

Syntactic agreement attraction reflects working memory processes

L. Robert Slevc^a  and Randi C. Martin^b

^aDepartment of Psychology, University of Maryland, College Park, MD, USA; ^bDepartment of Psychology, Rice University, Houston, TX, USA

ABSTRACT

Does producing syntactic agreement rely on syntactic or memory-based retrieval processes? The present study investigated the extent to which syntactic processing deficits and working memory (WM) deficits predict susceptibility to *agreement attraction* [Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93], where speakers tend to erroneously produce plural agreement for a singular subject when another noun in the sentence is grammatically plural. Four brain-injured patients with varying degrees of grammatical and WM deficits completed sentences with local nouns that matched or mismatched in number with the head noun, and that were plausible or implausible subjects. Both aspects of grammatical deficits and the extent of WM deficits predicted the extent of agreement attraction effects. These data are consistent with the proposal that producing an agreeing verb involves a cue-based search in WM for an appropriate controlling noun, which is subject to interference from other elements in memory with similar properties [cf. Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, 56(1), 65–85. doi:10.1016/j.jml.2006.08.004].

ARTICLE HISTORY

Received 6 May 2015
Accepted 10 June 2016

KEYWORDS

Syntactic agreement;
language production; short-
term memory; aphasia

Syntactic agreement links together separate elements of language that are conceptually related (Bock, 1995). Most of the world's languages use at least some type of agreement (Mallinson & Blake, 1981) and agreement relations must generally be computed quite often; for example, adult English speakers produce number agreement at least once in every 16 words (Bock, 2011). Agreement processes are important not just because of their ubiquity but also because of agreement's place at the interface of meaning and syntax, thus the processing of syntactic agreement has important implications for our understanding of syntactic processing and the syntax-semantics interface.

As is true for most aspects of language production, our understanding of the psycholinguistics of agreement processing has drawn important insights from the kinds of errors that speakers make. For example, take the following quote from former US president G.W. Bush: "Then you wake up at the high school level and find out that the

illiteracy level of our children are appalling" (G.W. Bush, 23 January 2004).¹ Although the subject (*level*) is singular, the verb (*are*) is plural, presumably because of some sort of interference from the intervening plural object of the prepositional phrase (*children*). This type of *agreement attraction* has been noted for many years (e.g. Francis, 1986), and is supported by experimental evidence from error elicitation paradigms in production (Bock & Miller, 1991; Eberhard, Cutting, & Bock, 2005; Franck, Soare, Frauenfelder, & Rizzi, 2010; Vigliocco & Nicol, 1998; *inter alia*) and from comprehenders' relative insensitivity to agreement errors in attraction contexts (Staub, 2009; Wagers, Lau, & Phillips, 2009; cf. Pearlmutter, Garnsey, & Bock, 1999).

Most experimental work on agreement attraction relies on a sentence completion paradigm pioneered by Bock and Miller (1991), where participants are presented with a sentence preamble that they then repeat and complete. In general, speakers are relatively likely to erroneously produce a plural

CONTACT L. Robert Slevc  slevc@umd.edu

This article was originally published with errors. This version has been corrected. Please see Erratum (10.1080/20445911.2016.1214441)

¹Retrieved from <http://politicalhumor.about.com/od/bushquotes/a/topbushisms2004.htm>.

marked verb in sentences with a singular head noun when there is an intervening plural noun (though note that speakers only rarely erroneously produce a singular verb in sentences with a plural head noun). That is, when participants hear preambles like *The key to the cabinets...* they sometimes produce sentences like that in (1).

- (1) The key to the cabinets are missing. (Bock & Miller, 1991).

Because subject–verb number agreement codifies the link between a predicate and its subject, it is often viewed as a purely syntactic process (e.g. Chomsky, 1995; Franck et al., 2010; Levelt, Roelofs, & Meyer, 1999). By the *hierarchical feature passing* account, relevant agreement features simply percolate through the hierarchical syntactic structure from the subject noun phrase (NP) to the verb phrase, and errors such as (1) occur when the number feature from the plural local noun (cabinets) gets passed too far up the tree and overwrites the number feature from the head noun (Franck, Vigliocco, & Nicol, 2002; Vigliocco & Nicol, 1998). Evidence supporting this feature passing account comes from findings that the likelihood of errors is not affected by linear/surface distance, but instead is affected by the hierarchical relationship between an agreement target and local attractor noun. For example, Franck et al. (2002) contrasted preambles like (2a), where the potential agreement attractor (*presidents*) is farther from the verb in terms of linear order, but closer in terms of syntactic structure (i.e. is less deeply embedded), with preambles like (2b), where the potential attractor (*companies*) is linearly proximal but syntactically more removed.

- (2a) The threat to the presidents of the company ...
 (2b) The threat to the president of the companies ...

Erroneous plural agreement was, in fact, more likely in sentences like (2a) than in sentences like (2b), suggesting that agreement features are computed in a way constrained by hierarchical syntactic structure (Franck et al., 2002; see also related proposals from Franck et al., 2010; Franck, Lassi, Frauenfelder, & Rizzi, 2006; but see Gillespie & Pearlmutter, 2013).

A second influential framework – the *marking and morphing* account (Bock, Eberhard, & Cutting, 2004; Eberhard et al., 2005) relies on both syntactic and lexical mechanisms. By this account, the likelihood of attraction errors is related both to the “strength” of the number value of a NP, which is a function of the notional number of the NP as a whole and of

the morphological specification of individual elements within the NP, and by the hierarchical distance of the NP controller from potentially interfering material. That is, a verb takes the number of the subject NP via spreading activation through the hierarchical syntactic structure. However, the number feature of the subject NP is graded, which influences the likelihood of agreement errors. For example, a notionally plural but grammatically singular collective noun like “team” must reconcile a plural notional *marking* with a singular grammatical number from its constituent morphemes (via *morphing*). The number value of such a notionally plural collective noun is therefore less strongly singular than the number value of a NP with a notionally (and grammatically) singular head, and indeed collective head nouns do lead to increased agreement error rates (e.g. Bock, Nicol, & Cutting, 1999). Similarly, more agreement errors occur to fragments like “The label on the bottles” where several labels are implied (one on each bottle) than to fragments like “The baby on the blankets” where only one baby is implied (though note that this effect appears to be relatively weak in English; Bock & Miller, 1991; Eberhard, 1999). Because the *marking and morphing* model involves both syntactic and semantic influences on agreement, it gains support from influences of these notional factors on agreement production.

According to both the *hierarchical feature passing* account and the *marking and morphing* account, agreement is computed over the full syntactic structure of a sentence (or at least the entire subject NP). However, as Gillespie and Pearlmutter (2011) point out, this is difficult to reconcile with the incremental nature of production. This sort of observation leads to a different conception of agreement production, where the influence of agreement attractors depends on processing dynamics rather than syntactic or semantic factors per se. That is, agreement errors might reflect interference during the encoding and/or retrieval of the subject NP from working memory (WM), fitting with a growing body of literature linking general properties of memory retrieval dynamics to aspects of parsing (e.g. Harris, 2015; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; McElree, Foraker, & Dyer, 2003; Nicenboim, Vasishth, Gattei, Sigman, & Kliegl, 2015; see Van Dyke & Johns, 2012, for a review) and production (e.g. Martin & Freedman, 2001; Slevc, 2011; see Martin & Slevc, 2014, for a review).

One proposal along these lines is that agreement errors result from interference during encoding of

the subject number feature, which occurs when head and local nouns are planned in parallel (Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004). By this account, interference between simultaneously activated lexical items can cause the incorrect number feature to be encoded into memory, thus leading to a verb agreeing with that incorrect number (but see Veenstra, Meyer, & Acheson, 2015). Alternatively, agreement errors might happen at memory *retrieval*. Specifically, the influence of a local noun on agreement might reflect the extent to which it is confusable during retrieval of an agreement controller from WM (Badecker & Kuminiak, 2007; Badecker & Lewis, 2007; Wagers et al., 2009). Under this *cue-based retrieval* approach to agreement, the need to produce an agreeing verb triggers a search in content-addressable WM for the verb's controller, based on a set of retrieval cues generated during verb selection. Agreement attraction effects emerge when a local noun with similar features to the target controller is erroneously retrieved and selected as the target for agreement (Badecker & Kuminiak, 2007; Badecker & Lewis, 2007). That is, the retrieval of the agreement controller is susceptible to interference from other items in memory, just as in any other kind of memory retrieval task (Lewis & Vasishth, 2005). This account has been developed primarily in terms of syntactic effects on agreement (i.e. retrieval based on syntactic cues); however, it could presumably account for some types of semantic effects on agreement as well. For example, retrieval cues generated from a verb might be more likely to pick out a semantically plausible than implausible controller NP, and indeed agreement attraction errors are more likely when the attracting local NP is a plausible rather than implausible agent of the verb (Thornton & MacDonald, 2003).

These memory-based models of agreement gain support from the finding that memory limitations can affect the likelihood of producing agreement errors (Hartsuiker & Barkhuysen, 2006) and from the finding that elements that are more closely linked in processing (e.g. that are semantically integrated and so likely planned together) exert a greater influence on agreement (Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004; but see Veenstra et al., 2015). These memory-based accounts can also successfully explain patterns of verb agreement attraction in comprehension – that is, cases where agreement errors do not

disrupt reading due to the presence of an attracting noun (e.g. Lago, Shalom, Sigman, Lau, & Phillips, 2015; Tanner, Nicol, & Brehm, 2014; Wagers et al., 2009) – and have been used to model other types of parsing deficits in aphasia (e.g. Patil, Hanne, Burchert, De Bleser, & Vasishth, 2016).

Although structurally based and memory-based models of agreement production are conceptually quite distinct, there is little work comparing these types of theories (although note that there is a growing body of work on memory-based models of agreement *comprehension*; see, e.g. Lago et al., 2015; Staub, 2009; Tanner et al., 2014; Wagers et al., 2009). The aim of the present study is to directly contrast grammatically based and memory-based accounts of agreement production by investigating susceptibility to agreement errors as a function of grammatical deficits and WM deficits in aphasia. If agreement is a primarily syntactic process (as per the hierarchical feature passing or the marking and morphing accounts), then susceptibility to agreement attraction should be related to the extent of patients' syntactic processing deficits (as long as those deficits are related to agreement computation) and unrelated to the extent of patients' memory deficits. In contrast, if agreement errors reflect processes of memory encoding or retrieval, then the extent of patients' WM deficits should predict susceptibility of agreement attraction, even when syntactic processing is preserved.

Agreement in aphasia

Of course, this is far from the first study to examine agreement processing in aphasia. Indeed, problems with verb inflection are one of the hallmarks of agrammatic aphasia (e.g. Faroqi-Shah & Thompson, 2004; Rochon, Saffran, Berndt, & Schwartz, 2000). Perhaps surprisingly, then, individuals classified as agrammatic speakers actually tend to do relatively well at agreement inflection (especially compared to tense inflection; e.g. Clahsen & Ali, 2009; Friedmann & Grodzinsky, 1997). Susceptibility to agreement errors in speakers with aphasia does vary as a function of cognitive load (Kok, Van Doorn, & Kolk, 2007), suggesting that agreement production is not a purely syntactic process. There is, however, little data directly comparable to the psycholinguistic work on agreement in neurally healthy populations as only a small body of work has investigated performance for aphasic individuals with grammatical difficulties in production in the

sorts of agreement attraction paradigms discussed above (Hartsuiker, Kolk, & Huinck, 1999; Vigliocco, Butterworth, Semenza, & Fossella, 1994; Vigliocco & Zilli, 1999). Findings from these studies have been mixed. Vigliocco and Zilli (1999) found that the likelihood of grammatical gender agreement errors in two Italian-speaking Broca's aphasics was influenced by both grammatical and conceptual gender, suggesting preserved syntactic and semantic influences on agreement processing. However, both patients showed exaggerated attraction effects relative to controls. Vigliocco et al. (1994) looked at errors in number agreement while also manipulating the semantic factor of distributivity, and found that a patient classified as agrammatic showed normal sensitivity to both syntactic and semantic factors in agreement errors, whereas a conduction aphasic showed sensitivity to neither syntactic nor semantic factors (again, both patients produced more agreement errors than normal overall). Hartsuiker et al. (1999) found that a group of Broca's aphasics were as likely as controls to make agreement errors based on a syntactic manipulation (i.e. number mismatch between the head and local nouns) but were unaffected by the semantic factor of distributivity.

Hartsuiker et al. (1999) suggested that Broca's aphasics have a computational resource restriction which prevents their being able to take into account both syntactic and semantic factors when determining agreement. Because of this, patients may focus only on syntactic information, although it is not entirely clear why syntactic factors should trump semantic ones for patients who are presumed to have grammatical difficulties in production. This account is similar to a memory-based explanation if the relevant computational resource is memory; for example, one might assume that patients rely on impoverished retrieval cues that include only syntactic, and not semantic, information. Given that the left inferior frontal damage implicated in Broca's aphasia is also associated with deficits in WM and in interference resolution (e.g. Hamilton & Martin, 2005, 2007; Hoffman, Jefferies, & Lambon Ralph, 2011; Martin & Allen, 2008; Novick, Kan, Trueswell, & Thompson-Schill, 2009; Thompson-Schill

et al., 2002), it is plausible that the exaggerated attraction effects for these patients reflect memory, rather than grammatical, deficits.

This previous work on agreement attraction in aphasia has not evaluated WM specifically nor has it related the degree of syntactic deficit to the degree of agreement deficit. This past work has also included only Broca's aphasics (with the exception of Vigliocco et al., 1994), thus one cannot tell whether patients with other classifications and other production patterns might show similar or different difficulties in agreement. Moreover, even within the agrammatic Broca's aphasics, no attempt has been made to relate the degree of grammatical deficit to the pattern of agreement production. For most of the patients, only a global rating of agrammatism from a standardised aphasia battery was reported. Although Kok et al. (2007) provided more specific information on verb inflection errors for their patients, it appears that there was little or no relation between the degree of these errors and the degree of agreement errors under load conditions for their patients (see Tables 3 and 4 in Kok et al., 2007). Thus, it is not yet clear whether agreement processing is influenced by grammatical processing deficits, memory deficits, or both.

Experiment

The goal of the present experiment was to directly investigate if and how agreement attraction relates to grammatical processing deficits and WM deficits in aphasia. Specifically, we elicited the production of agreement errors in patients with varying degrees of syntactic deficits in production (as assessed by the Quantitative Production Analysis (QPA) scoring system; Rochon et al., 2000) and WM/short-term memory (STM) deficits (assessed with memory probe tasks; Martin, Shelton, & Yaffee, 1994).²

Accounts of agreement production based on syntactic processes predict that susceptibility to agreement attraction should be especially pronounced in those with the most severe grammatical deficits, at least when those grammatical deficits are related to the processes involved in determining

²Memory-based accounts of agreement generally implicate WM rather than STM, which are dissociable constructs (e.g. Engle, Tuholski, Laughlin, & Conway, 1999). However, the semantic STM deficits investigated here and in a large body of previous work (see, e.g. Martin, 2005, for a review) not only involve problems in maintenance (i.e. STM), but also processing deficits characteristic of control processes (Hamilton & Martin, 2005, 2007; Hoffman, Jefferies, & Lambon Ralph, 2011; Martin & Allen, 2008; Novick et al., 2009; Thompson-Schill et al., 2002). That is, these patients' deficits likely reflect both storage and processing components of memory and so can be considered to be a type of WM deficit. However, these patients have historically been referred to as "semantic STM" patients and so, for comparability with past work, we use the terms WM and semantic STM largely interchangeably in this manuscript.

agreement – that is, determining the hierarchical structure of the NP such that the correct head noun is identified and then selecting the correct verb morphology for that head. The QPA measures that would appear most relevant for these processes are the sentence elaboration index and the inflection index. The sentence elaboration index reflects the number of content words in the NP and verb phrase and thus relates to the structural complexity of these phrases. Note that this is an indirect measure of structural complexity (e.g. active and passive sentences could have the same number of content words per phrase despite clearly differing in syntactic complexity). Nevertheless, sentences with higher sentence elaboration scores are likely to be more syntactically complex. For instance, a sentence with a head NP with adjective modifiers would have a higher elaboration score and greater syntactic complexity than a head NP with only a bare noun and, similarly, a sentence with a verb phrase including a direct object and adverbial phrase would have a higher elaboration score and greater syntactic complexity than a verb phrase including only a direct object. The inflection index refers to the proportion of inflectable verbs that were inflected.³ Patients showing grammatical deficits in production might only show pronounced attraction effects for grammatical number and not show any semantic effects on agreement, supporting purely syntactic accounts (e.g. Franck et al., 2002), or might also show normal semantic effects, supporting models with both syntactic and notional influences on agreement (e.g. Eberhard et al., 2005).

On the other hand, memory-based accounts of agreement production (i.e. where agreement computation relies on cue-based retrieval processes; Badecker & Kuminiak, 2007) suggest that susceptibility to agreement attraction should be especially pronounced in patients who are highly susceptible to interference in WM, even if they do not show obvious grammatical deficits. That is, when a patient who experiences especially high levels of interference (and/or is impaired in resolving interference) produces an agreeing verb, he/she would likely be strongly influenced by other items in memory that partially match the retrieval cues for the subject noun, thus tend to show highly exaggerated attraction effects.

In the present study, patients' abilities to maintain phonological and semantic information (i.e. phonological and semantic STM) were assessed separately. Previous research has shown that these abilities dissociate (Allen, Martin, & Martin, 2012; Barde, Schwartz, Chrysikou, & Thompson-Schill, 2010; Martin & He, 2004) and that a semantic STM deficit, in particular, is associated with (or perhaps caused by) exaggerated susceptibility to interference in memory (Hamilton & Martin, 2005, 2007; *inter alia*). Semantic STM deficits have also been shown to cause particular difficulty for production of adjective-noun phrases (Martin & Freedman, 2001) and conjoined noun phrases (Freedman, Martin, & Biegler, 2004; Martin, Miller, & Vu, 2004). The association between a semantic STM deficit and these production deficits has been made by assuming that planning proceeds on a phrase-by-phrase basis with this planning occurring at the syntactic and lexical-semantic levels. The absence of an effect of a phonological STM deficit on these production tasks may result because phonological planning has a much smaller scope, possibly only a single phonological word (Wheeldon & Lahiri, 1997), which may be within the capacity of patients with even very reduced phonological capacities. If agreement attraction errors reflect interference from items in recent memory that partially match retrieval cues for the agreement controller (i.e. the cue-based retrieval account), then patients with lexical-semantic STM deficits are predicted to show exaggerated attraction effects given that such patients are highly susceptible to interference in memory. Phonological STM seems less relevant for cue-based retrieval processes as the relevant cues are unlikely to be targeting phonology, however, note that phonological STM might play some role in *encoding* the to-be-produced sentence preamble (e.g. for encoding the presence or absence of the plural marker /s/). Thus both types of STM deficits might relate to agreement production if agreement errors reflect competition at memory encoding (e.g. Gillespie & Pearlmutter, 2011).

This experiment examined agreement production in four individuals with aphasia, who varied in degree of syntactic deficits and STM/WM deficits, and in 24 older adult control participants. Participants performed an agreement production

³Another QPA measure, the embedding index, which reflects the proportion of sentences with an embedded structure would also be relevant to the issue of the ability to form hierarchical structure. However, because the range of the embedding index for control participants includes 0 and because it has low reliability in scoring, this measure is not included in Table 1.

Table 1. Grammatical production and short-term memory span data from BB, EV, MB, and SJ.

	BB	EV	MB	SJ	Control mean (range)
<i>QPA Indices</i>					
<i>Morphological content</i>					
Proportion of closed class words	0.39*	0.58	0.53	0.47	0.54 (0.47–0.61)
Determiner index	0.62*	0.96	1.00	0.97	0.99 (0.94–1.00)
Proportion of pronouns	0.06*	0.42	0.55	0.27*	0.41 (0.29–0.55)
Proportion of verbs	0.21*	0.49	0.55	0.47	0.48 (0.35–0.63)
Inflection index	0.75	1.00	1.00	1.00	0.92 (0.53–1.00)
Auxiliary complexity index	0.51*	1.47	1.49	1.08	1.26 (0.80–1.71)
Words per minute	35.00*	155.87	73.15*	95.48*	160.82 (107.44–232)
<i>Structural analysis</i>					
Proportion of words in sentences	0.44*	1.00	0.96	0.97	0.98 (0.84–1.00)
Proportion of well-formed sentences	0.60*	0.90	0.76	0.81	0.95 (0.75–1.0)
Sentence elaboration	1.94*	2.01*	2.75	1.65*	3.06 (2.14–4.06)
Mean utterance length	4.42*	7.29	7.95	5.81*	8.17 (6.5–10.5)
<i>STM measures (span)</i>					
Category probe (Semantic STM)	1.50*	1.80*	2.46*	2.38*	5.38 (3.4–7.0)
Rhyme probe (Phonological STM)	3.34*	3.34*	5.00*	3.00*	7.02 (5.8–9.0)

* indicates scores outside the normal range (from Allen et al., 2012; Rochon et al., 2000)

task that manipulated both syntactic attraction – whether a local noun matched or mismatched with the subject noun – and semantic attraction – whether the local noun was a plausible or implausible agent for the sentence. The manipulation of semantic attraction was based on Thornton and MacDonald's (2003) findings that agent/verb plausibility influenced attraction rates; however, the present task manipulated whether a post-verbal adjective was a plausible or implausible modifier for the local noun (see below).

Method

Participants

Four patients (identified with subject codes) with good single-word processing and who could successfully perform the agreement production task (see below) participated in the study. In addition, 24 control participants (8 males) were recruited from a pool of older adults, ranging in age from 50 to 75 and with at least a high school education, who regularly participate in experiments at Rice University.

Patient description: BB. At the time of testing, BB was a 49-year-old man with a graduate-level education who suffered a left hemisphere cerebrovascular accident (CVA) five years prior to testing. A MRI scan revealed a left hemisphere lesion that included inferior frontal gyrus and lateral temporal cortex, extending into parietal cortex. BB performed reasonably well on tests of single-word processing, scoring 69% correct on the Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996) and scoring 94% correct on a picture-word matching task that included phonological and semantic

distractors (Martin, Lesch & Bartha, 1999). As can be seen in Table 1, BB had highly impaired semantic and phonological STM spans as assessed with the category probe and rhyme probe tasks (Martin et al., 1994; see below for task details). BB also had markedly impaired grammatical production as assessed with QPA (Saffran, Berndt, & Schwartz, 1989), as he scored below the normal range (from Rochon et al., 2000) on almost all of the morphological and structural measures reported in Table 1.

Patient description: EV. Patient EV was a 53-year-old woman at the time of testing with a college education who suffered a left hemisphere CVA 10 years prior to testing. A MRI scan showed an infarction in left inferior frontal cortex, constrained mostly to the pars triangularis (BA 45), as well as a small lesion in the left middle frontal gyrus. Like BB, EV performed relatively well on tests of single-word processing, scoring 85% on the PNT (Roach et al., 1996) and 95% correct on the picture-word matching task (Martin et al., 1999), but had impaired semantic and phonological STM spans according to the category and rhyme probe tasks (Table 1). Unlike BB, EV showed no evidence of agrammatism, with all but one score on the QPA within the normal range (the exception being the sentence elaboration index) and most scores near or even above the mean of control participants' scores (see Table 1).

Patient description: MB. Patient MB was a 60-year-old man at the time of testing who had completed approximately one year of college coursework and who suffered a left hemisphere CVA six years prior to testing. Structural MRI revealed a left temporal-parietal lesion, including damage to the left parietal and superior temporal lobes, plus some damage to the left posterior insula. MB had good single-word

processing, scoring 96% correct on the PNT (Roach et al., 1996) and 98% correct on picture-word matching (Martin et al., 1999). MB had an impaired semantic STM span but relatively preserved phonological STM, and showed little evidence of agrammatism, with only the number of words produced per minute below the range for controls (see Table 1).

Patient description: SJ. Patient SJ was a 61-year-old woman at the time of testing who had completed one year of college and who suffered a left hemisphere CVA three years prior to testing. Structural MRI revealed a left temporal–parietal infarction with damage to a large portion of her left temporal lobe (for more detail, see Baum, Martin, Hamilton, & Beauchamp, 2012). SJ performed well on tests of single-word processing, scoring 97% correct on the PNT and 97% on the picture-word matching task. SJ had an impaired STM span on both category and rhyme probe tasks and showed some evidence of agrammatism, with scores below the normal range on 4 of the 11 indices of the QPA (see Table 1). Notably, however, her inflection index (i.e. proportion of inflectable verbs inflected) was at ceiling (1.00), though her structural elaboration index was the lowest of the four patients.

Materials and procedure

Memory and syntactic production measures. Memory probe tasks (Martin et al., 1994) were chosen to assess memory span because they do not require list output and so performance is not impacted by difficulties in language production. In these tasks, participants heard a list of pre-recorded words separated by a 500 ms delay and then, after a one second delay, heard a probe word while seeing “???” on the screen. In the category probe task, participants judged whether the probe word was in the same category as any of the items in the list, and in the rhyme probe task, participants judged whether the probe word rhymed with any items on the list. Lists started with two words and progressed up to seven words in length, however, the task was terminated when accuracy dropped below 75%.

Extent of agrammatism was evaluated by analysing patients’ narrative speech using the QPA (Saffran et al., 1989). Specifically, patients were shown a picture book of the story *Cinderella* (without words) and were then asked to tell the story aloud without reference to the book. Stories were recorded, transcribed, and analysed for lexical, morphological, and structural measures (for more detail

on the QPA, see Rochon et al., 2000; Saffran et al., 1989).

Agreement production task. The critical items for the agreement production task consisted of 44 sentence preambles of the form [head NP] [preposition] [local NP] (e.g. *The box by the window...*), each paired with a single adjective that was always a plausible modifier of the head NP (e.g. *open* was selected for the item headed by *the box*). For each item, two local NPs were selected: one that could be plausibly modified by the adjective (*the window* can plausibly be open) and one that could not (*the table* is unlikely to be open). These two local nouns were matched as closely as possible on lexical frequency, number of syllables, and length, as calculated from the English Lexicon Project (Balota et al., 2007). Each local NP occurred both in singular and plural forms (note that the head NP in these critical items was always singular). There were thus four versions of each item resulting from crossing the plausibility and number of the local NP, which were counterbalanced across four lists such that each item appeared in every condition across lists and each list contained only one version of each item. (See appendix for the experimental items.) In addition, 44 filler preamble-adjective pairs were created, all with plural head NPs. Half of these filler items contained only one NP (i.e. only a head NP without a local NP) to provide some easier trials. The other 22 filler items had either a singular or plural local NP (11 items each). Thus the entire set of 88 stimuli had equal numbers of singular and plural head NPs, equal numbers of singular and plural local NPs, and equal numbers of preambles with matching and mismatching NP number. Filler and critical items were put into fixed random orders and interleaved such that filler and critical items alternated.

The agreement production task was administered with PsyScope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). On each trial, participants saw an adjective displayed in 48-point Helvetica font on a computer screen (e.g. *open*), and read this word aloud. The experimenter then read aloud the sentence preamble (e.g. *The box near the window...*), and participants repeated the preamble, added an auxiliary verb and finished with the initially presented adjective (*The box near the window was open*). The experimenter would repeat the preamble when asked. Before beginning the experimental trials, participants were given instructions by the experimenter (which were also printed on the

computer screen), and then did four practice trials on which feedback was given when necessary. Control participants were tested as above, except they were asked to try to complete their sentences within a two-second time limit (where the remaining time on each trial was indicated by a row of gradually disappearing Xs). This time pressure was included in hope of increasing the number of errors. The four patients were tested four times: once on each list, with at least one week separating each session. Control participants were tested only once, with equal numbers of participants tested on each list.

Design and analysis

Participants' utterances were digitally recorded, transcribed, and coded as containing a singular verb (*is* or *was*) or plural verb (*are* or *were*). Trials were excluded when the produced sentence did not use the target adjective, did not preserve the grammatical number of the nouns from the preamble, or otherwise deviated from the preamble in such a way that it changed the correct subject number marking (e.g. cases where "The box near the windows" was repeated as "The box and the windows"). These criteria led to the exclusion of 18.4% out of all trials from BB, 12.2% of trials from EV, 22.7% of trials from MB, 24.7% of trials from SJ and 11.2% of trials from control participants.⁴

Data from control participants were analysed in two ways. First, 2×2 repeated measures ANOVAs were fit to aggregated data with fixed effects factors for *local number* (the grammatical number of the local noun: singular or plural) and *local plausibility* (the plausibility of the local noun as the subject of the sentence: plausible or implausible), and with participants ($F1$) and items ($F2$) as random variables. Variability for these ANOVAs is reported with repeated measures 95% confidence-interval (CI) half-widths based on single degree-of-freedom comparisons (Loftus & Masson, 1994). These aggregate analyses were conducted both on untransformed proportions and on arcsine square root transformed proportions; for ease of interpretation, analyses and figures report untransformed proportions. In addition, a logistic linear mixed effects model was fit to the data (Bates, Machler, Bolker, & Walker, 2014; Jaeger, 2008) using the lme4

Table 2. Mixed effects model analysis for control participants, examining effects of the grammatical number and plausibility of local nouns.

Parameters	Fixed effects			Random effects	
	Estimate	SE	Z	By subjects SD	By items SD
Intercept	-4.69	0.56	-8.32*	1.28	0.00
Local number	3.54	0.87	4.07*	1.52	0.00
Local plausibility	0.01	0.74	0.01	0.00	0.00
Number \times plausibility	0.01	1.49	0.01	0.65	0.00

Notes: Correlations between random effect terms had to be removed for the model to converge. Participants' utterances were coded as 0 for accurate agreement and 1 for agreement errors. *Local Number* was coded as $-.5$ for singular and $.5$ for plural, and *Local Plausibility* was coded as $-.5$ for implausible and $.5$ for plausible. The model formula (with uncorrelated random effects) was: Utterance Code \sim Local Number * Local Plausibility + (1 + Local Number + Local Plausibility || Subject) + (1 + Local Number + Local Plausibility || Item). * p

package (version 1.1-8) in the R statistical software (version 3.2.1). This model, reported in Table 2, included the same fixed-effects structure (*local number* crossed with *local plausibility*) and the maximal random effects structure for both participants and items (Barr, Levy, Scheepers, & Tily, 2013), except that correlations between random effect terms had to be removed in order for the model to converge.

Effects of local number and of local plausibility for each patient were compared to the effects shown by control participants using Crawford and Howell's (1998) modified *t*-test along with the estimated effect size for the difference between patients and controls (z_{cc} ; an estimate of the average difference, in standard deviations, between a case's score and a randomly chosen control participant's score; Crawford, Garthwaite, & Porter, 2010). Agreement error rates for all conditions (including filler trials) are reported in Table 4.

Results and discussion

Control participants

Figure 1 shows the proportion of erroneous plural verbs produced by control participants (recall that all critical trials had singular head nouns) as a function of the grammatical number of the local noun and of the plausibility that the local noun could be modified by the end-of-sentence adjective. Participants showed a significant grammatical attraction effect, erroneously producing a plural verb on

⁴EV, MB, and SJ had more trials excluded from the plural local NP than from the singular local NP conditions (most strikingly, MB did not correctly produce the preamble for 57% of the locally-plural trials, in comparison to only 6% of the locally singular trials). In contrast, BB's exclusions were equal for sentences with plural and with singular local nouns. Excluded trials showed no consistent pattern as a function of local plausibility.

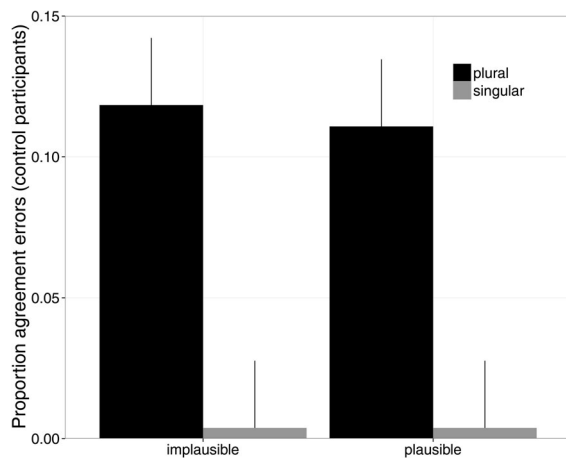


Figure 1. Proportion of erroneous plural verb agreement on critical trials for control participants as a function of the grammatical number and plausibility of the local noun. Data are plotted as untransformed proportions and error bars represent 95% confidence intervals (Loftus & Masson, 1994).

11.5% of critical trials with a plural local noun but only on 0.4% of critical trials with a singular local noun (a significant main effect of local number: $F(1, 23) = 12.35$, $CI = \pm 4.6\%$, $p < .01$; $F(1, 43) = 50.12$, $CI = \pm 2.8\%$, $p < .001$), but showed no effect of local plausibility and no interaction (all $F_s < 1$, ns). The logistic mixed effect model revealed the same pattern of effects (Table 2).

Although there is little work investigating agreement production in older adults, these grammatical attraction rates are in line with those previously reported in college-aged participants (see, e.g. meta-data reported in Eberhard et al., 2005). In contrast, errors were no more likely when the local noun was a semantically plausible subject than when it was an implausible subject for the sentence. This is surprising given that effects of plausibility on agreement errors have been documented with a relatively similar manipulation (Thornton & MacDonald, 2003). The lack of a plausibility effect in these data likely results from the specifics of the procedure used here. In particular, while Thornton and MacDonald (2003) manipulated the plausibility of a to-be-included verb (e.g. by having participants complete the fragment “*The album by the composers...*” with a verb like *played* versus a verb like *praised*), the current paradigm manipulated the plausibility of a following adjective. It may be the case that participants here could incrementally produce the preamble and main verb with little concurrent activation of (and thus influence from) the final adjective, whereas the tight relationship between the

number-bearing auxiliary verb and main verb in Thornton and MacDonald’s (2003) task would be more likely (or even obligatorily) planned together (cf. Solomon & Pearlmutter, 2004). Relatedly, the local syntactic coherence between the auxiliary and main verb is likely stronger than that between the main verb and following adjective (cf. Tabor, Galantucci, & Richardson, 2004), and so the plausibility of locally coherent local noun plus verb phrases in Thornton and MacDonald’s (2003) task might exert relatively large effects on agreement processing compared to the verb-adjective pairs used here. While unexpected, this null effect does not preclude the comparison of these data with the patient data reported below.

Patients

Figure 2 shows the proportion of erroneous plural verbs produced by the four patients, as well as the control data from above for comparison (note the different scale from Figure 1), as a function of the grammatical number of the local noun and of the plausibility that the local noun could be modified by the end-of-sentence adjective. As shown in the figure, all patients but MB showed very large grammatical attraction effects but, like controls, little effect of the plausibility manipulation. Statistical results, reported in Table 3, confirmed these impressions. MB’s error rates were within the range of the control group, and his attraction effect did not significantly differ from the control group. In contrast, BB, EV, and SJ showed a very large effect of the grammatical number of the local noun (44.3%, 49.2%, and 68.6% attraction effects, respectively); which all were significantly greater than the attraction effect of control participants. In contrast, the effect of the plausibility of the local noun did not significantly differ from the (null) effect shown by the control group for any patient.

If agreement attraction reflects a failure in a purely syntactic process (e.g. Franck et al., 2002) then extent of agrammatism should predict susceptibility to agreement attraction. However, we found little evidence that this is the case. Patient BB was clearly the most impaired on the QPA measures but yet showed an attraction effect similar in size to that of EV, who performed at a normal level on the QPA. Specifically, a two (patient) by two (local number) ANOVA treating patient as a fixed effect and item as a random effect revealed no evidence for a main effect of patient ($F(1, 37) = 1.40$, $MSE = 0.12$, ns) nor for an interaction of patient and local number ($F(1,$

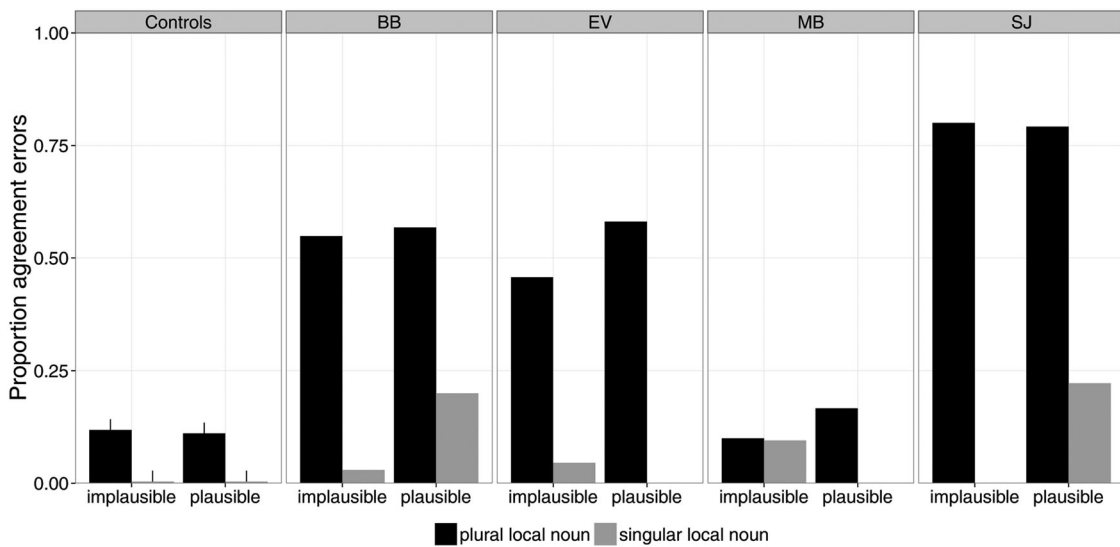


Figure 2. Proportion of erroneous plural verb agreement on critical trials for control participants (left panel; the same data as is shown in Figure 1) and the four patients as a function of the grammatical number and plausibility of the local noun.

Table 3. Statistical results comparing each patient's attraction effect to the control group.

Patient	Attraction effect (%)	Significance test ^a		Estimated effect size (Z_{cc}) ^b	
		<i>t</i>	<i>p</i>	Z_{cc}	(95% CI)
BB	44.30	2.14	.022	2.18	(1.43–2.92)
EV	49.20	2.45	.011	2.51	(1.68–3.32)
MB	8.34	–0.16	.44	–0.17	(–0.57–0.24)
SJ	68.60	3.69	.0006	3.77	(2.61–4.92)

^aCrawford and Howell (1998); *p* values are for one-tailed tests.

^bCrawford et al. (2010).

37) = 0.36, MSE = 0.074, *ns*). Similarly, although BB was considerably more impaired than SJ on the QPA measures, his attraction effect was numerically somewhat smaller and not statistically different from SJ's effect in an analysis across items (as above, there was no effect of patient ($F(1, 29) = 1.55$, MSE = 0.086, *ns*) or patient by local number interaction ($F(1, 29) = 1.11$, MSE = 0.12, *ns*)).⁵ There is thus no evidence that BB differed from EV or SJ in susceptibility to grammatical attraction from the local noun.

In terms of more specific aspects of grammatical difficulty in production reflected in the QPA, no relation is evident between the inflection index and the size of the attraction effect. EV and SJ scored 1.00 (at ceiling) on the inflection index, yet showed large attraction effects that were at least as large as those shown by BB, who had a lower inflection index score (although note that his score of .75 was

still within the range of the control participants reported by Rochon et al., 2000). The sentence elaboration index, on the other hand, does show some relation to the size of the attraction effect: SJ performed the worst on the sentence elaboration index and showed the largest attraction effect whereas MB scored the highest on sentence elaboration and showed the smallest attraction effect. BB and EV were more impaired than MB and closer to SJ on the sentence elaboration index and showed large attraction effects.

With regard to distinguishing between the grammatical and WM account, however, the interpretation of the relationship between attraction effects and the inflection and sentence elaboration indices is not entirely clear-cut. The inflection index reflects proportion of verbs inflected, but not whether those inflections are correct. Unfortunately, the raw data from QPA transcriptions also provide little relevant evidence: In the QPA, participants tell a story (in these cases, the story of Cinderella) and most often use past tense verbs, where the same form is used for singular and plural (e.g. “was”, “danced”, “rode”). Consequently, patients might have difficulties with number agreement, but not tense, which would not be evident in their stories. With respect to the sentence elaboration index, previous studies have demonstrated that patients with reduced semantic STM capacity have difficulty producing

⁵Note that 6 items had to be excluded from the comparison of BB and EV and 14 items from the comparison of BB and SJ due to missing values.

complex noun phrases consisting of two conjoined nouns (Freedman et al., 2004; Martin et al., 2004) or a noun preceded by adjectives (Martin & Freedman, 2001). The argument from these earlier findings is that sentence planning proceeds at a phrasal level for lexical–semantic representations and hence patients with reduced STM capacity have difficulty producing phrases with several content words. Thus while it is possible that sentence elaboration deficits contribute to the likelihood of agreement attraction, it might also be that both the reduced sentence elaboration measures and the large attraction effects are due to underlying WM deficits.

Another means of addressing patients' ability to produce agreement and to create hierarchical structure in production is to examine their performance on the trials where plural head nouns were presented either alone or with singular vs. plural attractors. Appropriate plural production in sentences with a plural head noun alone would provide support for the notion that the patients can choose the appropriate inflection in terms of number agreement. And if the patients show a smaller attraction effect for plural head nouns than for singular, this would suggest that they are able to structure the complex noun phrases appropriately. That is, one well-attested morphosyntactic effect on agreement production is that attraction is considerably more likely to occur when an attractor is grammatically plural than when grammatically singular. This is often described as an effect of plural markedness, where the marked plural form exerts relatively strong attraction (Bock & Miller, 1991; Eberhard, 1997). This asymmetry is syntactic in nature, as shown, for example, by findings that notionally plural but grammatically singular attractors do not enhance attraction effects in verb agreement (Bock and Eberhard 1993; Bock et al. 2001), thus, if patients show the typical singular/plural asymmetry in production, this would argue that they are able to structure the noun phrases appropriately and show a typical sensitivity to grammatical markedness.

Although these experiments were not designed for this comparison, the filler items did include preambles with grammatically plural head nouns alone and with both singular and plural local nouns (e.g.

Table 4. Agreement error rates as a function of the grammatical number of head and local NPs.

	Head NP	Local NP			Attraction effect (%)
		Singular (%)	Plural (%)	(none) (%)	
Controls	Singular	0.4	11.3	0.2	10.9
	Plural	3.6	0.0		3.6
	(fillers)				
BB	Singular	11.6*	55.9*	32.0*	44.3*
	Plural	61.5*	27.8*		33.8*
	(fillers)				
EV	Singular	2.3	51.5*	1.2	49.2*
	Plural	23.8*	7.4*		16.4
	(fillers)				
MB	Singular	4.8*	13.2	2.6*	8.3
	Plural	2.4	0.0		2.4
	(fillers)				
SJ	Singular	11.0*	79.6*	7.9*	68.6*
	Plural	16.7	0.0		16.7%
	(fillers)				

Note: For patients, * indicates scores significantly different than controls, as assessed with Crawford and Howell's (1998) modified *t*-test.

The answers to the homework...). For the plural head nouns alone, the controls produced very few agreement errors (0.2%). For the patients, only patient BB showed a large proportion of agreement errors, indicating that he had greater difficulty than the rest in choosing correct agreement morphology even in the simplest condition in this experiment.

With respect to the attraction effect for plural head nouns, control participants did occasionally produce errors like *The answers to the homework was wrong*; reflected in a significant difference between error rates in plural/singular and plural/plural preambles ($b = 0.035$, $SE = 0.015$, $t = 2.36$),⁶ however, this 3.6% effect was notably smaller than the approximately 11% attraction effect from plural local nouns in sentences with singular controllers, reflected in an interaction between the grammatical number of the head NP and the match/mismatch of the grammatical numbers of the head and local NP ($b = 0.058$, $SE = 0.20$, $t = 2.88$). Control participants thus showed the expected singular/plural asymmetry.

EV's, SJ's, and MB's attraction effect (Table 4) to singular local nouns with plural head nouns did not differ significantly from control participants' effects (EV: $t = 1.78$, *ns*, $z_{cc} = 1.82$; SJ: $t = 1.82$, *ns*, $z_{cc} = 1.86$; MB: $t = -0.18$, *ns*, $z_{cc} = -0.18$) however, BB's 33.8% singular attraction effect was significantly greater than controls' ($t = 4.21$, $p < .001$, $z_{cc} = 4.30$, z_{cc} CI = 3.00–5.59) and the proportion was similar

⁶Note that there were only half as many filler items with a plural head and a (singular or plural) local noun as critical items with a singular head NP (the other half of the filler items did not include a local noun) and that the grammatical number of the local noun was not counterbalanced across items. Because of this, these statistical comparisons relied on logistic mixed effects models with participants and items treated as crossed random effects.

(though somewhat lower) to that for singular head nouns. Thus, BB, unlike the other three patients, may have difficulty structuring the NP appropriately, resulting in a tendency to use the nearest noun in determining agreement. These data for the filler trials should be treated with some caution, however, as the identity of the attractor nouns was not matched the singular and plural conditions.

While BB's large attraction effects may result, at least in part, from grammatical impairments, it would be hard to make that claim for EV and SJ, given the absence of grammatical difficulties on the QPA for EV and the relatively normal performance of both patients on the trials with plural head nouns. With respect to the memory-based hypothesis, EV and SJ were impaired on the phonological STM task (rhyme probe) whereas MB performed at a near-normal level. In contrast, all three performed below the range of the control group on the semantic STM task (category probe). Thus, the most suggestive relation is between agreement attraction and phonological STM, rather than with semantic STM as hypothesised. According to the cue-based retrieval account of agreement attraction, preparing to produce a verb involves the retrieval of a representation for the head noun, while the local noun provides interference. The ability to retrieve the lexical/semantic representation of the head noun (including the fact that it was plural) would certainly aid in determining the correct agreement inflection. However, the ability to accurately encode the head noun, and/or to retain and retrieve a phonological representation of the head noun (in both the experimenter's and the patients' production of the preamble) could certainly help as well, in that patients could use this representation to re-compute singular or plural status before determining the verb inflection. In this view, preserved phonological STM (as for MB) helps to make up for deficits in the ability to retrieve lexical/semantic information. Another possibility is that the relevant impairment for EV and SJ is in some mechanism that biases the selection of one representation from competing representations, making them overly sensitive to competition from the local noun. Indeed there is mounting evidence suggesting that semantic STM deficits are related to (or perhaps arise from) such a deficit in overcoming interference in memory (Hamilton & Martin, 2005, 2007; Hoffman et al., 2011; Martin & Allen, 2008; Novick et al., 2009; Thompson-Schill et al., 2002).

Conclusions

The present results are most consistent with memory-based accounts of agreement production, rather than syntactically based accounts. Although it is likely that one patient's exaggerated agreement effect was due, at least in part, to a disruption in grammatical processes involved in computing the hierarchical structure of a phrase, such computations are necessary for the encoding of agreement features into memory and/or for cue-based retrieval of agreement information to succeed. That is, in the cue-based retrieval model of agreement (Badecker & Kuminiak, 2007), it is necessary that individuals be able to compute the syntactic and semantic features of words and phrase as they are processed. Cue-based retrieval depends on the existence of these features when cues generated from the verb are used to retrieve the head noun. Thus, if the features are not generated correctly, retrieval is bound to fail. However, beyond these computations, the cue-based retrieval account further implies that representations for words that have some overlap with the features matching the cues will provide interference. We would argue that EV and SJ have difficulty overcoming interference from non-target items in memory that overlap partially with the search cues for the appropriate controller, thus have trouble retrieving the correct number information when producing an agreeing verb. This difficulty in overcoming interference could also impact the ability to accurately *encode* agreement-relevant information into memory (cf. Gillespie & Pearlmutter, 2011; Solomon & Pearlmutter, 2004). Although the data presented here cannot distinguish whether these problems occur at encoding or retrieval (or both), they do show that memory limitations (plausibly arising from problems resolving interference in memory) impact agreement processing, lending support to both types of memory-based theories.

Still, the relationship between this deficit in resolving interference in memory (leading to what has often been called a semantic STM deficit; e.g. Martin, 2005) and susceptibility to agreement attraction is not completely straightforward. In particular, these data do not support the possibility that there is a simple lower limit of semantic STM required to support agreement production, given the good performance of MB whose semantic STM was similar to that of SJ, who showed large attraction effects. Instead, phonological STM appeared to be the more distinguishing factor, perhaps because

accurate phonological STM allowed review of a verbatim record of the NP. Another possibility is that the category probe task only indirectly captures patients' ability to manage interference in WM, and so a more fine-grained measure of cue-based retrieval ability might more clearly distinguish the interference-resolution capabilities of MB and SJ.

Somewhat surprisingly, the patients (and control participants) tested here showed no effect of local plausibility. That is, agreement errors were no more likely when the local noun (and potential attractor) was a plausible subject for the sentence (e.g. *The box near the windows was/were open*) than when it was an implausible subject (e.g. *The box near the tables was/were open*). This is surprising given other evidence for semantic effects on agreement (see Bock & Middleton, 2011), even with similar manipulations (Thornton & MacDonald, 2003). On one hand, this might indicate that STM processes are not responsible for semantic effects on agreement. Indeed this sort of cue-based memory retrieval account may not obviously capture semantic or notional effects on agreement production (see Bock & Middleton, 2011, for discussion), so one might take these data to support some non-STM-based process underlying semantic effects on agreement. However, it seems likely that this null effect reflects limitations of the materials used here. Thornton and MacDonald (2003) found greater attraction effects following preambles like *The album by the composers* when the main verb was a plausible predicate for the local noun (*praised*) compared to when it was implausible predicate (*played*). The manipulation used here was similar, except the plausibility was a function of the fit between the local noun and the final adjective (see above). If participants produced these sentences incrementally, they may have produced the number-bearing verb with little concurrent planning/activation of the final adjective. In contrast, the number-bearing auxiliary verbs and main verbs in Thornton and MacDonald's (2003) materials were likely planned together (cf. Solomon & Pearlmutter, 2004) and so more likely to affect agreement processes. Thus, given these limitations of the experimental materials, these data are unlikely to directly inform the relationship between memory retrieval mechanisms and semantic effects on agreement production.

However, these results do inform syntactic attraction effects on agreement production. A memory retrieval approach to agreement production has, until now, been tested only in terms of grammatical gender agreement (Badecker & Kuminiak, 2007) or, indirectly, by looking at error rates as a function of a

memory load (Hartsuiker & Barkhuysen, 2006; cf. Fayol, Largy, & Lemaire, 1994). In all of these previous experiments, the production tasks were paired with a concurrent memory load, however, it has been suggested that this sort of dual task could artificially increase reliance on memory processes (Bock & Middleton, 2011). The experiments reported here, in contrast, measure agreement production in a standard fragment completion task without external memory manipulations. Instead, we relied on performance of individuals with deficits in memory processing, who are unlikely to rely more on memory processes than unimpaired speakers (cf. evidence that patients with phonological STM deficits rely less on phonological and more on visual STM to perform memory tasks; e.g. Campbell & Butterworth, 1985; also evidence for compensatory processes involved in other types of memory decline; e.g. Buckner, 2004).

In sum, these data are consistent with a role of WM processes in agreement production and add support to memory-based models of agreement production (Badecker & Lewis, 2007; Badecker & Kuminiak, 2007; Gillespie & Pearlmutter, 2011). Of course, we cannot rule out the possibility that the patients have some other deficit that makes them rely on memory resources more so than is the case for healthy individuals. However, we have provided evidence against the most likely other explanation – that is, that agreement errors for all patients result from a deficit in syntactic processing per se. These data also complement a growing body of work linking agreement attraction effects in comprehension to cue-based retrieval processes (Lago et al., 2015; Staub, 2009; Tanner et al., 2014; Wagers et al., 2009). Such a cue-based retrieval account of agreement production is appealing as it relies on well-understood memory mechanisms that underlie multiple aspects of cognition (Lewis et al., 2006) and is compatible with the well-accepted idea that we produce sentences incrementally (cf. Gillespie & Pearlmutter, 2011). Of course, our understanding of the role WM plays in agreement processing is not complete; there are several aspects of agreement processing that do not yet have a straightforward explanation in terms of memory mechanisms. For example, the role of semantic factors in agreement production (see Bock & Middleton, 2011) and differences between pronoun and verb agreement (e.g. Bock et al., 2004) do not obviously emerge directly from a cue-based retrieval account. Nevertheless, these data bring us closer to an account of agreement production as a WM process and a better

understanding of the interface between memory systems and sentence production more generally.

Acknowledgements

We thank A. Cris Hamilton and Simon Fischer-Baum for helpful comments and advice, Corinne Allen, Kelly Banneyer, and Sanam Jivani for assistance with data collection, and the patients and their families for their participation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported in part by NIH [grant number F32 DC-008723].

ORCID

L. Robert Slevc  <http://orcid.org/0000-0002-5183-6786>

References

- Allen, C. M., Martin, R. C., & Martin, N. (2012). Relations between short-term memory deficits, semantic processing, and executive function. *Aphasiology*, 26(3–4), 428–461. doi:10.1080/02687038.2011.617436
- Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, 56(1), 65–85. doi:10.1016/j.jml.2006.08.004
- Badecker, W., & Lewis, R. (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 Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, ... Treiman, R. (2007). The English Lexicon project. *Behavior Research Methods*, 39, 445–459. doi:10.3758/BF03193014
- Barde, L. H. F., Schwartz, M. F., Chrysikou, E. G., & Thompson-Schill, S. L. (2010). Reduced short-term memory span in aphasia and susceptibility to interference: Contribution of material-specific maintenance deficits. *Neuropsychologia*, 48(4), 909–920. doi:10.1016/j.neuropsychologia.2009.11.010
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Baum, S. H., Martin, R. C., Hamilton, A. C., & Beauchamp, M. S. (2012). Multisensory speech perception without the left superior temporal sulcus. *NeuroImage*, 62(3), 1825–1832. doi:10.1016/j.neuroimage.2012.05.034
- Bock, J. K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language*, 51, 251–278. doi:10.1016/j.jml.2004.04.005
- Bock, J. K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language*, 40, 330–346. doi:10.1006/jmla.1998.2616
- Bock, K. (1995). Producing agreement. *Current Directions in Psychological Science*, 4, 56–61. doi:10.1111/1467-8721.ep10771165
- Bock, K. (2011). How much correction of syntactic errors are there, anyway? *Language and Linguistics Compass*, 5(6), 322–335. doi:10.1111/j.1749-818X.2011.00283.x
- Bock, K. & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, 8, 57–99. doi:10.1080/01690969308406949
- Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, 43, 83–128. doi:10.1006/cogp.2001.0753
- Bock, K., & Middleton, E. L. (2011). Reaching agreement. *Natural Language & Linguistic Theory*, 29(4), 1033–1069. doi:10.1007/s11049-011-9148-y
- Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93. doi:10.1016/0010-0285(91)90003-7
- Buckner, R. L. (2004). Memory and executive function in aging and AD: Multiple factors that cause decline and reserve factors that compensate. *Neuron*, 44, 195–208. doi:10.1016/j.neuron.2004.09.006
- Campbell, R., & Butterworth, B. (1985). Phonological dyslexia and dysgraphia in a highly literate subject: A developmental case with associated deficits of phonemic processing and awareness. *The Quarterly Journal of Experimental Psychology Section A*, 37(3), 435–475. doi:10.1080/14640748508400944
- Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.
- Clahsen, H., & Ali, M. (2009). Formal features in aphasia: Tense, agreement, and mood in English agrammatism. *Journal of Neurolinguistics*, 22, 436–450. doi:10.1016/j.jneuroling.2009.02.003
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25(2), 257–271. doi:10.3758/BF03204507
- Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, 27(3), 245–260. doi:10.1080/02643294.2010.513967
- Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small

- samples. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, 12, 482–486. doi:10.1076/clin.12.4.482.7241
- Eberhard, K. M. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language*, 36, 147–164. doi:10.1006/jmla.1996.2484
- Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject-verb agreement in English. *Journal of Memory and Language*, 41, 560–578. doi:10.1006/jmla.1999.2662
- Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, 112(3), 531–559. doi:10.1037/0033-295X.112.3.531
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331. doi:10.1037/0096-3445.128.3.309
- Faroqi-Shah, Y., & Thompson, C. K. (2004). Semantic, lexical, and phonological influences on the production of verb inflections in agrammatic aphasia. *Brain and Language*, 89(3), 484–498. doi:10.1016/j.bandl.2003.12.006
- Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject-verb agreement errors. A study in French written language. *The Quarterly Journal of Experimental Psychology Section A*, 47, 437–464. doi:10.1080/14640749408401119
- Francis, W. N. (1986). Proximity concord in English. *Journal of English Linguistics*, 19(2), 309–317. doi:10.1177/007542428601900212
- Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition*, 101(1), 173–216. doi:10.1016/j.cognition.2005.10.003
- Franck, J., Soare, G., Frauenfelder, U. H., & Rizzi, L. (2010). Object interference in subject-verb agreement: The role of intermediate traces of movement. *Journal of Memory and Language*, 62, 166–182. doi:10.1016/j.jml.2009.11.001
- Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, 17(4), 371–404. doi:10.1080/01690960
- Freedman, M. L., Martin, R. C., & Biegler, K. (2004). Semantic relatedness effects in conjoined noun phrase production: Implications for the role of short-term memory. *Cognitive Neuropsychology*, 21(2), 245–265. doi:10.1080/02643290342000528
- Friedmann, N., & Grodzinsky, Y. (1997). Tense and agreement in agrammatic production: Pruning the syntactic tree. *Brain and Language*, 56(3), 397–425. doi:10.1006/brln.1997.1795
- Gillespie, M., & Pearlmutter, N. J. (2011). Hierarchy and scope of planning in subject-verb agreement production. *Cognition*, 118(3), 377–397. doi:10.1016/j.cognition.2010.10.008
- Gillespie, M., & Pearlmutter, N. J. (2013). Against structural constraints in subject-verb agreement production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 515–528. doi:10.1037/a0029005
- Hamilton, A. C., & Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience*, 5(1), 1–13. doi:10.3758/CABN.5.1.1
- Hamilton, A. C., & Martin, R. C. (2007). Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex*, 43, 112–123. doi:10.1016/S0010-9452(08)70449-0
- Harris, J. A. (2015). Structure modulates similarity-based interference in sluicing: An eye tracking study. *Frontiers in Psychology*, 6, 1839. doi:10.3389/fpsyg.2015.01839
- Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, 21(1), 181–204. doi:10.1080/01690960400002117
- Hartsuiker, R. J., Kolk, H., & Huinck, W. J. (1999). Agrammatic production of subject-verb agreement: The effect of conceptual number. *Brain and Language*, 69(2), 119–160. doi:10.1006/brln.1999.2059
- Hoffman, P., Jefferies, E., & Lambon Ralph, M. A. (2011). Explaining semantic short-term memory deficits: Evidence for the critical role of semantic control. *Neuropsychologia*, 49, 368–381. doi:10.1016/j.neuropsychologia.2010.12.034
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446. doi:10.1016/j.jml.2007.11.007
- Kok, P., Van Doorn, A., & Kolk, H. (2007). Inflection and computational load in agrammatic speech. *Brain and Language*, 102(3), 273–283. doi:10.1016/j.bandl.2007.03.001
- Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149. doi:10.1016/j.jml.2015.02.002
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science*, 22, 1–75. doi:10.1017/S0140525X99001776
- Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375–419. doi:10.1207/s15516709cog0000_25
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454. doi:10.1016/j.tics.2006.08.007
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476–490. doi:10.3758/BF03210951
- Mallinson, G., & Blake, B. (1981). *Language typology: Cross-linguistic studies in syntax*. Amsterdam: North Holland.
- Martin, R., Shelton, J., & Yaffee, L. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, 33, 83–111. doi:10.1006/jmla.1994.1005

- Martin, R. C. (2005). Components of short-term memory and their relation to language processing. Evidence from neuropsychology and neuroimaging. *Current Directions in Psychological Science*, 14(4), 204–208. doi:10.1111/j.0963-7214.2005.00365.x
- Martin, R. C., & Allen, C. M. (2008). A disorder of executive function and its role in language processing. *Seminars in Speech and Language*, 29(3), 201–210; C 4–5. doi:10.1055/s-0028-1082884
- Martin, R. C., & Freedman, M. (2001). Short-term retention of lexical-semantic representations: Implications for speech production. *Memory*, 9(4–6), 261–280. doi:10.1080/09658210143000173
- Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language*, 89(1), 76–82. doi:10.1016/S0093-934X(03)00300-6
- Martin, R. C., Miller, M., & Vu, H. (2004). Lexical-semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology*, 21(6), 625–644. doi:10.1080/02643290342000302
- Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, 41, 3–29. doi:10.1006/jmla.1999.2637
- Martin, R. C., & Slevc, L. R. (2014). Language production and working memory. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 437–450). New York, NY: Oxford University Press.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67–91. doi:10.1016/S0749-596X(02)00515-6
- Nicenboim, B., Vasisht, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Frontiers in Psychology*, 6, 173. doi:10.3389/fpsyg.2015.00312
- Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, 26(6), 527–567. doi:10.1080/02643290903519367
- Patil, U., Hanne, S., Burchert, F., De Bleser, R., & Vasisht, S. (2016). A computational evaluation of sentence processing deficits in Aphasia. *Cognitive Science*, 40, 5–50. doi:10.1111/cogs.12250
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427–456. doi:10.1006/jmla.1999.2653
- Roach, A., Schwartz, M., Martin, N., Grewal, R., & Brecher, A. (1996). The Philadelphia naming test: Scoring and rationale. *Clinical Aphasiology*, 24, 121–133.
- Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, 72(3), 193–218. doi:10.1006/brln.1999.2285
- Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, 37(3), 440–479. doi:10.1016/0093-934X(89)90030-8
- Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1503–1514. doi:10.1037/a0024350
- Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, 49(1), 1–46. doi:10.1016/j.cogpsych.2003.10.001
- Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, 60, 308–327. doi:10.1016/j.jml.2008.11.002
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4), 355–370. doi:10.1016/j.jml.2004.01.001
- Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, 76, 195–215. doi:10.1016/j.jml.2014.07.003
- Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., ... Swick, D. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective, & Behavioral Neuroscience*, 2, 109–120. doi:10.3758/CABN.2.2.109
- Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, 48(4), 740–759. doi:10.1016/S0749-596X(03)00003-2
- Van Dyke, J. A. & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass* 6/4, 193–211. doi:10.1002/lnc3.330
- Veenstra, A., Meyer, A. S., & Acheson, D. J. (2015). Effects of parallel planning on agreement production. *Acta Psychologica*, 162, 29–39. doi:10.1016/j.actpsy.2015.09.011
- Vigliocco, G., Butterworth, B., Semenza, C., & Fossella, S. (1994). How two aphasic speakers construct subject-verb agreement. *Journal of Neurolinguistics*, 8(1), 19–25. doi:10.1016/0911-6044(94)90003-5
- Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, 68, B13–B29. doi:10.1016/S0010-0277(98)00041-9
- Vigliocco, G., & Zilli, T. (1999). Syntactic accuracy in sentence production: The case of gender disagreement in Italian language-impaired and unimpaired speakers. *Journal of Psycholinguistic Research*, 28(6), 623–648. doi:10.1023/A:1023268911312
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237. doi:10.1016/j.jml.2009.04.002
- Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37(3), 356–381. doi:10.1006/jmla.1997.2517

Appendix

Item	Local plausibility	Preamble [local noun: singular/plural]	Adjective
1	Plausible Implausible	The tree near the [statue/statues] The tree near the [meadow/meadows]	Enormous
2	Plausible Implausible	The chocolate in the [cookie/cookies] The chocolate in the [wrapper/wrappers]	Delicious
3	Plausible Implausible	The granola in the [cereal/cereals] The granola in the [container/containers]	Crunchy
4	Plausible Implausible	The flagpole next to the [tower/towers] The flagpole next to the [river/rivers]	Tall
5	Plausible Implausible	The flower by the [girl/girls] The flower by the [rock/rocks]	Beautiful
6	Plausible Implausible	The marble beside the [ball/balls] The marble beside the [book/books]	Round
7	Plausible Implausible	The insect under the [pebble/pebbles] The insect under the [boulder/boulders]	Small
8	Plausible Implausible	The seasoning in the [soup/soups] The seasoning in the [jar/jars]	Spicy
9	Plausible Implausible	The blade for the [razor/razors] The blade for the [blender/blenders]	Sharp
10	Plausible Implausible	The peak on the [mountain/mountains] The peak on the [graph/graphs]	High
11	Plausible Implausible	The lady with the [pet/pets] The lady with the [bag/bags]	Hungry
12	Plausible Implausible	The bunny next to the [doll/dolls] The bunny next to the [shed/sheds]	Cute
13	Plausible Implausible	The ice in the [freezer/freezers] The ice in the [sculpture/sculptures]	Cold
14	Plausible Implausible	The figure near the [alley/alleys] The figure near the [fountain/fountains]	Dark
15	Plausible Implausible	The glass in the [window/windows] The glass in the [frame/frames]	Clear
16	Plausible Implausible	The spell from the [witch/witches] The spell from the [scroll/scrolls]	Evil
17	Plausible Implausible	The sugar in the [cake/cakes] The sugar in the [bowl/bowls]	Sweet
18	Plausible Implausible	The assignment from the [textbook/textbooks] The assignment from the [teacher/teachers]	Long
19	Plausible Implausible	The furniture in the [house/houses] The furniture in the [sale/sales]	Old
20	Plausible Implausible	The box near the [window/windows] The box near the [table/tables]	Open
21	Plausible Implausible	The joke from the [comic/comics] The joke from the [newspaper/newspapers]	Funny
22	Plausible Implausible	The rash from the [fabric/fabrics] The rash from the [vaccine/vaccines]	Itchy
23	Plausible Implausible	The light from the [lamp/lamps] The light from the [clock/clocks]	Bright
24	Plausible Implausible	The ground under the [tire/tires] The ground under the [bush/bushes]	Flat
25	Plausible Implausible	The plate with the [egg/eggs] The plate with the [meat/meats]	Cracked
26	Plausible Implausible	The juice in the [smoothie/smoothies] The juice in the [tumbler/tumblers]	Tart
27	Plausible Implausible	The water in the [pool/pools] The water in the [glass/glasses]	Deep
28	Plausible Implausible	The riverbed near the [desert/deserts] The riverbed near the [forest/forests]	Dry
29	Plausible Implausible	The towel by the [shower/showers] The towel by the [oven/ovens]	Wet
30	Plausible Implausible	The charm of the [clover/clovers] The charm of the [flower/flowers]	Lucky
31	Plausible Implausible	The man with the [cat/cats] The man with the [tie/ties]	Fat
32	Plausible Implausible	The ball in the [bush/bushes] The ball in the [trail/trails]	Green
33	Plausible Implausible	The sticker on the [apple/apples] The sticker on the [window/windows]	Red

(Continued)

Appendix Continued.

Item	Local plausibility	Preamble [local noun: singular/plural]	Adjective
34	Plausible	The banana by the [lemon/lemons]	Yellow
	Implausible	The banana by the [onion/onions]	
35	Plausible	The outfit for the [kitten/kittens]	Adorable
	Implausible	The outfit for the [party/parties]	
36	Plausible	The diamond in the [ring/rings]	Shiny
	Implausible	The diamond in the [ad/ads]	
37	Plausible	The line to the [road/roads]	Long
	Implausible	The line to the [store/stores]	
38	Plausible	The down from the [coat/coats]	Soft
	Implausible	The down from the [bird/birds]	
39	Plausible	The sauce on the [burger/burgers]	Tasty
	Implausible	The sauce on the [platter/platters]	
40	Plausible	The water in the [pot/pots]	Hot
	Implausible	The water in the [tub/tubs]	
41	Plausible	The doll near the [shelf/shelves]	Wooden
	Implausible	The doll near the [pillow/pillows]	
42	Plausible	The blanket for the [sofa/sofas]	Comfortable
	Implausible	The blanket for the [picnic/picnics]	
43	Plausible	The towel by the [plate/plates]	Dirty
	Implausible	The towel by the [soap/soaps]	
44	Plausible	The food by the [oven/ovens]	Hot
	Implausible	The food by the [freezer/freezers]	