RESEARCH

Two minds are not always better than one: Modeling evidence for a single sentence analyzer

Dan Parker
College of William and Mary, Williamsburg, VA, US
dparker@wm.edu

A challenge for grammatical theories and models of language processing alike is to explain conflicting online and offline judgments about the acceptability of sentences. A prominent example of the online/offline mismatch involves "agreement attraction" in sentences like *The key to the cabinets were rusty*, which are often erroneously treated as acceptable in time-restricted "online" measures, but judged as less acceptable in untimed "offline" tasks. The prevailing assumption is that online/offline mismatches are the product of two linguistic analyzers: one analyzer for rapid communication (the "parser") and another, slower analyzer that classifies grammaticality (the "grammar"). A competing hypothesis states that online/offline mismatches reflect a single linguistic analyzer implemented in a noisy memory architecture that creates the opportunity for errors and conflicting judgments at different points in time. A challenge for the single-analyzer account is to explain why online and offline tasks sometimes yield conflicting responses if they are mediated by the same analyzer. The current study addresses this challenge by showing how agreement attraction effects might come and go over time in a single-analyzer architecture. Experiments 1 and 2 use an agreement attraction paradigm to directly compare online and offline judgments, and confirm that the online/offline contrast reflects the time restriction in online tasks. Experiment 3 then uses computational modeling to capture the mapping from online to offline responses as a process of sequential memory sampling in a single-analyzer framework. This demonstration provides some proof-of-concept for the single-analyzer account and offers an explicit process model for the mapping between online and offline responses.

Keywords: Grammar-parser distinction; cognitive architecture of language; agreement attraction; working memory; acceptability judgments; computational modeling

1 Introduction

A long-standing puzzle for theories of language concerns the relationship between "online" and "offline" judgments about the acceptability of sentences. Online and offline data are distinguished by the time sensitivity of the response: offline judgments are elicited with no time restrictions following presentation of the complete sentence, whereas online responses are elicited with time-restricted measures, usually in the middle of the sentence or in a short time window at the end of the sentence (see Lewis & Phillips 2015, for discussion).[1] Historically, linguists have focused on offline data to develop their grammatical theories, and psycholinguists have focused on online data as the basis of their process models. However, there has been little work to date to reconcile the claims based on these different types of data. The current study seeks to address parts of this gap.

---

[1] The term "online" is also often used to refer to linguistic processes that occur relatively quickly, but the focus of the current study will be on when the response is elicited, using the terms "online" and "offline" informally to refer to restricted and unrestricted time windows respectively.

A starting point to unite theories of online and offline data are cases where online and offline data actually diverge. There are numerous cases of close alignment between online and offline data (see Lewis & Phillips 2015, for a recent review), but there are also a handful of misalignments that have been presented as critical evidence for a dualistic architecture of the human linguistic system. One such type of misalignment that has received much attention recently involves so-called "linguistic illusions", where comprehenders temporarily accept ill-formed sentences in time-restricted online measures, but later judge those same sentences as less acceptable in untimed offline tasks (Phillips, Wagers & Lau 2011). A prominent example involves errors of "agreement attraction" in ungrammatical sentences like *The key to the cabinets were rusty*, which are often erroneously treated as acceptable in time-restricted online measures, but reliably judged as less acceptable in untimed offline tasks (Phillips, Wagers & Lau 2011; Lewis & Phillips 2015). The prevailing assumption is that conflicting judgments in online and offline tasks reflect the application of two distinct cognitive systems to interpret language. There is one system that contains the mental machinery for fast and efficient communication, traditionally referred to as the "parser", and a slower backup system that defines the precise rules of the language and classifies grammaticality, traditionally referred to as the "grammar". On this view, divergence between online and offline data reflects a parser-grammar misalignment (Lewis & Phillips 2015).

The dual-analyzers account received its classic formulation in the 1970s by Bever and colleagues (e.g., Bever 1970; Fodor, Bever & Garrett 1974), who argued that the relation between grammatical rules and perceptual operations is more "abstract rather than direct". Later, the dual-analyzers account was presented under the slogan *We understand everything twice* introduced by Townsend & Bever (2001), who claimed that we interpret sentences by first constructing a "quick-and-dirty" parse of the sentence using a set of superficial strategies, heuristics, and sentence-level templates, and then apply the grammar as a backup if those strategies fail. The assumption for multiple analyzers is adopted in many popular sentence processing theories, such as those that rely on "good-enough" representations (Ferreira, Bailey & Ferraro 2002; Ferreira & Patson 2007; Karimi & Ferreira 2016). According to these accounts, the properties of the parser are revealed in online data collected using time-sensitive measures (e.g., speeded acceptability judgments, self-paced reading, eye-tracking, ERPs), and the properties of the grammar are revealed in offline data collected using time-insensitive measures (e.g., untimed acceptability judgments, Likert ratings, magnitude estimation).

Recently, Karimi and Ferreira (2016) offered an explicit process model that adopts dual analyzers. In their model (illustrated in Figure 1), the parser uses superficial strategies to construct a quick-and-dirty parse of the sentence. The representations generated by the parser are complete enough to advance communication, but sometimes have errors that require revision. If revision is required, the initial output of the parser will be analyzed by the grammar, which is a slow-going process that fills in details that were missed in the first pass by the parser.

Under this account, the parser and grammar reflect separate cognitive systems because they have independent functions (rapid communication vs. knowledge representation), operate over representations of a distinct kind (noisy "good-enough" templates vs. detailed hierarchical structure), and use a distinct set of rules (fallible heuristics vs. grammatical constraints) that operate on different time scales (fast vs. slow).

Linguistic illusions like those involving agreement attraction can be taken to reinforce a grammar-parser distinction because they suggest that real-time processing builds representations that are not licensed by the grammar, consistent with a dual-analyzers account.
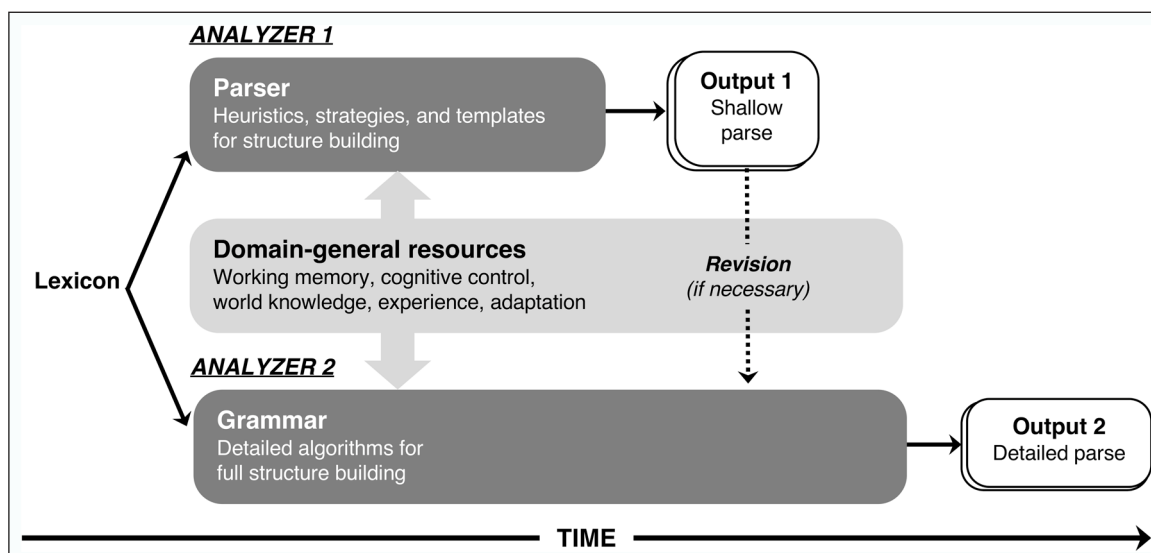
Figure 1: Graphical representation of the dual-analyzers account (adapted from Karimi & Ferreira 2016).

Although linguistic illusions were not originally part of the motivation for a dual-analyzers account, illusions have been presented as supporting evidence, as in Townsend & Bever (2001: 183–184). Consider the sentence in (1), which is ungrammatical because of the number mismatch between the verb and the head of its syntactic subject.

(1)   The key to **the cabinets** unsurprisingly   ("agreement attraction" configuration)
        *were rusty.

The claim in the literature on linguistic illusions is that there is a distinction between timed and untimed judgments for sentences like (1) (e.g., Phillips, Wagers & Lau 2011; Lewis & Phillips 2015). Comprehenders are often sensitive to number agreement errors when have they sufficient time to make their judgment. However, in time-restricted tasks, such as those involving speeded acceptability judgments, sentences like (1) are treated as acceptable on ~20–40% of trials due to the presence of the plural lure, i.e., the "attractor" (shown in bold in (1)). This effect constitutes an illusion of grammaticality because the lure creates the illusion that plural agreement is licensed. Importantly, attraction is not limited to subject-verb agreement: qualitatively similar effects have been shown for anaphora, ellipsis, case licensing, and negative polarity item (NPI) licensing (Drenhaus, Saddy & Frisch 2005; Arregui, Clifton, Frazier & Moulton 2006; Martin, Nieuwland & Carreiras 2012; 2014; Parker, Lago & Phillips 2015; Parker & Phillips 2016, 2017; Xiang, Dillon & Phillips 2009; Xiang, Grove & Giannakidou 2013). In each of these cases, illusions can arise in ungrammatical contexts, where the dependent element (reflexive, NPI, case marker, etc.) and target antecedent/licensor are incompatible (typically described in terms of feature match), but the presence of a non-target feature-matching lure tricks comprehenders into thinking that the dependency is licensed.

   To evaluate the claim that there is a distinction between timed and untimed responses, Table 1 provides a summary of findings in the field. This summary shows that in time-restricted binary ('yes/no') acceptability judgments, there is on average a 24% increase (range: 12–40%; median: 23%) in error rates for sentences with a feature-matching lure (computed as the increase from the ungrammatical condition that lacks a feature-matching lure). This effect drops to 12% (range: 9–17%; median 12%) in untimed binary

Table 1: Summary of judgments studies on linguistic illusions involving agreement attraction and illusory NPI licensing. "Attraction effect" is defined as the boost in acceptability for the critical ungrammatical plural attractor condition relative to the ungrammatical singular attractor condition (taken from the reported numerical values or estimated from figures when numerical values were not provided).

Citation	Dependency	Language	N	Attraction effect
Timed binary acceptability judgments				
Drenhaus et al. (2005), E1	NPI	German	24	13% boost
Xiang et al. (2006), E1	NPI	English	21	23% boost
Wagers (2008), E3	Agreement	English	16	40% boost
Wagers (2008), E5	Agreement	English	24	30% boost
Wagers (2008), E5	Agreement	English	24	22% boost
Wagers et al. (2009), E7	Agreement	English	16	30% boost
Franck et al. (2015), E3	Agreement	French	26	20% boost
Parker & Phillips (2016), E2	NPI	English	18	24% boost
Parker & Phillips (2016), E4	NPI	English	18	23% boost
Parker & Phillips (2016), E6	NPI	English	18	18% boost
Parker & Phillips (2016), E7	Agreement	English	18	21% boost
Parker & Phillips (2016), E7	Agreement	English	18	24% boost
de Dios Flores et al. (2017), E1	NPI	English	32	14% boost
Schlueter (2017), E10	Agreement	English	24	26% boost
Lago et al. (2018), E1	Agreement	Turkish	44	12% boost
Schlueter et al. (2018), E1	Agreement	English	30	28% boost
Schlueter et al. (2018), E3	Agreement	English	30	38% boost
Schlueter et al. (2018), E4	Agreement	English	30	27% boost
Hammerly et al. (2018), E1	Agreement	English	43	20% boost
Average				**24% boost**
Untimed binary acceptability judgments				
Xiang et al. (2006), E2	NPI	English	21	17% boost
Tanner (2011), E1	Agreement	English	17	10% boost
Xiang et al. (2013), E1	NPI	English	92	9% boost
Xiang et al. (2013), E1	Agreement	English	92	16% boost
Tanner et al. (2014), E1	Agreement	English	24	11% boost
Tanner et al. (2014), E2	Agreement	English	22	12% boost
Schlueter (2017), E11	Agreement	English	34	12% boost
Average				**12% boost**
Untimed scaled acceptability judgments				
Xiang et al. (2006)	NPI	English	14	.49 pt boost
Dillon et al. (2013)	Agreement	English	12	.75 pt boost
Parker & Phillips (2016) E1	NPI	English	18	.09 pt boost
Parker & Phillips (2016) E1	NPI	English	18	.09 pt boost
de Dios Flores et al. (2017), E2	NPI	English	16	.40 pt boost
Hammerly & Dillon (2017), E1	Agreement	English	64	.75 pt boost

Citation	Dependency	Language	N	Attraction effect
Hammerly & Dillon (2017), E1	Agreement	English	64	.39 pt boost
Hammerly & Dillon (2017), E3	Agreement	English	96	.64 pt boost
Hammerly & Dillon (2017), E3	Agreement	English	96	.63 pt boost
Yanilmaz & Drury (2018), E1	NPI	Turkish	38	.02 pt boost
Yanilmaz & Drury (2018), E1	NPI	Turkish	38	.02 pt boost
Average				**.38 pt boost**

acceptability judgments. Untimed scaled acceptability judgments show on average an increase of less than half a point in acceptability (along 5- and 7-point scales). Based on these findings, there is a distinction between timed and untimed judgments in comprehension, with the trend being an overall reduction in illusory licensing when participants are given more time to make their judgment. However, there is an unbalanced number of studies across methodologies, with most studies employing time-restricted judgments, and the effect sizes vary considerably across studies. Furthermore, none of these studies directly compared timed and untimed responses using the same set of items across methodologies, motivating the empirical basis of the current study.

The fact that we see different responses at different points in time for sentences like (1) is unsurprising if comprehenders engage multiple analyzers that rely on distinct rules and representations that operate on different time scales. For instance, agreement attraction effects might be expected if comprehenders apply template-based heuristics that rely on the proximity of the plural noun (Quirk et al. 1985), local syntactic coherence relations between the verb and plural lure (Tabor et al. 2004), or structural attachment preferences that are sensitive to competing non-target items (Villata, Tabor & Franck 2018). Application of these heuristics during rapid communication can produce error-prone representations that can initially appear acceptable, giving rise to illusions, but might later require revision by the slower, but more accurate grammatical system reflected in offline tasks.

A problem with a dual-analyzers account is that it does not provide a precise theory of how or when the grammar and parser interact in a predictable manner, e.g., how are errors detected? when are they revised?. Furthermore, if grammatical knowledge is applied on a time scale that is independent of speaking and understanding, then it is not possible to pinpoint grammatical processes in time using standard behavioral measures, making it difficult to develop and test linking hypotheses about the internal representations and grammatical behavior (Phillips 2004). By contrast, if grammatical knowledge is treated as a real-time system for constructing sentences, as the sole structure-building system, then the linking problem becomes more tractable (Phillips 1996; 2004; Lewis & Phillips 2015).

This alternative conception of the grammar as a structure-building system leads to a single-analyzer view of the cognitive architecture like that shown in Figure 2, in which both online and offline tasks rely on the same properties, namely the lexicon, the grammar, and limited general-purpose resources. On this view, the traditional notions of the "parser" and "grammar" simply reflect different descriptions of the same system: the grammar is just an abstraction from the processes involved in real-time sentence comprehension under the idealization of unbounded resources (Phillips 1996).

It would be more parsimonious, and maybe more cognitively efficient, if there were one linguistic analyzer for online and offline tasks. But it remains an empirical question how the cognitive architecture is organized. An important step to evaluate the plausibility of the single-analyzer hypothesis is to show that it can capture linguistic illusions.

Under a single-analyzer view, illusions arise due to limitations of the general-purpose memory access mechanisms that are recruited to implement grammatical computations (Lewis & Phillips 2015; Phillips et al. 2011). For instance, many researchers have argued that agreement attraction reflects error-prone memory retrieval mechanisms that are recruited by the grammar to implement long-distance syntactic dependencies (Wagers et al. 2009; Dillon et al. 2013; Tanner, Nicol & Brehm 2014; Lago et al. 2015; Tucker, Idrissi & Almeida 2015; Tucker & Almeida 2017). This account is based on memory studies showing that long-distance syntactic dependencies are implemented in real time by retrieving an antecedent/licensor from the preceding context using a cue-guided retrieval mechanism (Lewis 1996; McElree 2000; 2006; McElree, Foraker & Dyer 2003; Lewis & Vasishth 2005; Lewis, Vasishth & Van Dyke 2006; Van Dyke & McElree 2006; 2011; Jonides et al. 2008; Martin & McElree 2008; 2009; 2011). A key feature of this type of mechanism is that it is susceptible to interference from non-target items that match a subset of the retrieval cues, i.e., "partial matches". Drawing on these findings, Wagers et al. (2009) argued that agreement attraction errors likely reflect interference that stems from cue-based retrieval, as illustrated in Figure 3. In sentences like (1), encountering the plural marked verb *were* triggers a retrieval process that seeks a match to the required structural and morphological properties, e.g., [+subject] and [+plural]. On some trials, the attractor might be incorrectly retrieved due to a partial-match to the [+plural] cue, leading to
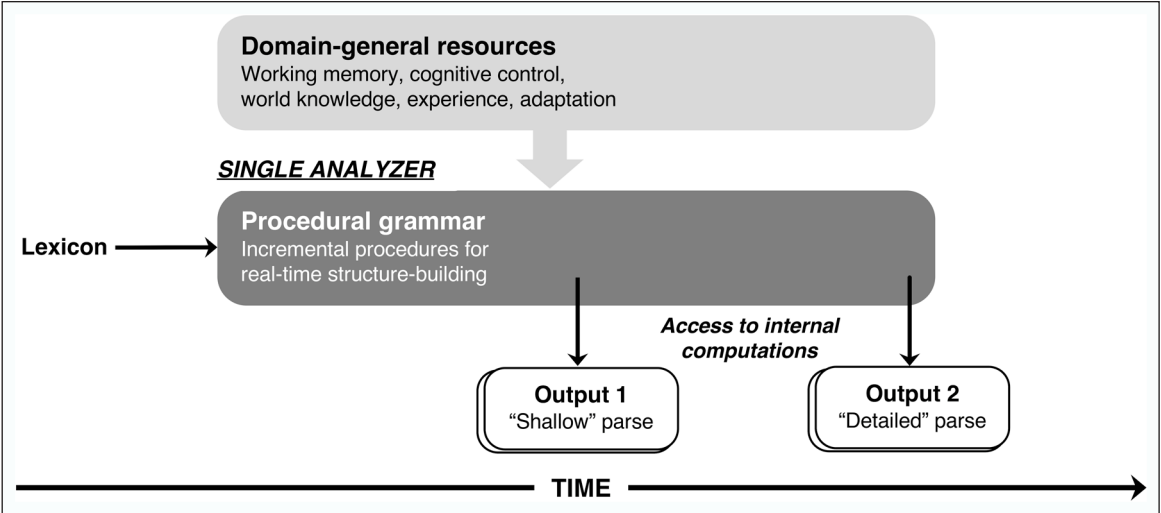
Figure 2: Graphical representation of the single-analyzer account (adapted from Phillips 1996).
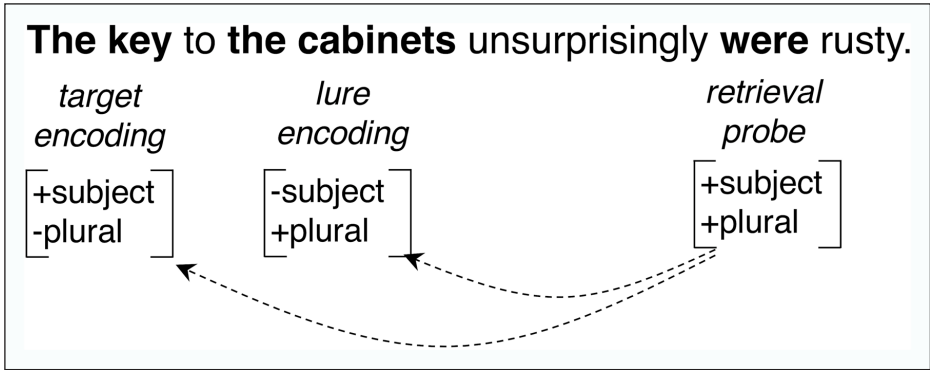
Figure 3: Graphical representation of the retrieval-based account of agreement attraction proposed by Wagers et al. (2009).

the false impression that agreement is licensed and boosting acceptability. On this view, agreement attraction errors reflect the exact constraints of grammar implemented by an error-prone memory retrieval mechanism, not the product of multiple analyzers.

The single-analyzer account provides an appealing explanation for why comprehenders are misled during online comprehension because it relies on independently motivated mechanisms, but it remains unclear why online and offline tasks yield conflicting responses if they are mediated by the same structure-building mechanism. One possibility suggested by Lewis & Phillips (2015) is that the increased grammatical accuracy observed in offline tasks might reflect improvement in the signal-to-noise ratio in grammatical processing over time. For instance, if offline judgments involve repeated attempts at retrieval over the same representation, then increased time for a judgment should yield improved grammatical accuracy, e.g., if there is a 25% chance of error on a single retrieval attempt, that outcome will become less dominant over multiple retrieval attempts to reprocess the sentence, yielding different outcomes at different points in time.

In the words of Lewis & Phillips (2015), mismatches between online and offline responses reflect different "snap-shots" of the internal steps involved in dependency formation. For instance, Lewis and Phillips reason that building a long-distance dependency involves multiple steps (lexical access, retrieval and/or prediction, integration, interpretation, discourse updating, etc.), and each of these steps take time to complete. If our experimental measures can tap into the results of the intermediate steps of those computations, we might sometimes elicit conflicting responses at different points in time. In short, online/offline mismatches may reflect the output of linguistic computations that are in various stages of completion, rather than the output of multiple analyzers.

Recently, similar proposals for iterative memory sampling has been invoked to explain certain timing effects that arise in long-distance dependency resolution. For instance, Dillon et al. (2014) found that in Mandarin Chinese, the processing of the long-distance reflexive *ziji* slows with increased syntactic distance to the target antecedent. To capture these effects, Dillon et al. (2014) presented a model of the antecedent retrieval process that relies on a series of serially executed, cue-based retrievals. Under this model, recovery of a distant antecedent takes more time than recovery of a local antecedent because more retrieval attempts are required to recover the distant antecedent. The notion of iterative memory sampling has also been implemented in a novel model of retrieval to capture effects of inhibitory interference, i.e., a slowdown at the retrieval site when multiple items match the retrieval cues (Nicenboim & Vasishth 2018). Beyond these studies though, the notion of iterative memory sampling has received little attention in research on linguistic dependency formation.

Lewis & Phillips' (2015) appeal to internal stages of computation to explain online/offline mismatches is intuitive, but it has not been tested yet because it does not provide enough detail about the computations to generate precise predictions. What is needed is an explicit process model that can explain how the internal states change over time, yielding both the cases of alignment and misalignment between online and offline responses. The current study seeks to address this issue.

1.1 The present study

The present study offers an explicit process model that is implemented in computational form to explain the mapping from online to offline responses in a single-analyzer architecture. The model is based on the proposal by Lewis & Phillips (2015) that the mapping from online to offline responses involves extended re-processing of the sentence in memory to minimize the signal-to-noise ratio.

Since the Lewis & Phillips (2015) proposal has not been implemented before, some architectural assumptions must be clarified. For explicitness, the proposal will be framed as

a process of sequential memory sampling in the cue-based memory retrieval framework (e.g., McElree 1993; 2000; McElree, Foraker & Dyer 2003; Lewis & Vasishth 2005; Lewis, Vasishth & Van Dyke 2006), in which a stimulus response is based on accumulation of evidence over time. In the cue-based memory framework, incorrect memory retrieval (i.e., retrieval of a non-target or "grammatically irrelevant" item) can trigger a "backtracking" process to reanalyze the sentence using sequential memory sampling (i.e., repeated retrieval attempts) (McElree 1993; McElree et al. 2003; Martin & McElree 2018). In the technical use of the term, backtracking refers to the process of returning to a choice point in the parse for reanalysis, and is often evoked to explain how the parser recovers from garden path effects (see Lewis 1998, for discussion). For present purposes, the notion of backtracking can be extended to memory retrieval processes, whereby retrieval mechanisms perform the same retrieval process multiple times over the same representation using the same set of cues used in the initial retrieval attempt, and aggregating the outcomes to minimize the signal-to-noise ratio, leading to more accurate representation of the current parser state (McElree 1993). This account is also inspired by "analysis-by-synthesis" models of perception, in which pattern recognition, symbolic generative processes, and hypothesis confirmations are performed by comparing a predicted pattern to the actual input, computing the error, and iterating the process until the error is minimized (see Bever & Poeppel 2010, for a review). Crucially, if linguistic dependency formation relies on cue-based retrieval, as previously claimed (Lewis 1996; McElree 2000; McElree et al. 2003; Lewis & Vasishth 2005; Lewis et al. 2006; Van Dyke & McElree 2006; 2007; 2011; Van Dyke 2007; Jonides et al. 2008; Martin & McElree 2008; 2009; 2011; Vasishth et al. 2008), then it is reasonable to assume that backtracking would apply uniformly to retrieval for linguistic dependencies, such as subject-verb agreement.

To provide a brief sketch of how this process plays out, consider again the sentence in (1). Here, incorrect retrieval of the attractor during online processing fails to satisfy the grammatical constraints on subject-verb agreement, e.g., it is not the subject of the verb, triggering a backtracking process to recover the target subject. Since backtracking takes time to complete, different outcomes are predicted at different points in time: initially, the wrong item can be retrieved, giving rise to agreement attraction in time-restricted online measures, but this retrieval error can be rectified via backtracking operations triggered by the grammar, eventually leading to the correct analysis reflected in offline judgments.

Three experiments were designed to test Lewis and Phillips' (2015) proposal that the mapping from online to offline responses reflects extended re-processing of sentences in memory. Experiments 1 and 2 used an agreement attraction paradigm to verify that online and offline measures yield contrasting profiles with respect to illusory licensing. The results of those experiments served as the basis for the computational implementation of the proposed process model in Experiment 3. To preview, the model generates a good fit to the data from Experiments 1 and 2, providing proof-of-concept for the single analyzer account.

2 Experiment 1: Timed judgments

A concern with previous research on agreement attraction is that few studies have directly compared speeded (timed, "online") responses and unspeeded (untimed "offline") responses using the same set of items across methodologies, making it difficult to assess existing generalizations about mismatches between time-sensitive and time-insensitive tasks. To address this issue, Experiments 1 and 2 directly compared the same set of items using timed and untimed forced-choice ('yes/no') acceptability judgments.

Experiment 1 used timed ("speeded") acceptability judgments to measure susceptibility to agreement attraction in a time-restricted task. In a speeded-acceptability judgment task,

sentences are presented one word at a time at a fixed rate. After the entire sentence has been presented, participants have up to three seconds to make a 'yes/no' response about the perceived acceptability of the sentence. Speeded acceptability judgments have been previously shown to reliably elicit attraction effects by restricting the amount of time that comprehenders have to reflect on acceptability intuitions (Drenhaus, Saddy & Frisch 2005; Wagers, Lau & Phillips 2009; Parker & Phillips 2016). As such, speeded acceptability tasks constitute an appropriate "online" measure, in the sense that they elicit a response relatively quickly, and offer a binary ('yes/no') measure that can be directly compared to the binary ('yes/no') untimed acceptability judgments in Experiment 2. Based on previous studies, agreement attraction is predicted to manifest in speeded judgments as increased rates of acceptance for ungrammatical sentences with an attractor that matches the number of the verb, relative to ungrammatical sentences that lack a number-matching attractor.

2.1 Method

2.1.1 Participants

Participants were 56 native speakers of English who were recruited using Amazon's Mechanical Turk web service. All participants provided informed consent and were screened for native speaker abilities. The screening probed knowledge of the constraints of English tense, modality, morphology, ellipsis, and syntactic islands. Participants were compensated $3.00 each. The experiment lasted approximately 20 minutes.

2.1.2 Materials

Experiment 1 used the same 24 item sets from Wagers et al. (2009) shown in Table 2, which represent the canonical agreement attraction paradigm. The experiment used a 2×2 factorial design, which crossed the factors grammaticality (grammatical vs. ungrammatical) and attractor number (singular vs. plural). In all conditions, the subject head noun was modified by a prepositional phrase that contained the attractor, and the agreeing verb was a past tense form of *be* (grammatical = *was*, ungrammatical = *were*). An adverb signaled the end of the prepositional phrase, and was included to delimit the effect of the verb (see Wagers et al. 2009, for discussion). Grammaticality was manipulated by varying the number of the verb such that it either matched or mismatched the number of the subject. Attractor number was manipulated such that the number of the attractor either matched or mismatched the number of the agreeing verb (plural vs. singular).

Each participant read 72 sentences, consisting of 24 agreement sentences and 48 filler sentences. Half of the fillers were ungrammatical resulting in an overall grammatical-to-ungrammatical ratio of 1:1. The ungrammatical fillers relied on a variety of grammatical errors, including unlicensed verbal morphology based on tense (e.g., *will laughing*) and unlicensed reflexive anaphors. The 24 sets of agreement items were distributed across 4 lists in a Latin square design. The filler sentences were of similar length and complexity to the agreement sentences. Materials were balanced such that half of the sentences were ungrammatical. The fill list of test sentences is provided in the Supplementary Materials.

Table 2: Sample set of materials from Experiment 1. PL = plural; SG = singular.

Grammatical, PL Attractor	The key to the cells unsurprisingly was dusty after many years of disuse.
Grammatical, SG Attractor	The key to the cell unsurprisingly was dusty after many years of disuse.
Ungrammatical, PL Attractor	The key to the cells unsurprisingly were dusty after many years of disuse.
Ungrammatical, SG Attractor	The key to the cell unsurprisingly were dusty after many years of disuse.

2.1.3 Procedure

Sentences were presented using the online presentation software Ibex Farm (Drummond 2018). Sentences were presented in the center of the screen, one word at a time, in a rapid serial visual presentation (RSVP) paradigm at a rate of 300 ms per word. Participants were instructed to judge whether each sentence was an acceptable sentence that a speaker of English might say. The full set of instructions for Experiments 1 and 2 are provided in the Supplementary Materials. A response screen appeared for 3 s at the end of each sentence during which participants made a 'yes/no' response by button press. If participants waited longer than 3 s to respond, they were given feedback that their response was too slow. The order of presentation was randomized for each participant.

2.1.4 Data analysis

Data were analyzed using logistic mixed-effects models, with maximal random effects structures. Each model included contrast coded fixed effects for experimental manipulations (±.5 for each factor), and their interaction, with random intercepts for participants and items (Baayen, Davidson & Bates 2008; Barr et al. 2013). Models were estimated using the *lmerTest* package in the R software environment (R Development Core Team, 2018). If there was a convergence failure, the random effects structure was simplified following Baayen et al. (2008).

2.1.5 Results

Figure 4 shows the percentage of 'yes' responses for the 4 experimental conditions. Average response times by condition are reported in Table 3. Results of the statistical analyses are reported in Table 4. A main effect of grammaticality, a main effect of attractor
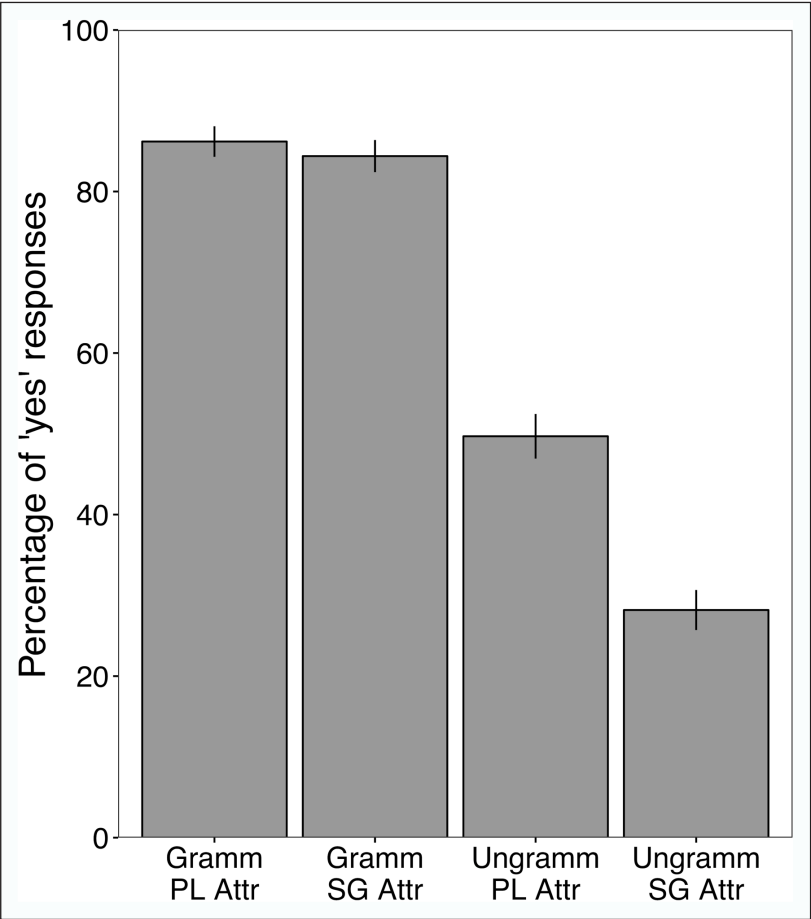

Figure 4: Speeded acceptability judgments and standard error by participants for Experiment 1.

Table 3: Average response times in milliseconds by condition for Experiment 1.

Grammatical, PL Attractor	585
Grammatical, SG Attractor	639
Ungrammatical, PL Attractor	592
Ungrammatical, SG Attractor	596

Table 4: Logistic mixed-effects model results for Experiment 1. Significant effects ($|z| > 2$ and $p < 0.05$) are in bold. Final model: glmer(rating ~ gram*attr + (1|item) + (1|participant), data = df, family = binomial).

	$\hat{\beta}$	**SE**	*z*	*p*
Intercept	**−1.10**	**0.20**	**−5.42**	**>0.01**
Grammaticality	**3.0**	**0.22**	**13.86**	**>0.01**
Attractor number	**1.10**	**0.18**	**6.08**	**>0.01**
Grammaticality × Attractor number	**0.95**	**0.29**	**3.24**	**>0.01**

number, and a significant interaction between grammaticality and attractor number were observed. Grammatical sentences were more likely to be accepted than ungrammatical sentences, and the interaction shows that the number of the attractor impacted grammatical and ungrammatical sentences differently. Planned pairwise comparisons revealed that the interaction was driven by a significant attraction effect in the ungrammatical conditions, as ungrammatical sentences with a plural attractor were more likely to be accepted than ungrammatical sentences with a singular attractor ($\hat{\beta}$ = 3.04, SE = 1.28, z = 2.36, p = 0.01). No such effect was observed in the grammatical conditions ($\hat{\beta}$ = 0.16, SE = 0.22, z = 0.70, p = 0.47).

2.2 Discussion

Results from Experiment 1 revealed n effect of agreement attraction in a time-restricted acceptability task, which appear as increased acceptability for ungrammatical sentences with an attractor that matched the number of the verb, relative to ungrammatical sentences that lacked a number-matching attractor. These results replicate those reported in previous studies that have used speeded acceptability judgments to elicit agreement attraction (e.g., Wagers et al. 2009; see also Parker & Phillips 2016), and provide a clear measure of time-restricted responses that will be directly compared to the untimed acceptability judgments in Experiment 2.

3 Experiment 2: Untimed judgments

Experiment 2 tested the same items from Experiment 1 using untimed forced-choice ('yes/no') acceptability judgments to obtain a measure of offline responses. Previous studies have reported that agreement attraction effects are reduced in offline tasks when participants have ample time to make their judgment (see Table 1). Experiment 2 sought to replicate this contrast using the same items in an RSVP forced-choice task. Typically, untimed acceptability judgment studies use Likert scale ratings, but Experiment 2 used a forced-choice ('yes/no') response design to provide a more direct comparison with the forced-choice speeded acceptability judgment data from Experiment 1. Based on previous untimed acceptability judgment studies (Table 1), ungrammatical sentences were predicted to show lower rates of acceptance relative to grammatical sentences, and unlike in the speeded judgments from Experiment 1, the presence of a plural attractor was expected not to modulate acceptability of the ungrammatical sentences.

3.1 Method

3.1.1 Participants

Participants were 56 native speakers of English from the College of William & Mary. Each participant provided informed consent and received credit in an introductory linguistics or psychology course. The experiment lasted approximately 25 minutes.

3.1.2 Materials

Experimental materials consisted of the same 24 sets of 4 items as in Experiment 1, with the same filler sentences.

3.1.3 Procedure

Sentences were presented using Ibex Farm, in RSVP mode, using the same parameters used in Experiment 1. However, unlike in Experiment 1, responses were not time-restricted, and participants were informed in the instructions that they could take as much time as they needed to record their response. Participants were instructed to read each sentence carefully, paying special attention to any errors that may be encountered. The instructions for Experiments 1 and 2 are provided in the Supplemental Materials. The order of presentation was randomized for each participant.

3.1.4 Data analysis

Data analysis followed the same steps as in Experiment 1. An additional model was built to test for an interaction of attraction (the effect of attractor number within the ungrammatical conditions) × task (timed judgments from Experiment 1 vs. untimed judgments from Experiment 2) to determine whether timed and untimed tasks yield contrasting profiles with respect to attraction effects.

3.1.5 Results

Figure 5 shows the percentage of 'yes' responses for the 4 experimental conditions. Average response times by condition are reported in Table 5. Results of the statistical analyses are reported in Table 6. A main effect of grammaticality was observed, as grammatical sentences were rated as more acceptable than ungrammatical sentences. Crucially, no effect of attractor number or an interaction between grammaticality and attractor number was observed ($ps > 0.1$), indicating that the presence of a plural attractor did not modulate ratings.

3.2 Discussion

Results from Experiment 2 showed that participants are sensitive to the number match between the subject head noun and the verb but are not misled by a number matching attractor when they are given ample time to make their judgment. These results replicate previous studies showing that attraction effects are reduced in untimed tasks (see Table 1). Crucially, Experiments 1 and 2 tested the same item sets and held constant the mode of presentation (RSVP) and the requirement for a forced-choice judgment, but showed contrasting profiles that hinged on whether or not judgments were elicited with a time restriction. This contrast is illustrated in Figure 6, which shows how much the presence of the plural attractor boosts (or fails to boost) acceptance rates in the ungrammatical conditions for timed and untimed judgments. This figure highlights that attraction is significantly reduced in untimed judgments. A statistical analysis supporting this contrast is presented in Table 7. In addition, average response times for Experiment 2 were also considerably longer than those from the speeded judgment task in Experiment 1, which is consistent with proposal that additional time for re-sampling reduces susceptibility to attraction. This proposal will be explored in-depth in the modeling experiment in the next section.
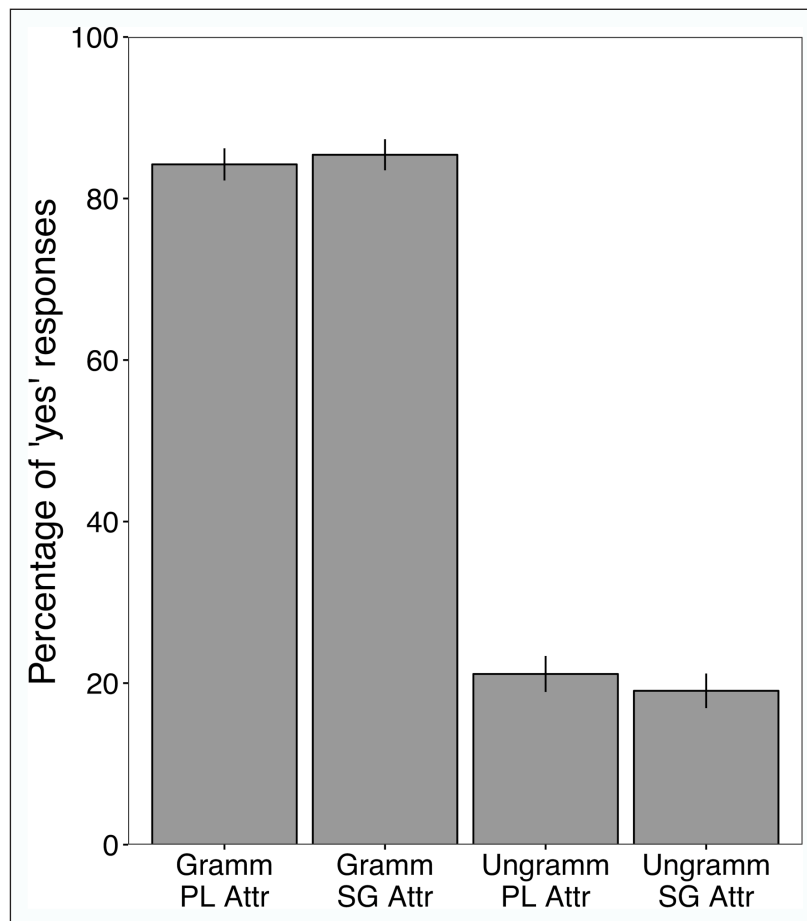
Figure 5: Mean untimed acceptability ratings and standard error by participants for Experiment 2.

Table 5: Average response times in seconds by condition for Experiment 2.

Grammatical, PL Attractor	2.33
Grammatical, SG Attractor	3.53
Ungrammatical, PL Attractor	2.05
Ungrammatical, SG Attractor	2.56

Table 6: Logistic mixed-effects model results for Experiment 2. Significant effects ($|z| > 2$ and $p < 0.05$) are in bold. Final model: glmer(rating ~ gram*attr + (1|item) + (1|participant), data = df, family = binomial).

	$\hat{\beta}$	SE	z	p
Intercept	**−1.86**	**0.24**	**−7.53**	**>0.01**
Grammaticality	**4.11**	**0.27**	**15.22**	**>0.01**
Attractor number	0.27	0.22	1.23	0.21
Grammaticality × Attractor number	−0.50	0.33	−1.49	0.13

   One surprising effect concerning the finishing times for Experiments 1 and 2 is that participants consistently took longer to respond in the grammatical condition with a singular attractor. A similar, albeit smaller effect is observed in the parallel ungrammatical conditions with a singular attractor. In these conditions, both the target and subject overlap
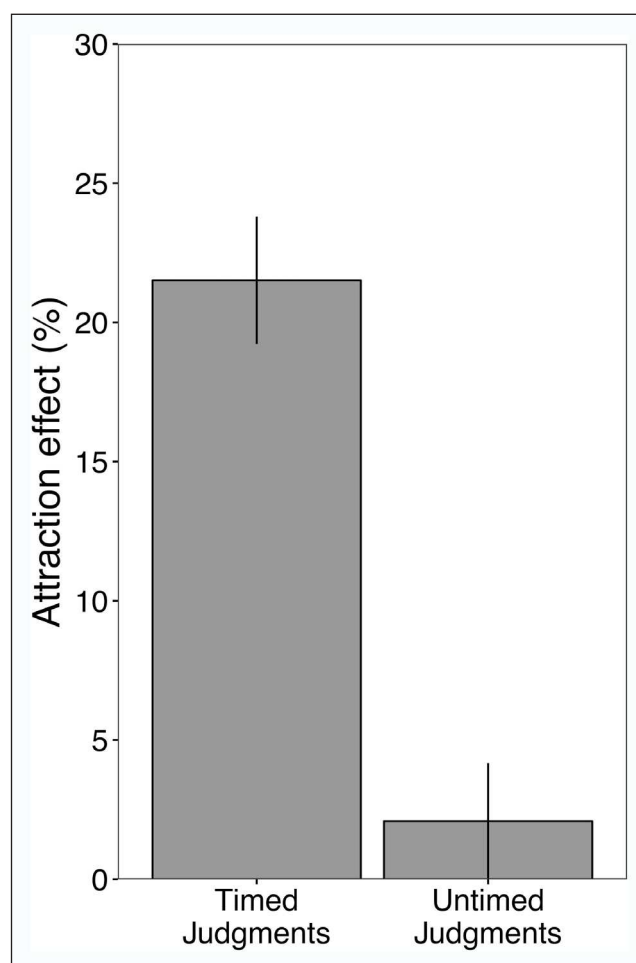
Figure 6: Comparison of effect sizes in the ungrammatical conditions from timed judgments (Experiment 1) and untimed judgments (Experiment 2).

Table 7: Logistic mixed-effects model comparing the effects of attraction between Experiments 1 and 2. The model was fit with the factor experiment as a between-participant factor. Significant effects ($|z| > 2$ and $p < 0.05$) are in bold. Final model: glmer(rating ~ attr*expt + (1|participant), data= df, family = binomial).

	$\hat{\beta}$	SE	z	p
Intercept	**−2.02**	**0.26**	**−755**	**>0.01**
Attraction	0.20	0.23	0.86	0.38
Experiment	**1.02**	**0.25**	**4.06**	**>0.01**
Attraction × Experiment	**1.07**	**0.30**	**3.57**	**>0.01**

in features with the retrieval cues (e.g., both are singular nouns). A likely possibility is that the increased time in these conditions reflects a "fan" effect at the stage of retrieval (Anderson 1974; Anderson & Reder 1999), which can lead to increased processing times when multiple items match the retrieval cues (Badecker & Straub 2002; Autry & Levine 2014; but cf. Chow, Lewis & Phillips 2014). Alternatively, it could reflect an effect of feature-overwriting at the stage of encoding (Nairne 1990; Vasishth, Jäger & Nicenboim 2017), where the overlap in features degrades the quality of the target representation, making recovery of the target more difficult at the stage of retrieval.

Taken together, Experiments 1 and 2 confirm that online/offline mismatches involving agreement attraction reflect the time sensitivity of the task (Lewis & Phillips 2015). These results will form the empirical basis of the single-analyzer process model developed and tested in Experiment 3.

4 Online/offline process model

Experiments 1 and 2 revealed a contrast between timed and untimed ("online" and "offline") judgments: attraction effects were observed in time-restricted judgments, but were reduced in untimed judgments when participants were given ample time to respond. Previously, online/offline mismatches of this sort have been presented as evidence for separate linguistic analyzers for online and offline tasks. However, recently, it has been argued that online/offline mismatches reflect a single linguistic analyzer for both online and offline tasks. According to this account, the increased grammatical accuracy observed in untimed offline tasks reflects extended re-processing of the sentence in memory to minimize the signal-to-noise ratio in grammatical processing over time (Lewis & Phillips 2015). This account is appealing for its simplicity, but it has not been explicitly tested.

Experiment 3 used computational modeling to test Lewis & Phillips' (2015) proposal. To make their account explicit, the mapping from online to offline responses was modeled as a process of sequential memory sampling in the independently-motivated cue-based retrieval framework (McElree 2000; McElree et al. 2003; Lewis & Vasishth 2005; Lewis et al. 2006). In this model, retrieval of a non-target item during online dependency formation, such as in the case of agreement attraction, triggers a backtracking process that involves sequential sampling using the same cues used in the initial retrieval attempt to recover the target subject. This process takes time to complete, predicting different outcomes at different points in time that can be mapped to online and offline judgments. Crucially, the model qualifies as a single-analyzer account because online and offline responses are generated using the same rules and representations to satisfy the grammatical constraints on subject-verb agreement. The following subsections describe the model in detail.

4.1 Description of the model

To derive quantitative predictions for timed and untimed responses, the current study used a variant of the ACT-R model of sentence processing described in Lewis & Vasishth (2005), which implements a cue-based retrieval mechanism for syntactic dependency formation [using code originally developed by Badecker & Lewis (2007)]. ACT-R (Adaptive Control of Thought—Rational; Anderson et al. 2004) is a general cognitive architecture based on independently motivated principles of memory and cognition, and has been applied to investigate a wide range of cognitive behavior involving memory access, attention, executive control, and learning. The ACT-R model of sentence processing applies the cognitive principles embodied in the general ACT-R framework to the task of sentence processing.

In the model, the words and phrases of a sentence are encoded as "chunks" (Miller 1956) in content-addressable memory (Kohonen 1980), and hierarchical sentence structure is represented using pointers that index the local relations between chunks. Chunks are encoded as bundles of feature-value pairs, which are inspired by the attribute-value matrices described in head-driven phrase structure grammars (Pollard & Sag 1994). Features are specified for lexical content (e.g., morpho-syntactic and semantic features), syntactic information (e.g., category, case), and local hierarchical relations (e.g., parent, daughter, sister). Values for features include symbols (e.g., ±singular, ±animate) or pointers to other chunks (e.g., NP1, VP2).

Linguistic dependencies, such as subject-verb agreement, are constructed using a domain-general cue-guided retrieval mechanism. This mechanism probes all previously encoded chunks in memory to recover the left part of the dependency (i.e., the target/licensor) using a set of retrieval cues that are compiled into a retrieval probe. Retrieval cues are derived from the current word, the linguistic context, and grammatical constraints, and correspond to a subset of the features of the target (Lewis et al. 2006).

The current model falls under the class of "activation-based" models of memory access, as chunks are differentially activated based on their match to the retrieval cues (see Jonides et al. 2008, for a review). In this class of models, the probability of retrieving a chunk is proportional to the chunk's overall activation at the time of retrieval, modulated by decay and similarity-based interference from other items that match the retrieval cues. The activation of an item Ai is defined in Equation 1, which makes explicit four principles that are known to impact memory access: (i) an item's baseline activation B_i, (ii) the match between the item and each of the j retrieval cues in the retrieval probe S_{ji}, (iii) the penalty for partial matches PM between the cues of the retrieval probe and the item's feature values, and (iv) stochastic noise. [2]

(2)   Equation 1

$$A_i = B_i + \sum_{j=1}^{m} W_j S_{ji} - \sum_{k=1}^{p} PM_{ki} + \epsilon$$

Baseline activation B_i is calculated according to Equation 2, which describes the usage history of chunk i as the summation of n successful retrievals of i, where t_j reflects the time since the jth successful retrieval of i to the power of the negated decay parameter d. The output is passed through a logarithmic transformation to approximate the log odds that the chunk will be needed at the time of retrieval, based on its usage history. After a chunk has been retrieved, the chunk receives an activation boost, followed by decay.

(3)   Equation 2

$$B_i = \ln \left(\sum_{j=1}^{n} t_j^{-d} \right)$$

The degree of match between chunk i and the retrieval cues reflects the weight W associated with each retrieval cue j, which defaults to the total amount of goal activation G available divided by the number of cues (G/j). Weights are assumed to be equal across all cues. The degree of match between chunk i and the retrieval cues is the sum of the weighted associative boosts for each retrieval cue S_j that matches a feature value of chunk i. The associative boost that a cue contributes to a matching chunk is reduced as a function of the "fan" of that cue, i.e., the number of competitor items in memory that also match the cue (Anderson 1974; Anderson & Reder 1999), according to Equation 3.

(4)   Equation 3

$$S_{ji} = S - \ln \left(\text{fan}_j \right)$$

Partial matching makes it possible to retrieve a chunk that matches only some of the cues (Anderson & Matessa 1997; Anderson et al. 2004), creating the opportunity for retrieval

---

[2] Equations 1–7 are based on ACT-R 6.0. Readers familiar with the Lewis & Vasishth (2005) ACT-R model may notice the non-standard presentation of Equation 1: the sign on the partial match component has been moved outside of the summation to indicate its penalizing nature.

interference of the sort that leads to agreement attraction errors (Wagers et al. 2009). Partial matching is calculated as the matching summation over the k feature values of the retrieval cues. P is a match scale, and M_{ki} reflects the similarity between the retrieval cue value k and the value of the corresponding feature of chunk i, expressed by maximum similarity and maximum difference.

Lastly, stochastic noise contributes to the activation level of chunk i. Noise is generated from logistic distribution with a mean of 0, controlled by the noise parameter s, which is related to the variance of the distribution, according to Equations 4 and 5. Noise is recomputed at each retrieval attempt. Activation noise plays a critical role in the current analysis. Activation creates the opportunity for memory errors (Anderson & Matessa 1997), such as agreement attraction in real-time comprehension. The notion of noise in this framework is based on the hypothesis that memory trace activation fluctuates over time both randomly and as a function of usage (see Lewis & Vasishth 2005, for discussion).

(5)   Equation 4

$$\epsilon \sim \text{logistic}\left(0, \sigma^2\right)$$

(6)   Equation 5

$$\sigma^2 = \frac{\pi^2}{3} s^2$$

Ultimately, activation A_i determines the probability of retrieving a chunk according to Equation 6. The probability of retrieving chunk i is a logistic function of its activation with gain $1/s$ and threshold τ. Chunks with a higher activation are more likely to be retrieved.

(7)   Equation 6

$$P\left(\text{recall}\right) = \frac{1}{1 + e^{(-A_i - \tau)/s}}$$

Typically, the target item will have the highest probability of retrieval, because it has the highest degree of activation at retrieval due to its match to the retrieval cues. However, non-target items, such as attractors, can be activated based on a partial match to the retrieval cues (see the third term of Equation 1) and subsequently retrieved if their activation is higher than that of the target due to noise, giving rise to attraction effects.

Once an item is accessed in memory as described in Equations 1–6, it is checked by the grammar (the sole structure-building system) to determine whether it meets the grammatical requirements for dependency formation. If the item satisfies these requirements, it will be integrated into the current context by combining the cues and contents of the item to form a new memory trace (Eich 1982; Murdock 1983; Dosher & Rosedale 1989) with a feature reflecting its downstream dependency (Parker, Shvartsman & Van Dyke 2017). However, if the item does not satisfy grammatical requirements, e.g., because it is not in the required structural position, then the grammar will trigger a subsequent retrieval process that engages in iterative sampling over the same representation using the same cues that were used in the initial retrieval attempt to recover the target. This process sequentially aggregates the outcomes from each retrieval iteration to minimize the signal-to-noise ratio such that retrieval of the target becomes the dominant outcome over time.

In an attraction configuration such as (1), if a non-target item has a subset of the required features, such as a plural feature for plural subject-verb agreement, the process of checking the plural feature can temporarily boost acceptability, giving rise to attraction effects, but resampling will still occur because the attractor does not satisfy the grammatical

constraints on subject-verb agreement, i.e., it is not the subject of the verb. Crucially, sequential sampling will decrease the probability of retrieval error over time, eventually leading to the grammatically correct analysis revealed in later offline judgments. This process will take time to complete, predicting different outcomes depending on the amount of time that comprehenders have to process the sentence, e.g., initial time-sensitive vs. untimed responses. Importantly, during reprocessing, the model relies on the same rules and representations used in the initial retrieval attempt, consistent with the single-analyzer account of sentence comprehension proposed by Lewis & Phillips (2015). That is, sequential sampling in the model does not resort to different rules, build a different set of representations, or invoke different mechanisms for timed and untimed tasks.

4.1.1 Procedure for the simulations

The goal of the computational simulations was to determine whether sequential memory sampling could capture the conflicting responses observed in timed and untimed measures for the critical attractor conditions from Experiments 1 and 2. Simulations modeled retrieval for all four conditions in Table 2. Here, it is important to spell out a key assumption regarding the role of retrieval in agreement processing. As discussed in the introduction, previous studies have shown that agreement attraction arises in ungrammatical, but not in grammatical sentences (e.g., Wagers et al. 2009; Dillon et al. 2013). Wagers and colleagues offered two suggestions for how a retrieval-based account could capture this grammatical asymmetry. One possibility is that retrieval functions as an error-driven repair mechanism that is triggered by the detection of an agreement violation. In the items in Table 2, the subject NP predicts the number of the verb. When the verb violates this prediction, as in the ungrammatical conditions, the parser engages cue-based retrieval at the verb to recover a number matching noun to license agreement. In the ungrammatical conditions with a plural verb and plural attractor, the attractor should sometimes be incorrectly retrieved because it matches the verb in number, leading to the false impression that agreement is licensed. In the grammatical conditions, the verb fulfills the number prediction made by the subject NP, and therefore retrieval is not engaged. Another possibility is that retrieval is always engaged, regardless of grammaticality. On this view, no attraction is expected in the grammatical condition, since the fully matching target NP should strongly outcompete partial matches. Although current time course evidence favors a prediction-based account of agreement processing (see Parker et al. 2018, for a review), I report the results of the retrieval simulations for both the grammatical and ungrammatical conditions for completeness. However, it is the changes in behavior over time for the ungrammatical condition with the plural attractor that is of key theoretical interest for the current study.

To model online responses, 100 Monte Carlo simulations were run for each condition, with each trial representing a single, independent retrieval attempt for dependency formation. To model offline responses, an additional 100 simulations were run using the same mechanisms, retrieval cues, and memory encodings that were used for online measures, with each trial repeating the same retrieval process up to 20 times (each trial reflects the aggregate outcome of 20 retrieval attempts, in which each of the 20 retrieval attempts was sequentially averaged together to yield the aggregate outcome). The results of each retrieval attempt were sequentially averaged together to minimize the signal-to-noise ratio over time. All trials averaged together yield the aggregate response reflected in offline tasks.

Some important questions regarding this implementation concern how the system determines whether iterative memory sampling is required and how acceptability is decided. For the current study, it was taken as a given that iterative sampling was required, and

that iterative sampling would terminate after a pre-determined number of samples. This approach was taken to evaluate what the overall process would achieve. There are several ways in which the triggering and evaluation processes might play out in actual comprehension. One possibility is that iterative sampling is triggered when the structural features of the retrieved item do not match the corresponding structural cues of retrieval probe. On this view, initial acceptability is based on the match between the number feature of the retrieved item and the corresponding number cue in the retrieval probe. Alternatively, it could be the error signal from the violation of the number prediction made by the target subject that triggers iterative sampling. For example, a violated prediction signals that something is amiss and that more information about the sentence is needed, motivating additional retrievals.

Also important to note is that the current model did not simulate the activation boost that arises with additional retrievals. In the ACT-R framework, each time an item is retrieved, that item receives a boost in activation. On this view, iterative memory sampling would quickly boost the activation levels of the target and attractor (depending on their individual rates of retrieval), which might modulate the outcome. In the current implementation, each sample was treated as an independent event, such that the activation boosts associated with retrieval did not feed subsequent samples.[3]

Two measures are reported for online and offline data: (i) activation values for the target (i.e., head subject noun) and the attractor, and (ii) predicted retrieval error rate. Since activation directly determines the probability of retrieval for the target and attractor, showing the underlying activation values across simulations provides insight into the amount of competition between the target and attractor during online vs. offline processing. Crucially, these activation values feed the main measure of interest, which is the predicted retrieval error rate. Predicted retrieval error rate reflects the percentage of runs for which the attractor was retrieved, rather than the target. Following previous studies, predicted retrieval error rate is assumed to map monotonically to human acceptability judgments, with higher retrieval error rates corresponding to increased rates of judgment errors (Vasishth et al. 2008; see also Kush & Phillips 2014; Parker & Lantz 2017).

All simulations used the default parameter setting reported in Lewis & Vasishth (2005) to ensure that the model would be predictive, rather than post-hoc. This method demonstrates that the predicted profiles are not the product of a special parameter setting that was hand-selected to approximate the data, but rather an accurate representation of the independently- and empirically-motivated principles of working memory embodied in the architecture.

4.1.2 Simulation results

Simulation results for the grammatical conditions are shown in Figures 7 and 8, and the results for the critical ungrammatical conditions are shown in Figures 9 and 11. These figures show the distribution of activation values for the target (solid line) and the attractor (dashed line) over all simulations for time-restricted online measures, where each trial is based on a single retrieval attempt (top), and untimed offline measures, where each trial reflects the aggregate outcome based on iterative memory sampling (bottom).

Results for the grammatical conditions show an initial activation advantage for the target in online measures (initial overlap between the target and attractor activation distributions were less than 2% in both the grammatical singular and plural attractor conditions), which persists into the offline judgments. Simulations predicted less than 2%

---

[3] Based on the simulation results presented in the next subsection, modeling the activation boosts such that they feed forward would likely lead to a more rapid increase in grammatically accurate judgments, due to the initial activation advantage for the target.
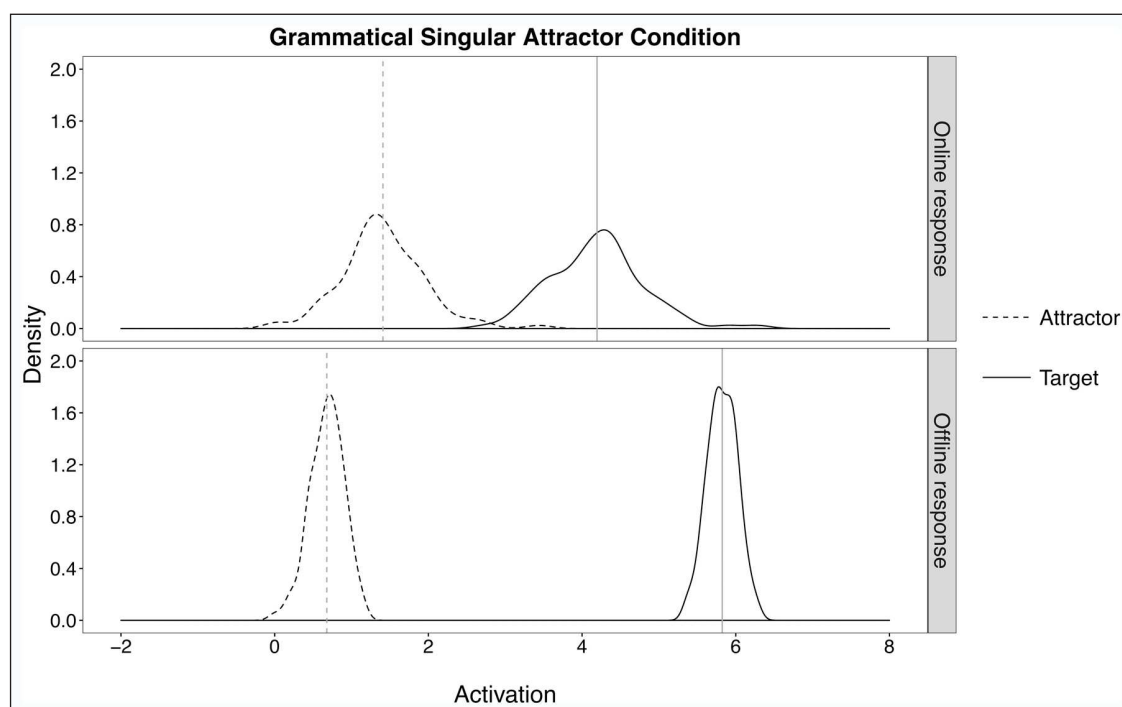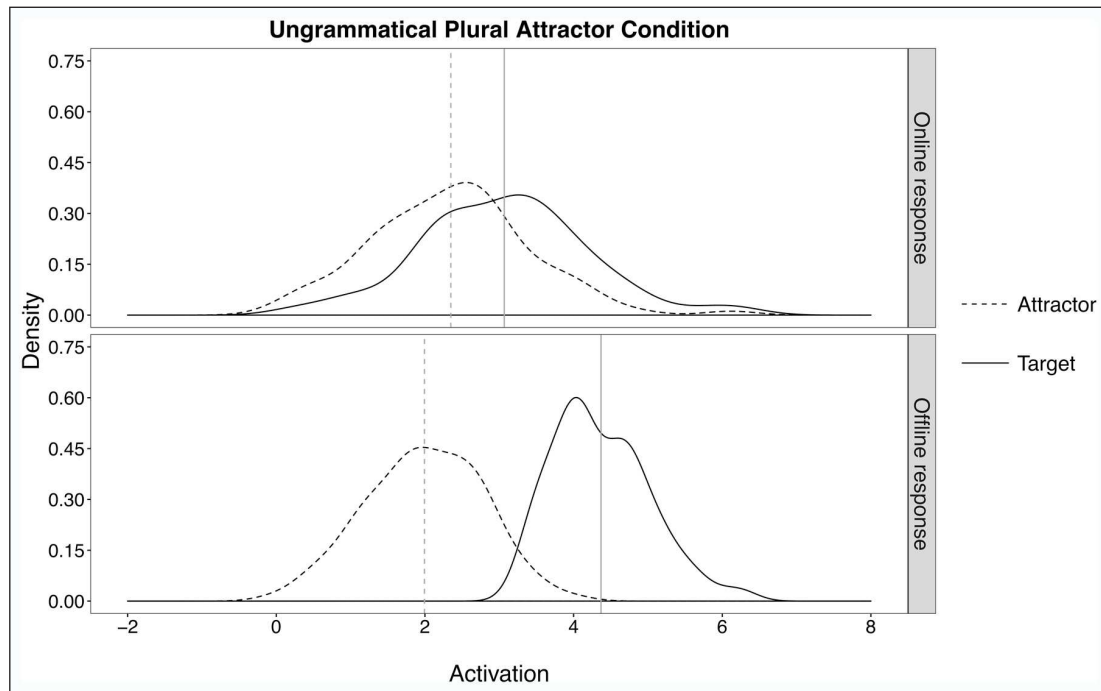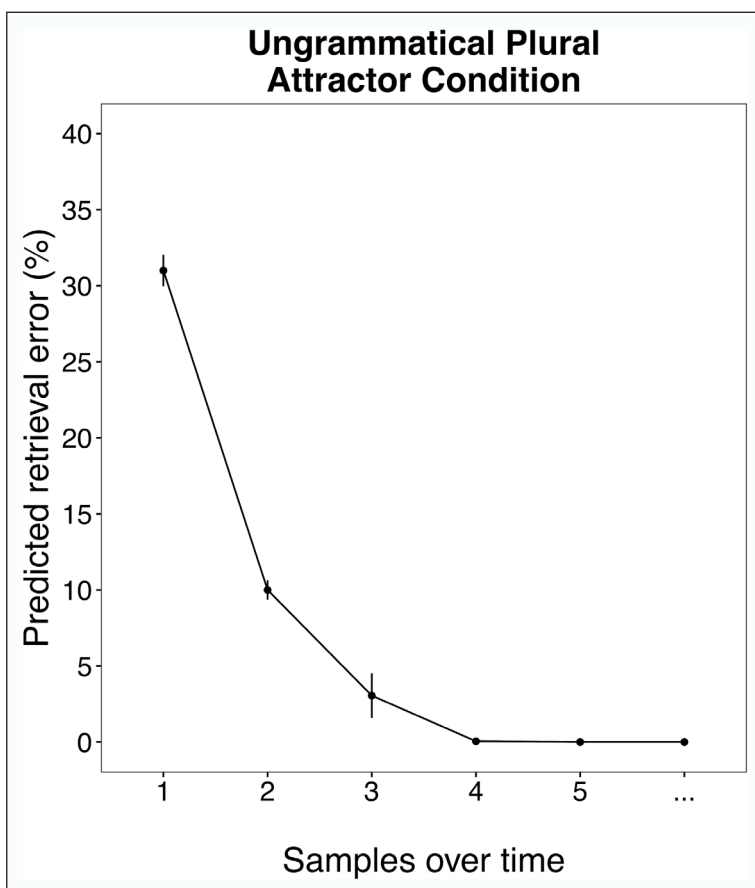
Figure 7: Grammatical singular attractor condition. Predicted activation distributions for the target (solid line) and the attractor (dashed line) for online responses (top) and offline responses (bottom). Vertical gray lines indicate the means for the corresponding distributions.
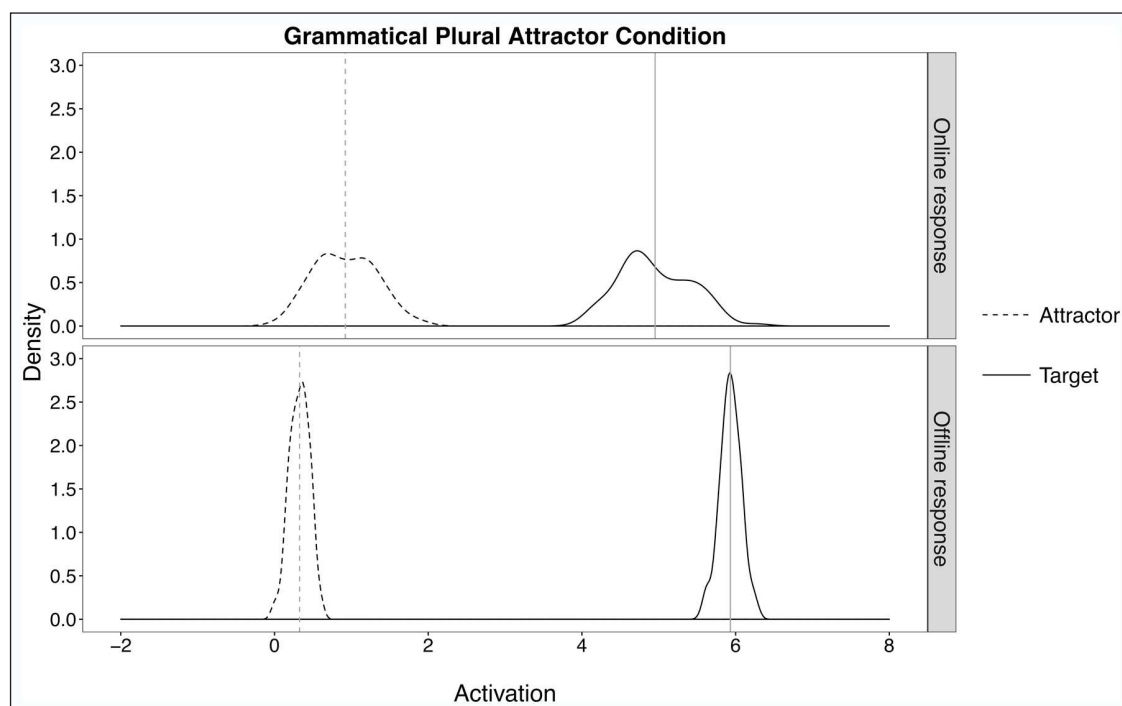
Figure 8: Grammatical plural attractor condition. Predicted activation distributions for the target (solid line) and the attractor (dashed line) for online responses (top) and offline responses (bottom). Vertical gray lines indicate the means for the corresponding distributions.

chance of retrieval error (i.e., retrieval of the attractor) in the online measures, which carries through to offline measures. These results suggest that retrieval accuracy is already at or near ceiling in online measures. Overall, simulations predicted high rates of accuracy

Figure 9: Ungrammatical plural attractor condition. Predicted activation distributions for the target (solid line) and the attractor (dashed line) for online responses (top) and offline responses (bottom). Vertical gray lines indicate the means for the corresponding distributions.

in both of the grammatical conditions, with no major changes in accuracy predicted in the transition from online to offline responses, as observed in Experiments 1 and 2.

Results for the ungrammatical conditions show a different profile. In particular, results for the critical ungrammatical plural attractor condition revealed a striking contrast between online and offline measures. Online measures show substantial overlap between the activation distributions for the target and distractor, increasing the opportunity for retrieval error, i.e., agreement attraction (percentage of overlap between the activation distributions: 74%). By contrast, offline measures show a separation between the activation distributions for the target and attractor, with a clear activation advantage for the target that reduces the opportunity for error in offline responses (percentage of overlap between the activation distributions: 8%).

The impact of sequential sampling on retrieval error is illustrated in Figure 10, which shows that retrieval error decreases as the number of memory samples increases over time. Given that each retrieval attempt takes time to complete (a single retrieval attempt requires on average 300–1200 ms in the current simulations), the increased accuracy predicted by sequential sampling will be most clearly reflected in later measures involving untimed judgments. In sum, the modeling results are closely aligned with the behavioral data, showing a clear attraction effect in online measures after a single retrieval attempt, and the eventual nullification of attraction in offline measures due to sequential sampling over time.
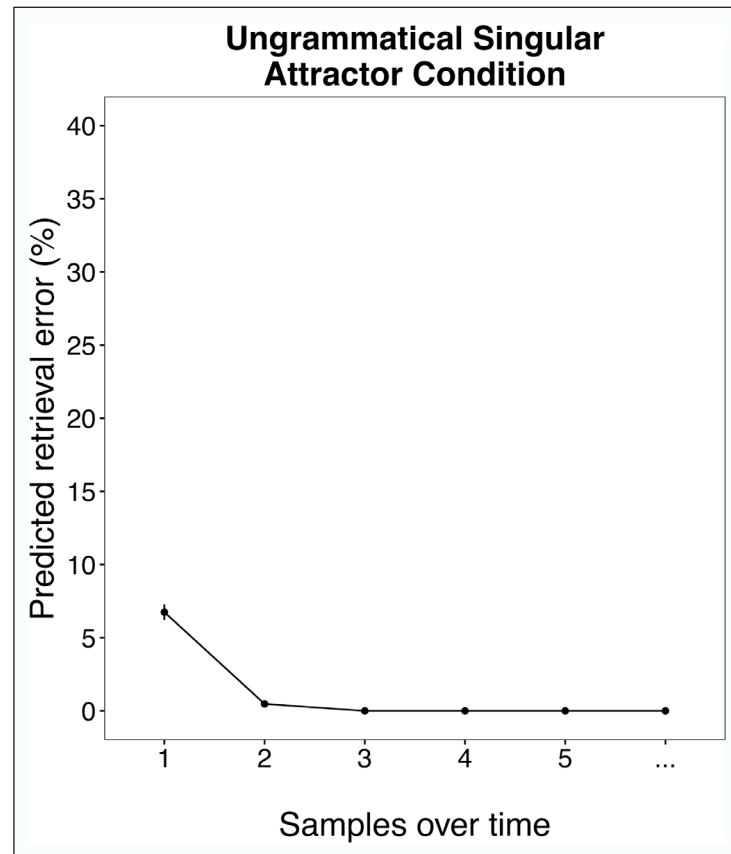
Results for the fully ungrammatical, singular attractor condition also showed improvement with repeated sampling. There was an initial advantage for the target in online measures, as shown in Figure 9 (initial overlap: 17%). These results map well to the relatively low rates of acceptance observed for this condition in Experiment 1. Importantly, the activation advantage for the target increased with repeated sampling, as shown in Figure 10 (overlap: 0%), leading to the slightly improved accuracy observed for this condition in untimed judgments from Experiment 2.

Figure 10: Ungrammatical plural attractor condition. Predicted retrieval error as a function of memory sampling over time. Error bars indicate standard error of the mean.
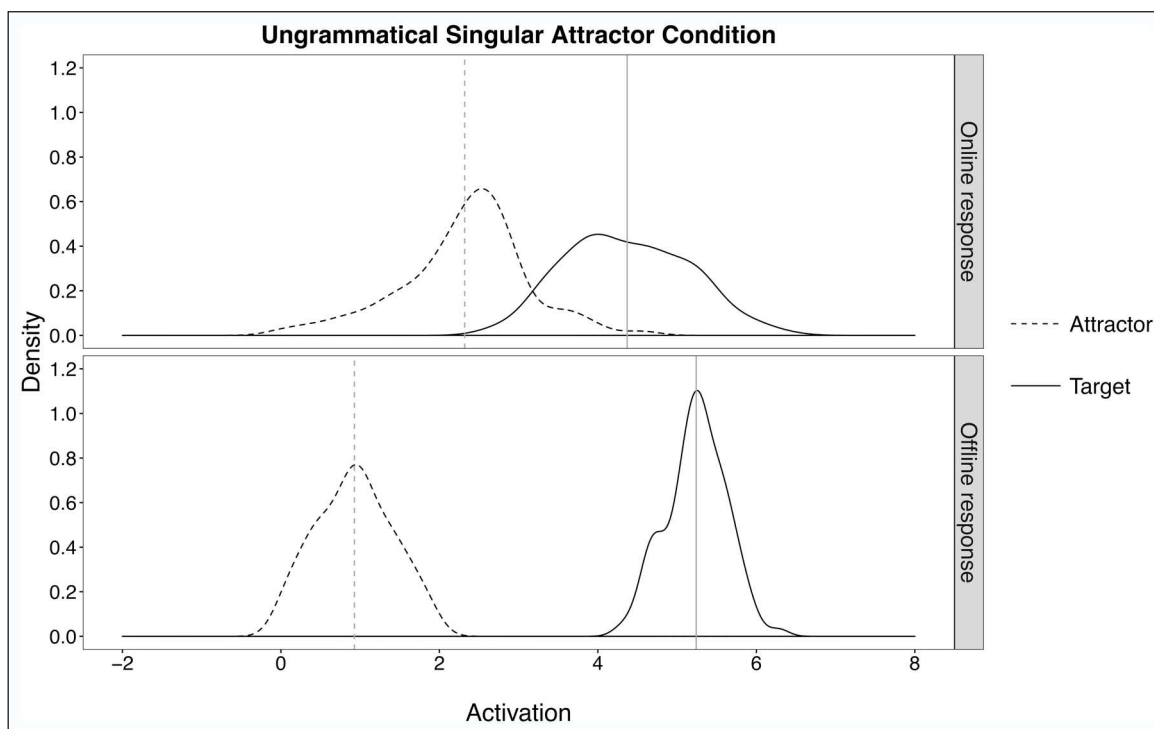
Figure 11: Ungrammatical singular attractor condition. Predicted activation distributions for the target (solid line) and the attractor (dashed line) for online responses (top) and offline responses (bottom). Vertical gray lines indicate the means for the corresponding distributions.

Figure 12: Ungrammatical singular attractor condition. Predicted retrieval error as a function of memory sampling over time. Error bars indicate standard error of the mean.

5 General discussion

5.1 Summary of results

The goal of the present study was to sharpen the issues concerning the debate over the cognitive architecture of language by testing the hypothesis that online and offline responses for sentence comprehension are the product of a single structure-building system embedded in a noisy cognitive architecture, and that mismatches between timed and untimed judgments about a sentence reflect extended re-processing to minimize the signal-to-noise ratio in grammatical processing over time (Lewis & Phillips 2015). To test this hypothesis, the current study focused on a specific type of online/offline mismatch involving agreement attraction.

Experiments 1 and 2 verified the online and offline generalizations reported in the literature using a single set of items across experimental methods: comprehenders treat ill-formed agreement dependencies with a feature-matching attractor as acceptable in time-restricted measures, but judge those same sentences as less acceptable in untimed measures. Experiment 3 then offered an explicit process model based on the single-analyzer account of the linguistic cognitive architecture (Phillips 2004; 2013; Phillips et al. 2011; Lewis & Phillips 2015). The model captured the mapping between online and offline responses as a process of error-driven sequential sampling in the cue-based memory retrieval framework (Lewis & Vasishth 2005; Lewis et al. 2006). The key prediction of the model is that different outcomes are expected at different points in time, which can be tracked by timed and untimed measures. Modeling results were closely aligned with the behavioral data, showing attraction in initial timed judgments, and a rapid reduction and eventual nullification of attraction in offline tasks as a function of sequential sampling over time.

The current results have several implications for our understanding of the source of agreement attraction effects and the cognitive architecture of language. First, the behavioral experiments from the current study (Experiments 1–2) sharpened the empirical issue concerning contrast between online and offline tasks involving agreement attraction by isolating the effect of timing in a way that previous studies on agreement attraction had not. Holding constant the mode of presentation, Experiments 1 and 2 provided empirical support for the claim that previously observed contrasts between online and offline data are distinguished by the time sensitivity of the response. Second, and more importantly, the results of the current study provide proof-of-concept that one type of online/offline mismatches involving agreement attraction can be captured in the single-analyzer framework (illustrated in Figure 2), without positing separate analyzers for online and offline tasks. Specifically, the current study drew on a widely-used model (ACT-R) and showed that by extending the model to perform iterative memory sampling, we are able to capture the contrast between online and offline data without recourse to a special class of extra-grammatical strategies or heuristics. In this way, the notion of resampling provides an explicit proposal for what constitutes "reflection" in linguistic judgment tasks, namely that it might involve repeated re-sampling of an activation-based memory to better distinguish between grammatical and ungrammatical strings. More broadly, the current results lend further support to the claims that reanalysis entails additional processing time (Martin & McElree 2018), and that multiple retrieval attempts can account for reanalysis effects without recourse to a specialized reanalysis mechanism (Van Dyke & Lewis 2003; Martin & McElree 2018).

A concern with the current study is that the proposed model does not predict acceptability judgments *per se*. In the current study, it was simply assumed that memory activations and the output of retrieval processes feed judgments in a monotonic fashion. However, there are alternative ways in which differences in activation for the target vs. attractor could impact judgments, and an important task for future research is to test this assumption more rigorously. For instance, activation values may have a non-monotonic, probabilistic relation with judgments that incorporates uncertainty at various levels of representations, starting at the level of the input and ending with motor command for the button press. What is needed is an "end-to-end" model that maps directly from input to the button press for the judgment. A modest next step would be to integrate the current model with recent modeling efforts that simulate judgment distributions (e.g., Dillon et al. 2015), which would draw directly from the activation distributions observed in the current study.

5.2 Broader implications for theories of sentence comprehension

The current results do not disconfirm the dual-analyzers account. But they do provide the necessary proof-of-concept that at least one piece of evidence taken to support the dual-analyzers account can be captured in a single- analyzer architecture by drawing on independently motivated principles of general cognition. The current single-system account offers several advantages over the dual-analyzers account. First, the current account offers a plausible explanation for why grammatically accurate judgments are often slow or delayed. According to the dual- analyzers account, slow but accurate judgments are taken to reflect a grammatical analyzer that is distinct from the fast acting parser (Townsend & Bever 2001). However, slow responses do not necessarily entail a separate linguistic analyzer. Under the current single-analyzer account, the reason we sometimes see delayed accuracy is because comprehension relies on complex, multiple-step computations (constraint application, cue-generation, memory access, retrieval, integration, interpretation, etc.) that take time to complete, even for a relatively straightforward dependency like subject-verb agreement. If online and offline measures can access the internal stages of those computations, then it should be unsurprising to find different responses at different points in time.

Second, the current proposal offers a detailed linking hypothesis that relates the underlying cognitive architecture with observable linguistic behavior. If the grammar operates independently of observable online parsing behavior, as assumed under a dual-analyzers account, then grammatical computations will be difficult to pinpoint in time, making it impossible to develop or test linking hypotheses for linguistic knowledge and behavior (see Phillips 2004, for discussion). However, if online and offline phenomena are treated as different reflections of the same system, then the mental operations of the grammar become easier to pinpoint in time.

5.3 Extensions of the current proposal

The current process model captured the mapping from online to offline responses for agreement attraction effects. Importantly, the model is not simply a "one off" model built to explain a narrow range of effects for subject-verb agreement. As noted in the Introduction, attraction effects are observed for a wide range of dependencies involving anaphora, ellipsis, case licensing, and negative polarity items (Drenhaus et al. 2005; Vasishth et al. 2008; Xiang et al. 2009; Martin et al. 2012; Sloggett 2013; Xiang et al. 2013; Parker et al. 2015; Parker & Phillips 2016; 2017). The current model can be applied similarly to capture attraction effects for each of these dependencies. However, recent work suggests that there are subtle, qualitative differences in attraction effects across dependencies (Dillon et al. 2013; Parker & Phillips 2016; 2017), and an important task for future research is to test whether those nuances are captured in the current model.

Lastly, it is also worth noting that the proposed process model is compatible with the broader conclusions drawn in the perceptual and cognitive domains. For instance, Keren & Schul (2009) argued that in the visual system, conflicting responses at different points in time, such as those involving visual illusions, reflect a single representational system that relies on two different types of criteria to evaluate the system's output, resulting in contrasting percepts, rather than the output of multiple visual systems. Under the current single-analyzer view, the conflicting responses observed for agreement attraction also reflect different evaluation criteria, such as the initial feature match at retrieval for online measures, and the aggregate response based on sequential sampling for offline measures (when there is an initial retrieval error), resulting in contrasting percepts. A potentially fruitful line of future research would be to examine the extent to which the evaluation criteria are structured similarly across cognitive domains.

6 Conclusion

This paper argued that it is possible to capture mismatches between online and offline sentence acceptability judgments with a single structure-building system (the grammar) implemented in a noisy memory architecture, and provided a computational model as proof-of-concept. Although the current study has not directly ruled out the possibility of multiple linguistic analyzers, the results of the current study show that multiple analyzers are not necessary to capture online/offline mismatches, at least in the case of agreement attraction. These results provide new insight into the cognitive architecture for language and contribute to the development of an explicit linking hypothesis that relates the underlying cognitive system with observable linguistic behavior.

Additional File

The additional file for this article can be found as follows:

- **[[Title]]**. Experimental items from Experiments 1 and 2. DOI: https://doi.org/10.5334/gjgl.766.s1

Acknowledgements

Competing Interests

The author has no competing interests to declare.

References

Anderson, John Robert. 1974. Retrieval of propositional information from long-term memory. *Cognitive Psychology* 6. 451–474. DOI: https://doi.org/10.1016/0010-0285(74)90021-8

Anderson, John Robert, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere & Yulin Qin. 2004. An integrated theory of the mind. *Psychology Review* 111. 1036–1060. DOI: https://doi.org/10.1037/0033-295X.111.4.1036

Anderson, John Robert & Michael Matessa. 1997. A production system theory of serial memory. *Psychological Review* 104. 728–748. DOI: https://doi.org/10.1037/0033-295X.104.4.728

Anderson, John Robert & Lynne M. Reder. 1999. The fan effect: New results and new theories. *Journal of Experimental Psychology: General* 128. 186–197. DOI: https://doi.org/10.1037/0096-3445.128.2.186

Arregui, Ana, Chuck Clifton Jr., Lyn Frazier & Keir Moulton. 2006. Processing elided VPs with flawed antecedents. *Journal of Memory and Language* 55. 232–246. DOI: https://doi.org/10.1016/j.jml.2006.02.005

Baayen, Rolf Harald, Douglas Davidson & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412. DOI: https://doi.org/10.1016/j.jml.2007.12.005

Badecker, William & Richard L. Lewis. 2007. A new theory and computational model of working memory in sentence production: Agreement errors as failures of cue-based retrieval. *Paper presented at the 20th CUNY Conference on Human Sentence Processing.* University of California, San Diego.

Badecker, W. & Kathleen Straub. 2002. The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology* 28. 748–769. DOI: https://doi.org/10.1037//0278-7393.28.4.748

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013.Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the development of language*, 279–362. New York: Wiley.

Bever, Thomas G. & David Poeppel. 2010. Analysis by synthesis: A (re-)emerging program of research for language and vision. *Biolinguistics* 4. 174–200.

Chow, Wing-Yee, Shevaun Lewis & Colin Phillips. 2014. Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology* 27. 1–16. DOI: https://doi.org/10.3389/fpsyg.2014.00630

de Dios Flores, Iria, Hanna Muller & Colin Phillips. 2017. Negative polarity illusions: Licensors that don't cause illusions, and blockers that do. Poster at the 30th CUNY Conference on Human Sentence Processing. MIT.

Dillon, Brian, Alan Mishler, Shayne Sloggett & Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69. 85–103. DOI: https://doi.org/10.1016/j.jml.2013.04.003

Dillon, Brian, Wing-Yee Chow, Matthew W. Wagers, Taomei Guo, Fengqin Liu & Colin Phillips. 2014. The structure-sensitivity of memory access: Evidence from Mandarin Chinese. *Frontiers in Psychology* 5. 1–16. DOI: https://doi.org/10.3389/fpsyg.2014.01025

Dosher, Barbara A. & Glenda Rosendale. 1989. Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General* 118. 191–211. DOI: https://doi.org/10.1037/0096-3445.118.2.191

Drenhaus, Heiner, Douglas Saddy & Stefan Frisch. 2005. Processing negative polarity items: When negation comes through the backdoor. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives,* 145–165. Berlin: de Gruyter. DOI: https://doi.org/10.1515/9783110197549.145

Drummond, Alex. 2018. Ibex Farm. Retrieved from: http://spellout.net/ibexfarm.

Eberhard, Kathleen M., J. Cutting & Kathryn Bock. 2005. Making syntax of sense: Number agreement in sentence production. *Psychological Review* 112. 531–559. DOI: https://doi.org/10.1037/0033-295X.112.3.531

Eich, Janet Metcalfe. 1982. A composite holographic associated recall model. *Psychological Review* 89. 627–661. DOI: https://doi.org/10.1037/0033-295X.89.6.627

Ferreira, Fernanda, Karl G. D. Bailey & Vittoria Ferraro. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11. 11–15. DOI: https://doi.org/10.1111/1467-8721.00158

Ferreira, Fernanda & Nikole D. Patson. 2007. The 'good enough' approach to language comprehension. *Language and Linguistics Compass* 1. 71–83. DOI: https://doi.org/10.1111/j.1749-818X.2007.00007.x

Fodor, Jerry A., Thomas G. Bever & Merrill F. Garrett. 1974. *The psychology of language.* New York: McGraw-Hill.

Franck, Julie, Saveria Colonna & Luigi Rizzi. 2015. Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in Psychology* 6. 1–15. DOI: https://doi.org/10.3389/fpsyg.2015.00349

Hammerly, Christopher, Adrian Staub & Brian Dillon. 2018. The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology* 110. 70–104. DOI: https://doi.org/10.1016/j.cogpsych.2019.01.001

Hammerly, Christopher & Brian Dillon. 2017. Restricting domains of retrieval: Evidence for clause-bound processing from agreement attraction. Poster at the 30th CUNY Conference on Human Sentence Processing, MIT

Jonides, John, Richard L. Lewis, Derek Evan Nee, Cindy A. Lustig, Mark G. Berman & Katherine Sledge Moore. 2008. The mind and brain of short-term memory. *Annual Review of Psychology* 59. 193–224. DOI: https://doi.org/10.1146/annurev.psych.59.103006.093615

Karimi, Hossein & Fernanda Ferreira. 2016. Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology* 69. 1013–1040. DOI: https://doi.org/10.1080/17470218.2015.1053951

Keren, Gideon & Yaacov Schul. 2009. Two is not slways better than one: A critical evaluation of two-systems theories. *Perspectives on Psychological Science* 4. 533–550. DOI: https://doi.org/10.1111/j.1745-6924.2009.01164.x

Kohonen, Teuvo. 1980. *Content-addressable memories.* Berlin: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-642-96552-4

Kush, Dave & Colin Phillips. 2014. Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology* 5. 1–12. DOI: https://doi.org/10.3389/fpsyg.2014.01252

Lago, Sol, Diego Shalom, Mariano Sigman, Ellen F. Lau & Colin Phillips. 2015. Agreement processes in Spanish comprehension. *Journal of Memory and Language* 82. 133–149. DOI: https://doi.org/10.1016/j.jml.2015.02.002

Lewis, Richard L. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25. 93–115. DOI: https://doi.org/10.1007/BF01708421

Lewis, Richard L. 1998. Reanalysis and limited repair parsing: Leaping off the garden path. In Janet Dean Fodor & Ferreira Fernanda (eds.), *Reanalysis in sentence processing. Studies in theoretical psycholinguistics* 21. 247–284. Dordrecht: Springer. DOI: https://doi.org/10.1007/978-94-015-9070-9_8

Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29. 375–419. DOI: https://doi.org/10.1207/s15516709cog0000_25

Lewis, Richard L., Shravan Vasishth & Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10. 447–454. DOI: https://doi.org/10.1016/j.tics.2006.08.007

Lewis, Shevaun & Colin Phillips. 2015. Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research* 44. 27–46. DOI: https://doi.org/10.1007/s10936-014-9329-z

Linzen, Tal & Brian Leonard. 2018. Agreement attraction does not depend on time pressure. Poster at the 31st Annual CUNY Conference on Human Sentence Processing. UC-Davis.

Martin, Andrea E. & Brian McElree. 2008. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language* 58. 879–906. DOI: https://doi.org/10.1016/j.jml.2007.06.010

Martin, Andrea E. & Brian McElree. 2009. Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35. 1231–1239. DOI: https://doi.org/10.1037/a0016271

Martin, Andrea E. & Brian McElree. 2018. Retrieval cues and syntactic ambiguity resolution: Speed-accuracy tradeoff evidence. *Language, Cognition and Neuroscience* 6. 1–20. DOI: https://doi.org/10.1080/23273798.2018.1427877

Martin, Andrea E. & Brian McElree. 2011. Direct-access retrieval during sentence comprehension: Evidence from Sluicing. *Journal of Memory and Language* 64. 327–343. DOI: https://doi.org/10.1016/j.jml.2010.12.006

Martin, Andrea E., Mante S. Nieuwland & Manuel Carreiras. 2012. Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage* 59. 1859–1869. DOI: https://doi.org/10.1016/j.neuroimage.2011.08.057

Martin, A. E., Mante S. Nieuwland & Manuel Carreiras. 2014. Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain & Language* 135. 42–51. DOI: https://doi.org/10.1016/j.bandl.2014.05.001

McElree, Brian. 2000. Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research* 29. 155–200. DOI: https://doi.org/10.1023/A:1005184709695

McElree, Brian. 2006. Accessing recent events. In Brian H. Ross (ed.), *The psychology of learning and motivation – Advances in research and theory*, 155–200. San Diego, CA: Academic Press. DOI: https://doi.org/10.1016/S0079-7421(06)46005-9

McElree, Brian, Stephanie Foraker & Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48. 67–91. DOI: https://doi.org/10.1016/S0749-596X(02)00515-6

Miller, George A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 101. 343–352. DOI: https://doi.org/10.1037/0033-295X.101.2.343

Murdock Jr, Bennet B. 1982. A theory for the storage and retrieval of item and associative information. *Psychological Review* 89. 609–626. DOI: https://doi.org/10.1037/0033-295X.89.6.609

Nairne, James S. 1990. A feature model of immediate memory. *Memory & Cognition* 18. 251–269. DOI: https://doi.org/10.3758/BF03213879

Nicenboim, Bruno & Shravan Vasishth. 2018. Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language* 99. 1–34. DOI: https://doi.org/10.1016/j.jml.2017.08.004

Parker, Dan & Colin Phillips. 2016. Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition* 157. 321–339. DOI: https://doi.org/10.1016/j.cognition.2016.08.016

Parker, Dan & Colin Phillips. 2017. Reflexive attraction in comprehension is selective. *Journal of Memory and Language* 94. 272–290. DOI: https://doi.org/10.1016/j.jml.2017.01.002

Parker, Dan & Daniel Lantz. 2017. Encoding and accessing linguistic representations in a dynamically structured holographic memory system. *Topics in Cognitive Science* 9. 51–68. DOI: https://doi.org/10.1111/tops.12246

Parker, Dan, Michael Shvartsman & Julie A. Van Dyke. 2017. The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In Linda Escobar, Vicenç Torrens & Teresa Parodi (eds.), *Language processing and disorders*, 121–144. Newcastle: Cambridge Scholars Publishing.

Parker, Dan, Sol Lago & Colin Phillips. 2015. Interference in the processing of adjunct control. *Frontiers in Psychology* 6. 1–13. DOI: https://doi.org/10.3389/fpsyg.2015.01346

Phillips, Colin. 1996. *Order and structure*. Cambridge, MA: Massachusetts Institute of Technology Dissertation.

Phillips, Colin. 2004. Linguistics and linking problems. In Mabel Rice & Steven F. Warren (eds.), *Developmental language disorders: From phenotypes to etiologies*, 241–287. Mahwah, NJ: Lawrence Erlbaum Associates.

Phillips, Colin. 2013. Parser-grammar relations: We don't understand everything twice. In Montserrat Sanz, Itziar Laka & Michael K. Tanenhaus (eds.), *Language down the garden path: The cognitive basis for linguistic structure*, 294–315. Oxford: Oxford University Press.

Phillips, Colin, Matthew W. Wagers & Ellen F. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In Jeffrey Runner (ed.), *Experiments at the interfaces (Vol. 37)*, 147–180. Bingley: Emerald Publications. DOI: https://doi.org/10.1108/S0092-4563(2011)0000037009

Pollard, Carl & Ivan Sag. 1994. *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. A comprehensive grammar of the English language. New York: Longman.

R Development Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org.

Schlueter, Zoe. 2017. *Memory retrieval in parsing and interpretation*. College Park, MD: University of Maryland Dissertation.

Schlueter, Zoe, Alexander Williams & Ellen F. Lau. 2018. Exploring the abstractness of number retrieval cues in the computation of subject-verb agreement in comprehension. *Journal of Memory and Language* 99. 74–89. DOI: https://doi.org/10.1016/j.jml.2017.10.002

Sloggett, Shayne. 2013. Case licensing in processing: Evidence from German. Poster at the 26th CUNY Conference on Human Sentence Processing. University of South Carolina.

Smith, Garrett, Julie Franck & Whitney Tabor. 2018. A self-organizing approach to subject-verb number agreement. *Cognitive Science* 42. 1043–1074. DOI: https://doi.org/10.1111/cogs.12591

Tabor, Whitney & Sean Hutchins. 2004. Evidence for self-organized sentence processing: Digging in effects. *Journal of Experimental Psychology: Learning, Memory & Cognition* 30. 431–450. DOI: https://doi.org/10.1037/0278-7393.30.2.431

Tanner, Darren. 2011. *Agreement mechanisms in native and nonnative language processing: Electrophysiological correlates of complexity and interference.* Seattle, WA: University of Washington Dissertation.

Tanner, Darren, Janet Nicol & Laurel Brehm. 2014. The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language* 76. 195–215. DOI: https://doi.org/10.1016/j.jml.2014.07.003

Townsend, David J. & Thomas G. Bever. 2001. *Sentence comprehension: The integration of habits and rules.* MA: MIT Press.

Tucker, Matthew A., Ali Idrissi & Diogo Almeida. 2015. Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology* 6. 1–21. DOI: https://doi.org/10.3389/fpsyg.2015.00347

Tucker, Matthew A. & Diogo Almeida. 2017. The complex structure of agreement errors: Evidence from distributional analyses of agreement attraction in Arabic. In Andrew Lamont & Katerina Tetzloff (eds.), *Proceedings of the 47th Meeting of the North East Linguistics Society*, 45–54. MA: GLSA.

Van Dyke, Julie A. & Brian McElree. 2007. Similarity-based proactive and retroactive interference reduces quality of linguistic representations. Poster presented at the CUNY Conference on Human Sentence Processing. San Diego.

Van Dyke, Julie A. & Brian McElree. 2011. Cue-dependent interference in comprehension. *Journal of Memory and Language* 65. 247–263. DOI: https://doi.org/10.1016/j.jml.2011.05.002

Van Dyke, Julie A. & Richard L. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49. 285–316. DOI: https://doi.org/10.1016/S0749-596X(03)00081-0

Vasishth, Shravan, Lena Jäger & Bruno Nicenboim. 2017. Feature overwriting as a finite mixture process: Evidence from comprehension data. In *Proceedings of MathPsych/ICCM*. Warwick.

Vasishth, Shravan, Sven Brüssow, Richard L. Lewis & Heiner Drenhaus. 2008. Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science* 32. 685–712. DOI: https://doi.org/10.1080/03640210802066865

Villata, Sandra, Whitney Tabor & Julie Franck. 2018. Encoding and retrieval interference in Sentence Comprehension: Evidence from Agreement. *Frontiers in Psychology* 9. 1–16. DOI: https://doi.org/10.3389/fpsyg.2018.00002

Wagers, Matthew W., Ellen F. Lau & Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61. 206–237. DOI: https://doi.org/10.1016/j.jml.2009.04.002

Xiang, Ming, Brian Dillon & Colin Phillips. 2006. Testing the strength of the spurious licensing effect for negative polarity items. New York: Talk at the 19th CUNY Conference on Human Sentence Processing.

Xiang, Ming, Julian Grove & Anastasia Giannakidou. 2013. Dependency dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology* 4. 1–19. DOI: https://doi.org/10.3389/fpsyg.2013.00708

Yanilmaz, Aydogan & John Drury. 2018. Intervening and non-intervening interference. Poster at the 31st CUNY Conference on Human Sentence Processing, UC-Davis.

Typesetting queries

1. Figures 11 and 12 are not cited within the text. Please advise where they should be cited.

2. The following items have been included within the reference list, but are not cited within the text. For each un-cited reference, please advise where it should be cited in the text, or confirm that it can be removed from the reference list.
   a. Ref no. "Eberhard, Kathleen M., J. Cutting & Kathryn Bock. 2005."
   b. Ref no. "Linzen, Tal & Brian Leonard. 2018."
   c. Ref no. "Murdock Jr, Bennet B. 1982."
   d. Ref no. "Smith, Garrett, Julie Franck & Whitney Tabor. 2018."

3. In Additional File section, we have maintained the placeholder for the title. Kindly provide the detail.