sentence, discourse, and visual context, and that they also activate representations that are outside the scope of these constraints. While anticipatory effects (e.g., Altmann & Kamide, 1999; Chambers & San Juan, 2008; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Knoeferle & Crocker, 2006, 2007) reveal that the language processing system rapidly uses the contextual constraints to predict upcoming linguistic input, effects of context-independent information reveal that the system simultaneously activates representations that are not consistent with the context (e.g., Allopenna et al., 1998; Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005; Kukona et al., 2011; Myung et al., 2006; Swinney, 1979; Tabor et al., 2004; Tanenhaus et al., 1979; Yee & Sedivy, 2006). In two simulations, we showed that the self-organizing neural network framework predicts classic results showing both bottom-up interference from context-independent constraints (e.g., Swinney, 1979; Tanenhaus et al., 1979) and rapid context integration (e.g., Altmann & Kamide, 1999). In two experiments, we also found evidence for bottom-up interference even in highly constraining anticipatory VWP contexts involving strong verb selection restrictions, as predicted by the simulations. For example, listeners hearing "The boy will eat the white . . ." predominately looked to targets (e.g., *white cake*) predicted by the joint contextual constraints from the verb and adjective (consistent with, e.g., Kamide, Altmann, & Haywood, 2003). However, they also made looks following the adjective to locally coherent competitors (e.g., *white car* vs. unrelated distractors). These results suggest that listeners were sensitive to the fit between "white" and the *white car*, although this object was not predicted by the verb selectional restrictions of the sentence context.

Note that our self-organizing neural networks do not have a grammar in the form of a system of rules that operate on symbolic objects. Rather, the networks' interacting units encode "grammatical constraints," which the network learns through the interactions of its units with the language. Critically, these grammatical constraints, which shape how the units interact, do not enforce total grammatical coherence, thus predicting the activation of the verb sense of "rose" in a noun context in Simulation 1, and fixations to *white car* following "Eat [the] white . . ." in Simulation 2. By contrast, our experimental findings provide evidence against approaches that *do* use a grammar to enforce global coherence by only allowing for the activation of structures that fully satisfy the global constraints (we describe Bayesian models that make this assumption below; e.g., Hale, 2001; Levy, 2008a).

Moreover, the pattern of results we observed over the course of training in the self-organizing neural network (see Figure 4) also provides rich predictions about both human language development and individual differences in language skill. In Simulation 2, we found that activation of the locally coherent competitor (e.g., *white car*) was largest early in training and diminished over the course of training. One prediction that follows from this pattern, given the assumption that the amount of training in the network maps onto *age*, is that bottom-up interference will be largest early in development (e.g., in children) and will diminish over development (e.g., in adults). Another prediction, given the assumption that the amount of error in the network maps (inversely) onto language *skill*, is that bottom-up interference will be larger in less skilled language users (i.e., with higher error) as compared to more skilled language users (i.e., with lower error).

Both of these predictions were supported by recent experimental findings by Borovsky et al. (2012). They found that both adults and children hearing "The pirate hides the . . ." while viewing a visual display with a *treasure chest*, *ship*, *dog bone*, and *cat* anticipatorily fixated the *treasure chest*, which was predicted by the combined contextual constraints from the subject and verb (compatible with Kamide, Altmann, & Haywood, 2003). However, their visual display also included a locally coherent competitor: *dog bone* was predicted by the verb "hides," but not the subject "pirate." They observed a numerically larger interference effect (as reflected in the length of the time window in which there were reliably more fixations to *dog bone* vs. distractors like *cat*) in children than adults, and in language users with lower vocabulary skill as compared to higher vocabulary skill (although they did not directly compare these groups; bottom-up interference was not a central focus of their study). These findings further suggest that our self-organizing predictions should not be limited to the behavior of the network at asymptote (i.e., where it predicts no *white car* vs. *brown car* difference).

Relation to Previous Findings

Our experimental findings are consistent with the results of Kukona et al. (2011), who found that listeners hearing a sentence like "Toby arrests the . . ." anticipatorily fixated a predictable (i.e., based on the contextual constraints) patient of the verb (e.g., *crook*). However, listeners were approximately as likely to make anticipatory fixations to a good agent of the verb (e.g., *policeman*), based on the thematic fit between the verb and agent, despite the fact that the agent role was already filled. In a second experiment with passive sentences (e.g., "Toby was arrested by the . . ."), in which more time elapsed between the verb and direct object noun and more syntactic cues were available (e.g., "was . . ."), Kukona et al. found even clearer evidence for graded influences, with more anticipatory fixations to the *policeman* (e.g., the predictable direct object of the passive sentence) than the *crook*, but also more fixations to the *crook* than the unrelated distractors. These gradient effects with *policeman* and *crook* closely parallel our results with *white cake* and *white car*, suggesting that there is a balance between top-down context effects and bottom-up interference effects that varies with the strength of the two sources of information.

In addition, we have implied throughout that the activation of representations that are at odds with contextual constraints (i.e., outcomes with zero probability in a sentence context) represents a kind of interference that impedes processing (e.g., by shifting listeners' attention away from target objects and toward nontarget objects). However, there may also be facilitory side effects of this activation, as emphasized in Kukona et al. (2011). For example, the activation of *policeman* when hearing "Toby will arrest the" may facilitate discourse processing because situations that involve "arrest" often also involve policemen (although, as we have suggested, this activation is highly transient). Indeed, activating representations that are not explicitly referred to linguistically, but which are associated or related to the linguistic input, may also

facilitate integration with prior knowledge and subsequent memory. On the other hand, the current results suggest that not all bottom-up interference is necessarily facilitative (i.e., it is not clear that eye movements to *white car* similarly facilitate discourse processing).

Our results are also consistent with evidence for "verb effects" in Kamide, Altmann, and Haywood (2003; see also Borovsky et al., 2012). Again, they found that listeners hearing "The girl will ride the . . ." anticipatorily fixated a *carousel* more than a *motorbike*, based on the contextual constraints. However, they also found that listeners anticipatorily fixated the subject-conflicting *motorbike* more during "the" when hearing "The girl will ride the . . ." as compared to "The girl will taste the . . ." These results suggest that listeners were sensitive to the thematic fit between "ride" and *motorbike*, and that they were activating this object even though it conflicted with the context, consistent with both Kukona et al. (2011) and our current results.

However, our findings extend the results of Borovsky et al. (2012); Kamide, Altmann, and Haywood (2003); and Kukona et al. (2011), as well as other evidence for context-independent processing in language, in a number of important ways. The current results reveal that these effects compete against strongly constraining verb selectional restrictions, which have previously been shown to drive robust anticipatory effects (e.g., Altmann & Kamide, 1999). For example, our sentence contexts involved a constraining verb (e.g., "The boy will eat the . . ."), whereas the preceding sentence contexts in Borovsky et al.; Kamide, Altmann, and Haywood; and Kukona et al. involved only a weakly constraining noun phrase (e.g., "The girl . . ." or "Toby . . ."). As well, the effects in the current study were based on adjectival information (e.g., linguistic–visual match that runs counter to the sentence context), rather than thematic association.

Finally, additional findings on ambiguity resolution in the VWP also provide a means for integrating the results of Tanenhaus et al. (1995) and Dahan and Tanenhaus (2004), which we described in the introduction. Recall that Tanenhaus et al. demonstrated that contextual constraints from the visual context (e.g., two referents) continuously and immediately constrained the interpretation of syntactic ambiguities (e.g., "Put the apple on the towel in the box"). However, Snedeker and Trueswell (2004; see also Novick, Thompson-Schill, & Trueswell, 2008) showed that these contextual effects can be modulated by lexical biases in adult listeners. They compared ambiguous sentences like "Choose/Tickle the cow with the stick": Whereas the verb "Choose" favors a modifier interpretation (e.g., Choose the cow that is holding the stick), the verb "Tickle" favors an instrument interpretation (e.g., Tickle the cow by using the stick). In line with Tanenhaus et al., they found a reliable effect of visual context: Two-referent contexts (e.g., a visual display with a cow holding a stick and a cow holding nothing) favored modifier interpretations relative to one-referent contexts. However, they also found a reliable effect of verb bias: Instrument-biased verbs like "Tickle" favored instrument interpretations relative to modifier-biased verbs like "Choose."

Similarly, recall that Dahan and Tanenhaus (2004) demonstrated that contextual constraints from a sentence context (e.g., "Nog nooit klom een bok . . ."/"Never before climbed a goat . . .") continuously and immediately constrained spoken word recognition (e.g., eliminating cohort effects from competitors like "bot"/"bone"). The self-organization approach, in keeping with

constraint-based frameworks in general, predicts that such effects should show evidence of being graded. Indeed, in a second experiment in which they cross-spliced the onset and vowel of a cohort competitor (e.g., "bo" from "bot") with the offset of the target noun (e.g., "k" from "bok"), they found more cohort fixations. Although these results suggest that contextual constraints can overwhelm bottom-up constraints, they also reveal that context-independent processing effects can be made more pronounced by shifting the relative balance of the constraints (e.g., by using verbs that bias context-inconsistent structures, Snedeker & Trueswell, 2004; or by providing more bottom-up support for context-inconsistent competitors via cross-splicing, Dahan & Tanenhaus, 2004).

Bayesian Belief Update

Bayesian models of sentence processing (e.g., Bicknell & Levy, 2010; Levy, 2008b; Narayanan, & Jurafsky, 2002; see also closely related probabilistic models: e.g., Hale, 2001; Jurafsky, 1996; Levy, 2008a) have also been widely used to model human sentence processing. Recent Bayesian implementations have assumed that language users have a probabilistic model of the way that words are sequenced in their language, such that during online processing, language users distribute probabilities over grammatical structures that are consistent with the observed input. Crucially, this probability distribution also allows language users to assign probabilities to future events (i.e., structures that are likely to follow from the current input). This approach is called "Bayesian belief update" because Bayes' rule specifies the way that each new piece of information revises current expectations.

A number of recent proposals (e.g., Hale, 2001; Levy, 2008a) have assumed that contextual information places very strong constraints on processing: Language users are assumed to be "optimal" in the sense that they only distribute probabilities over structures that are fully consistent with the context. These proposals very naturally predict findings showing rapid context integration (see Rapid, Immediate Context Integration). However, they also predict that language users should not show bottom-up interference effects (see Bottom-Up Interference). For example, listeners in the current experiments would be predicted to distribute probabilities over cakes only, which satisfy the selectional restrictions of "eat," and not cars, thus predicting no *white car* versus *brown car* difference. Thus, the current findings provide strong evidence against this class of models.

However, more recent proposals (e.g., Bicknell & Levy, 2010; Gibson, Bergen, & Piantadosi, 2013; Levy, 2008b; for a closely related proposal concerned with spoken word recognition, see Norris & McQueen, 2008) have assumed that language users also have "uncertainty" about contextual information. Consequently, these proposals allow for probabilities to be distributed over alternative grammatical structures that are not fully consistent with the context (i.e., as specified by the bottom-up input). For example, Levy, Bicknell, Slattery, and Rayner (2009) suggested that the local coherence effects observed by Tabor et al. (2004) are a consequence of language users' uncertainty about the sentence context "The coach smiled at . . ." They suggested that when readers encounter "tossed" in "The coach smiled at the player tossed the Frisbee," they entertain alternative interpretations of the prior words in the sentence (e.g., instead of "at," they read "and,"

which is orthographically similar) due to uncertainty about this input (crucially, "and" also allows for the more predictable sentence "The coach smiled and the player tossed the Frisbee"). Indeed, such an approach offers a very clear alternative to self-organization: Rather than assume that language users activate structures that conflict with the context (which is implicitly assumed to have no uncertainty) due to the self-organizing nature of structure formation, such an approach assumes that context-conflicting structures are activated because contextual information itself is uncertain. Nevertheless, Levy et al.'s proposal relies on the fact that the sentence context in Tabor et al. is (orthographically) confusable with other sentence contexts that do allow for the locally coherent parse. By contrast, it is not apparent that there are also phonologically similar parses in the case of Swinney (1979) and Tanenhaus et al. (1979; i.e., Simulation 1), or Experiments 1 and 2 (i.e., Simulation 2), that would likewise lead language users to expect the respective locally coherent structures.

Relation to Other Theories of Language Processing

Our experimental results also bear on a number of other theories of sentence processing that are closely related to self-organization. Ferreira and colleagues (e.g., Ferreira & Patson, 2007) proposed the Good-Enough theory of sentence processing that, like self-organization, assumes that contextual coherence of linguistic representations is not fully maintained. According to this theory, linguistic representations or structures may be neither accurate nor detailed; rather, representations are often cobbled together, using quick and dirty heuristics, so as to be "good enough" for the task at hand. In keeping with Good-Enough theory, they have shown that readers misinterpret garden path sentences like "While Mary bathed the baby played in the crib" to mean both that Mary bathed the baby and that Mary bathed herself (although only the latter is grammatically accurate; Christianson, Hollingworth, Halliwell, & Ferreira, 2001). Our results are generally consistent with good-enough theory's claims that structures can be cobbled together despite their contextual incoherence. However, our results suggest that such heuristics may even be used for straightforward, routine sentences, which presumably place very little burden on the language processing system (e.g., relative to NP-Z garden path sentences), consistent with evidence for closely related misinterpretation effects in passive sentences (e.g., Ferreira, 2003). Potentially, our results suggest that listeners may have a heuristic that leads them to rely on information within the noun phrase (e.g., adjective information), rather than outside it (e.g., verb information), in anticipating nouns. However, in our view these effects are not due to the system literally employing heuristics, but instead are emergent effects of a self-organizing system in which structures that are not fully coherent or compatible with the global context can become strongly activated.

Our results are also consistent with the feature-based approach of Altmann and Kamide (2007). They describe a linking hypothesis for language-mediated eye movements that accounts for findings showing that listeners fixate items in a visual context that are variously related to an unfolding utterance (e.g., Allopenna et al., 1998; Altmann & Kamide, 1999; Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005; Myung et al., 2006; Yee & Sedivy, 2006). According to their approach, listeners activate representations in a "mental world" that correspond to input from both language (e.g., words) and the visual world (e.g., objects within a visual display to which the discourse refers). Critically, these representations are featural in nature, and they include information like phonological form, verb selectional restrictions, physical form (visual, tactile, etc.), category membership, function, and so forth. Visual representations receive an activation boost when they share features with the linguistic input, which increases the likelihood of a fixation to the relevant item. Very much in this line, the behavior of the self-organizing neural network in Simulation 2 appears to similarly reflect the influence of verb- and adjective-based featural-type (e.g., edible/rideable; white/brown) information.

In summary, we have presented new evidence for bottom-up interference in highly constraining anticipatory contexts in the VWP. These results, in conjunction with prior findings, suggest that the language systems acts on, and maintains, transient information structures at many levels of language structure that are outside the scope of, or in conflict with, the contextual constraints. Critically, these phenomena are hallmarks of the self-organizing neural network framework. Moreover, our simulations, which address the integration of incoming lexical information (i.e., an incoming word) with sentence context information (i.e., from preceding words in an unfolding utterance), explicitly show that the framework predicts a range of findings that involve gradient effects of both context-dependent and context-independent information.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439. doi:10.1006/jmla.1997.2558

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

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502–518. doi:10.1016/j.jml.2006.12.004

Barr, D. J. (2008a). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59,* 457–474. doi:10.1016/j.jml.2007.09.002

Barr, D. J. (2008b). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109,* 18–40. doi:10.1016/j.cognition.2008.07.005

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. doi:10.1016/j.jml.2012.11.001

Bicknell, K., & Levy, R. (2010). A rational model of eye movement control in reading. In J. Hajic, S. Carberry, & S. Clark (Eds.), *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 1168–1178). Stroudsburg, PA: Association for Computational Linguistics.

Bicknell, K., Levy, R., & Demberg, V. (2010). Correcting the incorrect: Local coherence effects modeled with prior belief update. In *Proceedings of the 35th Annual Meeting of the Berkeley Linguistics Society* (pp. 13–24). Berkeley, CA: Berkeley Linguistics Society.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112,* 417–436. doi:10.1016/j.jecp.2012.01.005

Chambers, C. G., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition, 108,* 26–50. doi:10.1016/j.cognition.2007.12.009

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42,* 368–407. doi:10.1006/cogp.2001.0752

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84–107. doi:10.1016/0010-0285(74)90005-X

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Evidence from immediate effects of verb-based constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 498–513. doi:10.1037/0278-7393.30.2.498

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12,* 453–459. doi:10.3758/BF03193787

Dell, G. S., & Chang, F. (in press). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences.*

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24,* 409–436. doi:10.1007/BF02143160

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211. doi:10.1207/s15516709cog1402_1

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7,* 195–225. doi:10.1007/BF00114844

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47,* 164–203. doi:10.1016/S0010-0285(03)00005-7

Ferreira, F., & Patson, N. D. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass, 1,* 71–83. doi:10.1111/j.1749-818X.2007.00007.x

Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science, 37,* 578–591. doi:10.1111/cogs.12022

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology.* Cambridge, MA: MIT Press.

Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Inquiry, 29,* 1–36. doi:10.1162/002438998553644

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 8051–8056. doi:10.1073/pnas.1216438110

Gordon, D. M. (2010). *Ant encounters: Interaction networks and colony behavior.* Princeton, NJ: Princeton University Press.

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 of Computational Linguistics* (Vol. 2, pp. 159–166). Stroudsburg, PA: Association for Computational Linguistics.

Hölldobler, B., & Wilson, E. O. (2008). *The superorganism: The beauty, elegance, and strangeness of insect societies.* New York, NY: Norton.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96,* B23–B32. doi:10.1016/j.cognition.2004.10.003

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. doi:10.1016/j.cognition.2012.10.013

Jensen, H. J. (1998). *Self-organized criticality: Emergent complex behavior in physical and biological systems.* New York, NY: Cambridge University Press. doi:10.1017/CBO9780511622717

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137–194. doi:10.1207/s15516709cog2002_1

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156. doi:10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32,* 37–55. doi:10.1023/A:1021933015362

Keller, E. F., & Segel, L. A. (1970). Conflict between positive and negative feedback as an explanation for the initiation of aggregation in slime mould amoebae. *Nature, 227,* 1365–1366. doi:10.1038/2271365a0

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30,* 481–529. doi:10.1207/s15516709cog0000_65

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language, 57,* 519–543. doi:10.1016/j.jml.2007.01.003

Konieczny, L., Müller, D., Hachmann, W., Schwarzkopf, S., & Wolfer, S. (2009). Local syntactic coherence interpretation: Evidence from a visual world study. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp. 1133–1138). Austin, TX: Cognitive Science Society.

Konieczny, L., Weldle, H., Wolfer, S., Müller, D., & Baumann, P. (2010). Anaphora and local coherences. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1204–1209). Austin, TX: Cognitive Science Society.

Kukona, A., Fang, S.-Y, Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition, 119,* 23–42. doi:10.1016/j.cognition.2010.12.002

Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science, 35,* 1009–1051. doi:10.1111/j.1551-6709.2011.01180.x

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

Levy, R. (2008b). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the 13th Conference on Empirical Methods in Natural Language Processing* (pp. 234–243). Stroudsburg, PA: Association for Computational Linguistics.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 21086–21090. doi:10.1073/pnas.0907664106

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703. doi:10.1037/0033-295X.101.4.676

Magnuson, J. S., Dixon, J., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31,* 133–156. doi:10.1080/03640210709336987

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2000). Simple recurrent networks and competition effects in spoken word recognition. *University of Rochester Working Papers in the Language Sciences, 1,* 56–71.

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition, 108,* 866–873. doi:10.1016/j.cognition.2008.06.005

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word recognition and learning: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132,* 202–227. doi:10.1037/0096-3445.132.2.202

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

Marée, A. F. M., & Hogeweg, P. (2001). How amoeboids self-organize into a fruiting body: Multicellular coordination in *Dictyostelium discoideum*. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 3879–3883. doi:10.1073/pnas.061535198

McClelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech Perception. *Cognitive Psychology, 18,* 1–86. doi:10.1016/0010-0285(86)90015-0

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407. doi:10.1037/0033-295X.88.5.375

Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. Cambridge, MA: MIT Press.

Myung, J., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition, 98,* 223–243. doi:10.1016/j.cognition.2004.11.010

Narayanan, S., & Jurafsky, D. (2002). A Bayesian model predicts human parse preference and reading times in sentence processing. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems* (Vol. 14, pp. 59–65). Cambridge, MA: MIT Press.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395. doi:10.1037/0033-295X.115.2.357

Novick, J. M., Thompson-Schill, S., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition, 107,* 850–903. doi:10.1016/j.cognition.2007.12.011

Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21,* 25–34. doi:10.1145/37402.37406

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32,* 3–23. doi:10.1023/A:1021928914454

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109–147. doi:10.1016/S0010-0277(99)00025-6

Shillcock, R., & Bard, E. G. (1993). Modularity and the processing of closed-class words. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 163–185). Mahwah, NJ: Erlbaum.

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

Solé, R. V., & Bascompte, J. (2006). *Self-organization in complex ecosystems*. Princeton, NJ: Princeton University Press.

Spivey, M. J. (2007). *The continuity of mind*. New York, NY: Oxford University Press.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1521–1543. doi:10.1037/0278-7393.24.6.1521

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–659. doi:10.1016/S0022-5371(79)90355-4

Tabor, W. (2000). Fractal encoding of context-free grammars in connectionist networks. *Expert Systems: The International Journal of Knowledge Engineering and Neural Networks, 17,* 41–56. doi:10.1111/1468-0394.00126

Tabor, W., Cho, P. W., & Szkudlarek, E. (2013). Fractal analysis illuminates the form of connectionist structural gradualness. *Topics in Cognitive Science, 5,* 634–667. doi:10.1111/tops.12036

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50,* 355–370. doi:10.1016/j.jml.2004.01.001

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 431–450. doi:10.1037/0278-7393.30.2.431

Tanenhaus, M. K., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427–440. doi:10.1016/S0022-5371(79)90237-8

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. doi:10.1126/science.7777863

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives in sentence processing* (pp. 155–179). Hillsdale, NJ: Erlbaum.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition, 75,* 105–143. doi:10.1016/S0010-0277(00)00063-9

Yee, E., & Sedivy, J. C. (2006). Eye movements reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1–14. doi:10.1037/0278-7393.32.1.1

(*Appendices follow*)

Appendix A

Grammar Used to Generate the Training Sentences in Simulation 1

1.0 S → NP VP
.5 NP → Det N
.5 NP → Det Adj N
.5 VP → V
.5 VP → Adv V

Adjective (Adj): pretty, white
Adverb (Adv): slowly, quickly
Determiner (Det): the, your
Noun (N): dandelion, *rose*, tulip
Verb (V): grew, *rose*, wilted

Note. Within each syntactic class (e.g., Adj, Adv, Det, N, V), lexical items were used with equal probabilities. S = sentence; NP = noun phrase; VP = verb phrase.

Appendix B

Materials Used in Experiment 1

Item	Sentence	Competitor
1	The boy will eat the brown/white cake.	car
2	The woman will drink the white/yellow beverage.	cheese
3	The officer will arrest the green/purple man.	house
4	The woman will bathe the red/yellow puppy.	stool
5	The boy will bounce the blue/red ball.	plane
6	The man will climb the brown/gray ladder.	bunny
7	The housewife will fry the brown/gray mushroom.	knife
8	The doctor will inject the black/orange cat.	microscope
9	The woman will play the black/brown piano.	table
10	The woman will read the green/purple book.	bag
11	The man will repair the gray/green washer.	trash
12	The girl will ring the purple/red bell.	bricks
13	The man will sail the red/yellow boat.	bird
14	The man will smoke the blue/orange pipe.	glasses
15	The boy will walk the brown/white dog.	duck
16	The businessman will wear the black/green hat.	chair

Note. Sentences were closely based on materials from Altmann and Kamide (1999). Sentences always had a predictable direct object target and one of two possible color adjectives. Visual displays included four objects, reflecting the crossing of the target (e.g., cake) and competitor (e.g., car) objects with the color adjectives (e.g., brown/white).

(*Appendices continue*)

Appendix C

Materials Used in Experiment 2

Item	Competitor, unrelated distractor
1	white car, pink train
2	yellow cheese, brown plant
3	purple house, black broom
4	yellow stool, blue rocker
5	red plane, orange bike
6	gray bunny, yellow cactus
7	brown knife, pink scale
8	orange microscope, gray TV
9	black table, orange phone
10	green bag, blue cup
11	gray trash, pink mirror
12	purple bricks, black drum
13	red bird, white sun
14	blue glasses, green hen
15	brown duck, gray clock
16	green chair, yellow magnifier

Note. Sentences were identical to those in Experiment 1. Visual displays included four objects, including two objects reflecting the crossing of the target (e.g., cake) with the color adjectives (e.g., brown/white). However, the visual display also included an object that was consistent with one of the color adjectives (e.g., white car) as well as an unrelated distractor (e.g., pink train).